DOCUMENT RESUME

ED 449 163 TM 031 841

AUTHOR	Glass, Gene V., Ed.
TITLE	Education Policy Analysis Archives, 2000.
ISSN	ISSN-1068-2341
PUB DATE	2000-00-00
NOTE	984p.; "Education Policy Analysis Archives" is an electronic-only journal covered on an article-by-article basis in "Current Index to Journals in Education." (CIJE). For descriptions of articles in Volume 8, see TM 522 614-622, TM 522 947-963, TM 523 139-152, and TM 523 216-222 (CIJE0. The Associate Editor for Spanish Language is Roberto Rodriguez Gomez.
AVAILABLE FROM	For full text:: http://olam.ed.asu.edu/epaa.
PUB TYPE	Collected Works - Serials (022)
LANGUAGE	English, Spanish
JOURNAL CIT	Education Policy Analysis Archives; v8 n53 2000
EDRS PRICE	MF07/PC40 Plus Postage.
DESCRIPTORS	*Academic Achievement; Achievement Gains; Court Litigation; Educational Change; Educational Finance; *Educational Policy; Educational Research; Elementary Secondary Education; Equal Education; *Faculty Development; Foreign Countries; Higher Education; Low Income Groups; Rural Schools; Sex Differences; *State Standards; *Student Evaluation; Tax Credits; *Teacher Effectiveness; Teacher Salaries; Testing Programs; User Needs (Information)
IDENTIFIERS	Arizona; *ERIC; Texas; Uruguay

ABSTRACT

This document consists of the 2000 edition of the "Education Policy Analysis Archives." The papers include: (1) "Teacher Quality and Student Achievement: A Review of State Policy Evidence" (Linda Darling-Hammond); (2) "America Y2K: The Obsolescence of Educational Reforms" (Sherman Dorn); (3) "Forces for Change in Mathematics Education: The Case of TIMSS" (Donald S. Macnab); (4) "The Influence of Scale on School Performance: A Multi-Level Extension of the Matthew Principle" (Robert Bickel and Craig Howley); (5) "Student Assessment as a Political Construction: The Case of Uruguay" (Luis Benveniste); (6) "Implementation of the Kentucky Nongraded Primary Program" (Patricia J. Kannapel, Lola Aagaard, Pamelia Coe, and Cynthia A. Reeves); (7) "Should Achievement Tests Be Used To Judge School Quality?" (Scott C. Bauer); (8) "What Do Test Scores in Texas Tell Us?" (Stephen P. Klein, Laura S. Hamilton, Daniel F. McCaffrey, and Brian M. Stecher); (9) "Apoyo a la Participacion de padres en las escuelas primarias: Un estudio etnografico sobre un grupo latinoamericano en Canada" (Judith K. Bernhard, Marlinda Freire, and Veronica Pacini-Ketchabaw); and (10) "The Use of Logic in Educational Research and Policy Making" (Rick Garlikov). (AA)

Reproductions supplied by EDRS are the best that can be made
from the original document.

Education Policy Analysis Archives. 2000

Volume 8, Numbers 1-53

Gene V. Glass, Editor

ED 449 163

U.S. DEPARTMENT OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER ERIC

G. Glass

TM 031 841

BEST COPY AVAILABLE

This article has been retrieved 6906 times since January 1, 2000

Education Policy Analysis Archives

Volume 8 Number 1 January 1, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Teacher Quality and Student Achievement: A Review of State Policy Evidence

Linda Darling-Hammond
Stanford University

Abstract

Using data from a 50-state survey of policies, state case study analyses, the 1993-94 Schools and Staffing Surveys (SASS), and the National Assessment of Educational Progress (NAEP), this study examines the ways in which teacher qualifications and other school inputs are related to student achievement across states. The findings of both the qualitative and quantitative analyses suggest that policy investments in the quality of teachers may be related to improvements in student performance. Quantitative analyses indicate that measures of teacher preparation and certification are by far the strongest correlates of student achievement in reading and mathematics, both before and after controlling for student poverty and language status. State policy surveys and case study data are used to evaluate policies that influence the overall level of teacher qualifications within and across states. This analysis suggests that policies adopted by states regarding teacher education, licensing, hiring, and professional development may make an important difference in the qualifications and capacities that teachers bring to their work. The implications for state efforts to enhance quality and equity in public education are discussed. (Note 1)

Introduction

For many years, educators and researchers have debated which school variables influence student achievement. As policymakers become more involved in school reform, this question takes on new importance since their many initiatives rely on presumed relationships between various education-related factors and learning outcomes. Some research has suggested that "schools bring little influence to bear upon a child's achievement that is independent of his background and general social context" (Coleman et al., 1966, p. 325; see also Jencks et al., 1972). Other evidence suggests that factors like class size (Glass et al., 1982; Mosteller, 1995), teacher qualifications (Ferguson, 1991), school size (Haller, 1993), and other school variables may play an important role in what students learn.

As new standards for student learning have been introduced across the states, greater attention has been given to the role that teacher quality plays in student achievement (National Commission on Teaching and America's Future, 1996; National Education Goals Panel, 1998). In the last few years, more than 25 states have enacted legislation to improve teacher recruitment, education, certification, or professional development (Darling-Hammond, 1997a). While some evidence suggests that better qualified teachers may make a difference for student learning at the classroom, school, and district levels, there has been little inquiry into the effects on achievement that may be associated with large-scale policies and institutional practices that affect the overall level of teachers' knowledge and skills in a state or region. This paper reports on one such study, which combines state level case studies and quantitative analyses of state-level achievement data to examine whether and how state policies may influence teachers' capabilities and student learning.

Using data from a 50-state policy survey conducted by the National Commission on Teaching and America's Future, case studies of selected states conducted under the auspices of the Center for the Study of Teaching and Policy, the 1993-94 Schools and Staffing Surveys (SASS), and the National Assessment of Educational Progress (NAEP) sponsored by the National Center for Education Statistics, the study examines the ways in which teacher qualifications and other school inputs, such as class size, are related to student achievement across states, taking student characteristics into account. In addition, these data and state case study data are used to evaluate policies that influence the overall level of teacher qualifications within and across states.

Previous Research

Despite conventional wisdom that school inputs make little difference in student learning, a growing body of research suggests that schools can make a difference, and a substantial portion of that difference is attributable to teachers. Recent studies of teacher effects at the classroom level using the Tennessee Value-Added Assessment System and a similar data base in Dallas, Texas, have found that differential teacher effectiveness is a strong determinant of differences in student learning, far outweighing the effects of differences in class size and heterogenity (Sanders & Rivers, 1996; Wright, Horn, & Sanders, 1997; Jordan, Mendro, & Weerasinghe, 1997). Students who are assigned to several ineffective teachers in a row have significantly lower achievement and gains in achievement than those who are assigned to several highly effective teachers in sequence (Sanders & Rivers, 1996). Teacher effects appear to be additive and cumulative, and generally not compensatory. These studies also find troubling indicators for educational equity, noting evidence of strong bias in assignment of students to teachers of different effectiveness levels (Jordan, Mendro, & Weerasinghe, 1997), including indications that African American students are nearly twice as likely to be assigned to the most ineffective teachers and half as likely to be assigned to the most effective teachers (Sanders & Rivers, 1996). These studies did not, however, examine the characteristics or practices of more and less effective teachers.

These issues have been the topic of much other research over the last 50 years. Variables presumed to be indicative of teachers' competence which have been examined for their relationship to student learning include measures of academic ability, years of education, years of teaching experience, measures of subject matter and teaching knowledge, certification status, and teaching behaviors in the

classroom. The results of these studies have been mixed; however, some trends have emerged in recent years.

General Academic Ability and Intelligence While studies as long ago as the 1940s have found positive correlations between teaching performance and measures of teachers' intelligence (usually measured by IQ) or general academic ability (Hellfritsch, 1945; LaDuke, 1945; Rostker, 1945; Skinner, 1947), most relationships are small and statistically insignificant. Two reviews of such studies concluded that there is little or no relationship between teachers' measured intelligence and their students' achievement (Schalock, 1979; Soar, Medley, & Coker, 1983). Explanations for the lack of strong relationship between measures of IQ and teacher effectiveness have included the lack of variability among teachers in this measure and its tenuous relationship to actual performance (Vernon, 1965; Murnane, 1985). However, other studies have suggested that teachers' verbal ability is related to student achievement (e.g., Bowles & Levin, 1968; Coleman et al., 1966; Hanushek, 1971), and that this relationship may be differentially strong for teachers of different types of students (Summers & Wolfe, 1975). Verbal ability, it is hypothesized, may be a more sensitive measure of teachers' abilities to convey ideas in clear and convincing ways (Murnane, 1985).

Subject Matter Knowledge Subject matter knowledge is another variable that one might think could be related to teacher effectiveness. While there is some support for this assumption, the findings are not as strong and consistent as one might suppose. Studies of teachers' scores on the subject matter tests of the National Teacher Examinations (NTE) have found no consistent relationship between this measure of subject matter knowledge and teacher performance as measured by student outcomes or supervisory ratings. Most studies show small, statistically insignificant relationships, both positive and negative (Andrews, Blackmon & Mackey, 1980; Ayers & Qualls, 1979; Haney, Madaus, & Kreitzer, 1986; Quirk, Witten, & Weinberg, 1973; Summers & Wolfe, 1975).

Byrne (1983) summarized the results of thirty studies relating teachers' subject matter knowledge to student achievement. The teacher knowledge measures were either a subject knowledge test (standardized or researcher-constructed) or number of college courses taken within the subject area. The results of these studies were mixed, with 17 showing a positive relationship and 14 showing no relationship. However, many of the "no relationship" studies, Byrne noted, had so little variability in the teacher knowledge measure that insignificant findings were almost inevitable. Ashton and Crocker (1987) found only 5 of 14 studies they reviewed exhibited a positive relationship between measures of subject matter knowledge and teacher performance.

It may be that these results are mixed because subject matter knowledge is a positive influence up to some level of basic competence in the subject but is less important thereafter. For example, a controlled study of middle school mathematics teachers, matched by years of experience and school setting, found that students of fully certified mathematics teachers experienced significantly larger gains in achievement than those taught by teachers not certified in mathematics. The differences in student gains were greater for algebra classes than general mathematics (Hawk, Coble, & Swanson, 1985). However, Begle and Geeslin (1972) found in a review of mathematics teaching that the absolute number of course credits in mathematics was not linearly related to teacher performance.

It makes sense that knowledge of the material to be taught is essential to good teaching, but also that returns to subject matter expertise would grow smaller beyond some minimal essential level which exceeds the demands of the curriculum being taught. This interpretation is supported by Monk's (1994) more recent study of mathematics and science achievement. Using data on 2,829 students from the Longitudinal Study of American Youth, Monk (1994) found that teachers' content preparation, as measured by coursework in the subject field, is positively related to student achievement in mathematics and science but that the relationship is curvilinear, with diminishing returns to student achievement of teachers' subject matter courses above a threshold level (e.g., five courses in mathematics).

In a multilevel analysis of the same data set, Monk and King (1994) found both positive and negative, generally insignificant effects of teachers' subject matter preparation on student achievement. They did find some evidence of cumulative

effects of prior as well as proximate teachers' subject matter preparation on student performance in mathematics. Effects differed for high- and low-achieving students and for different grade levels. In a review of 65 studies of science teachers' characteristics and behaviors, Druva and Anderson (1983) found students' science achievement was positively related to the teachers' course taking background in both education and in science. The relationship between teachers' training in science and student achievement was greater in higher level science courses, a result similar to that found by Hawk, Coble, and Swanson (1985) in mathematics.

It may also be that the measure of subject matter knowledge makes a difference in the findings. Measures of course-taking in a subject area have more frequently been found to be related to teacher performance than have scores on tests of subject matter knowledge. This might be because tests necessarily capture a narrower slice of any domain. Furthermore, in the United States, most teacher tests have used multiple-choice measures that are not very useful for assessing teachers' ability to analyze and apply knowledge. More authentic measures may capture more of the influence of subject matter knowledge on student learning. For example, a test of French language teachers' speaking skill was found to have significant correlation to students' achievement in speaking and listening (Carroll, 1975).

Despite concerns that education majors may be less well prepared in their subject areas than are academic majors (Galambos, 1985), comparisons of teachers with degrees in education vs. those with degrees in disciplinary fields have found no relationship between degree type and teacher performance (Murnane, 1985). This may be because certification requirements reduce the variability in course backgrounds found for teachers with different degree types. For example, many states require the equivalent of an academic major or minor in the field to be taught as part of the education degree for high school teachers, regardless of the department granting the degree (NASDTEC, 1997). Given the standardizing influences of licensing requirements within states but substantial differences in licensing requirements across states, within-state studies are likely to find less variation in teachers' education backgrounds than might be found in cross-state studies.

Knowledge of Teaching and Learning Studies have found a somewhat stronger and more consistently positive influence of education coursework on teachers' effectiveness. Ashton and Crocker (1987) found significant positive relationships between education coursework and teacher performance in four of seven studies they reviewed—a larger share than those showing subject matter relationships. Evertson, Hawley, and Zlotnik (1985) reported a consistent positive effect of teachers' formal education training on supervisory ratings and student learning, with 11 of 13 studies showing greater effectiveness for fully prepared and certified vs. uncertified or provisionally certified teachers. With respect to subject matter coursework, 5 of 8 studies they reviewed found no relationship and the other three found small associations.

Reviewing findings of the National Longitudinal Study of Mathematical Abilities, Begle (1979) found that the number of credits a teacher had in mathematics methods courses was a stronger correlate of student performance than was the number of credits in mathematics courses or other indicators of preparation. Similarly, Monk's (1994) study of student's mathematics and science achievement found that teacher education coursework had a positive effect on student learning and was sometimes more influential than additional subject matter preparation. In an analysis of science teaching, Perkes (1967- 68) found that teachers' coursework credits in science were not significantly related to student learning, but coursework in science education was significantly related to students' achievement on tasks requiring problem solving and applications of science knowledge. Teachers with greater training in science teaching were more likely to use laboratory techniques and discussions and to emphasize conceptual applications of ideas, while those with less education training placed more emphasis on memorization.

In a study of more than 200 graduates of a single teacher education program, Ferguson and Womack (1993) examined the influences on 13 dimensions of teaching performance of education and subject matter coursework, NTE subject matter test scores, and GPA in the student's major. They found that the amount of education coursework completed by teachers explained more than four times the variance in teacher performance (16.5 percent) than did measures of content

knowledge (NTE scores and GPA in the major), which explained less than 4 percent. In a similar study which compared relative influences of different kinds of knowledge on 12 dimensions of teacher performance for more than 270 teachers, Guyton and Farokhi (1987) found consistent strong, positive relationships between teacher education coursework performance and teacher performance in the classroom as measured through a standardized observation instrument, while relationships between classroom performance and subject matter test scores were positive but insignificant and relationships between classroom performance and basic skill scores were almost nonexistent. Another program-based study by Denton and Lacina (1984) found positive relationships between the extent of teachers' professional education coursework and their teaching performance, including their students' achievement.

It may be that the positive effects of subject matter knowledge are augmented or offset by knowledge of how to teach the subject to various kinds of students. That is, the degree of pedagogical skill may interact with subject matter knowledge to bolster or reduce teacher performance. As Byrne (1983) suggested:

> It is surely plausible to suggest that insofar as a teacher's knowledge provides the basis for his or her effectiveness, the most relevant knowledge will be that which concerns the particular topic being taught and the relevant pedagogical strategies for teaching it to the particular types of pupils to whom it will be taught. If the teacher is to teach fractions, then it is knowledge of fractions and perhaps of closely associated topics which is of major importance.... Similarly, knowledge of teaching strategies relevant to teaching fractions will be important. (p. 14)

The kind and quality of in-service professional development as well as pre-service education may make a difference in developing this knowledge. Several recent studies have found that higher levels of student achievement are associated with mathematics teachers' opportunities to participate in sustained professional development grounded in content-specific pedagogy linked to the new curriculum they are learning to teach (Cohen & Hill, 1997; Wiley & Yoon, 1995; Brown, Smith, & Stein, 1995). In these studies, both the kind and extent of professional development mattered for teaching practice and for student achievement.

The National Assessment of Educational Progress has also documented how specific kinds of teacher learning opportunities correlate with their students' reading achievement. On average, in the 1992 and 1994 assessments, 4th grade students of teachers who were fully certified, who had master's degrees, and who had had professional coursework in literature-based instruction did better than other students on reading assessments (NCES, 1994; NCES, n.d.). While these relationships were modest, the relationships between specific teaching practices and student achievement were often quite pronounced, and these practices were in turn related to teacher learning opportunities. NAEP analyses found that teachers who had had more professional training were more likely to use teaching practices that are associated with higher reading achievement on the NAEP tests--use of trade books and literature, integration of reading and writing, and frequent visits to the library--and were less likely to engage in extensive of use of reading kits, basal readers, workbooks, and multiple choice tests for assessing reading, practices that the NAEP analyses found to be associated with lower levels of student achievement. Interestingly, students of teachers who had had more training in phonics instruction did noticeably less well than other students in both years. Often, this kind of training, narrowly cast, is focused heavily on the use of basal readers and workbooks rather than an integrated approach that teaches decoding skills in the context of other important reading skills and language development strategies.

Other studies have found that students achieve at higher levels and are less likely to drop out when they are taught by teachers with certification in their teaching field, by those with master's degrees, and by those enrolled in graduate studies (Council for School Performance, 1997; Knoblock, 1986; Sanders, Skonie-Hardin, & Phelps, 1994). However, like the NAEP analyses described above,

BEST COPY AVAILABLE

these are simple correlational analyses that do not take into account other school resources or student characteristics like poverty or language background that may affect student performance.

Continuity of teachers' learning may also matter. In earlier work, Hanushek (1971) demonstrated that the *recency* of voluntary educational experience was related to teacher performance. Penick and Yager (1983) found that teachers in exemplary science programs had higher levels of education and more recent educational experiences than others, even though they were older than the average science teacher. As Murnane (1985) suggests, these findings may indicate that it is not only the knowledge acquired with ongoing professional development (which may represent more recent advances in the knowledge base) but also the teacher's enthusiasm for learning that relates to increased student achievement.

Teaching Experience Other studies of the effects of teacher experience on student learning have found a relationship between teachers' effectiveness and their years of experience (Murnane & Phillips, 1981; Klitgaard & Hall, 1974), but not always a significant one or an entirely linear one. While many studies have established that inexperienced teachers (those with less than three years of experience) are typically less effective than more senior teachers, the benefits of experience appear to level off after about five years, especially in non-collegial work settings (Rosenholtz, 1986). A possible cause of this curvilinear trend in experience effects is that older teachers do not always continue to grow and learn and may grow tired in their jobs. Furthermore, the benefits of experience may interact with educational opportunities. Veteran teachers in settings that emphasize continual learning and collaboration continue to improve their performance (Rosenholtz, 1984). Similarly, very well-prepared beginning teachers can be highly effective. For example, some recent studies of 5-year teacher education programs--programs that include a bachelor's degree in the discipline and master's in education as well as a year-long student teaching placement--have found graduates to be more confident than graduates of 4-year programs and as effective as more senior teachers (Andrew & Schwab, 1995; Denton & Peters, 1988).

It is also possible that uneven effects of experience in cross-sectional studies can be the result of cohort effects (for example, cohorts of teachers hired in times of shortage may be less well-qualified than those hired when schools can be more selective) or of attrition effects (for example, disproportionate early attrition of more able teachers may leave a less capable senior force on average) (Murnane & Phillips, 198; Vance & Schlechty, 1982). Presumably, the direction of this effect would change if retention policies kept the most able beginning teachers in the profession. Since experience is also correlated with teacher education and certification status, these variables may be confounded in some analyses.

Certification Status Certification or licensing status is a measure of teacher qualifications that combines aspects of knowledge about subject matter and about teaching and learning. Its meaning varies across the states because of differences in licensing requirements, but a standard certificate generally means that a teacher has been prepared in a state-approved teacher education program at the undergraduate or graduate level and has completed either a major or a minor in the field(s) to be taught plus anywhere from 18 to 40 education credits, depending on the state and the certificate area, including between 8 and 18 weeks of student teaching. (The norm is about 30 education credits and about 12 to 15 weeks of student teaching.) There are only a few states that have requirements outside these parameters; however, individual teacher education programs often require more preparation than the state demands in education, in clinical practice, and in the content area(s) to be taught. Most states now also require one or more tests of basic skills, subject matter knowledge, and/or teaching knowledge or skills as the basis for the initial or continuing license or for admission to teacher education.

While most states have been increasing their standards since the 1980s, more than 30 states still allow the hiring of teachers who have not met their licensing standards, a practice that has been on the increase in some states as demand has grown in recent years. Some allow the hiring of teachers with no license. Others issue emergency, temporary, or provisional licenses to candidates who, depending on the state, may or may not have met varying requirements (e.g., a bachelors degree, a certificate in another teaching field, a basic skills test). More than 40 states

have also initiated alternate route provisions for candidates who enter through postbaccalaureate programs. Most of these are master's degree programs which offer an education degree that meets all of the normal state requirements but does so in a fashion tailored to individuals, like mid-career entrants, who already have a bachelor's degree. Some states allow candidates to complete a short summer course of study and assume full teaching responsibilities, with or without completing additional coursework.

In times of relatively low demand, like most of the 1980s, virtually all teachers were certified and there was too little variability to find effects of this variable in large-scale studies. Most studies of the influence of training and certification on teacher performance are from the high-demand era of the 1960s and 1970s and from the 1990s when demand increased again. Studies in different subject matter fields that compare teachers with and without preparation have typically found higher ratings and greater student learning gains for teachers who have more formal preparation for teaching. In addition to the studies of science and mathematics teachers cited earlier, these include reading and elementary education (Hice, 1970; LuPone, 1961; McNeil, 1974), early childhood education (Roupp et al., 1979), gifted education (Hansen, 1988), and vocational education (Erekson and Barr, 1985). In a review of research, Evertson, Hawley, and Zlotnik (1985) concluded:

> (T)he available research suggests that among students who become teachers, those enrolled in formal preservice preparation programs are more likely to be effective than those who do not have such training. Moreover, almost all well planned and executed efforts within teacher preparation programs to teach students specific knowledge or skills seem to succeed, at least in the short run (p.8).

Other studies point out the differences in the perceptions and practices of teachers with differing amounts and kinds of preparation. A number of studies suggest that the typical problems of beginning teachers are lessened for those who have had adequate preparation prior to entry (Adams, Hutchinson, & Martray, 1980; Glassberg, 1980; Taylor & Dale, 1971). Studies of teachers admitted with less than full preparation--with no teacher preparation or through very short alternate routes--have found that such recruits tend to be less satisfied with their training (Darling-Hammond, Hudson, & Kirby, 1987; Jelmberg, 1995), and they tend to have greater difficulties planning curriculum, teaching, managing the classroom, and diagnosing students' learning needs (Bents & Bents, 1990; Darling-Hammond, 1992; Lenk, 1989; Feiman-Nemser & Parker, 1990; Gomez & Grobe, 1990; Grady, Collins, & Grady, 1991; Grossman, 1989; Mitchell, 1987; National Center for Research on Teacher Learning, 1992; Rottenberg & Berliner, 1990). Principals, supervisors, and colleagues tend to rate them less highly on their instructional skills (Bents & Bents, 1990; Jelmberg, 1995; Lenk, 1989; Feiman-Nemser & Parker, 1990; Gomez & Grobe, 1990; Mitchell, 1987; Texas Education Agency, 1993), and they tend to leave teaching at higher-than- average rates (Darling-Hammond, 1992; Lutz & Hutton, 1989; Stoddart, 1992).

These findings are reflected in Gomez and Grobe's (1990) study of the performance of alternate route candidates in Dallas, who receive a few weeks of summer training from the district before they assume full teaching responsibilities. Although these candidates were rated near the average on some aspects of teaching, they were rated lower on such factors as their knowledge of instructional techniques and instructional models. The performance of alternate route candidates was also much more uneven than that of trained teachers, with a much greater proportion of them--from 2 to 16 times as many--rated "poor" on each of the teaching factors evaluated. The strongest effects of this unevenness were seen in students' achievement in language arts, where the achievement gains of students of alternate route teachers, adjusted for initial student scores, were significantly lower than those of students of traditionally trained teachers.

Two studies of alternative certification in Texas have reportedly failed to find such gaps in the performance of students of alternative and traditionally licensed

teachers (cited in Goldhaber & Brewer, 1999). A study of Houston's alternative certification program by Goebel, Romacher, and Sanchez (1989) reported no evidence of differential student outcomes and little evidence of teacher effects. However, this study did not control for students' initial test scores and did not match comparison teachers by years of experience. First year traditionally trained teachers were compared to two groups of alternative certification recruits, one with 1-4 years of experience and the other with 5-7 years of experience. Thus, this study did not include adequate controls to allow measurement of effects. Another study by Barnes, Salmon, and Wale (1989) reported second-hand that two districts reported equivalent outcomes for alternative and traditional program teachers but did not present any empirical data or discussion of methodology. The study's table listing program types evaluated included 1 to 2-year university-based master's programs (which are called "alternative" in Texas because they are not undergraduate models) as well as district alternative programs that generally offer only a few weeks of summer training. In this case, the "alternative" group included programs providing extensive graduate level training along with those with very little preparation, thus preventing assessment of the effects of preparation on teacher effectiveness. With non-comparable groups and no controls, it is impossible to draw inferences from either of these studies.

Some recent multivariate studies of student achievement at the school and district level have found a substantial influence of teachers' qualifications on what students learn, especially when scores on licensing examinations are included. In an analysis of nearly 900 Texas school districts that evaluated the effects of many school input variables and controlled for student background and district characteristics, Ronald Ferguson (1991) found that combined measures of teachers' expertise--scores on a licensing examination, master's degrees, and experience--accounted for more of the inter-district variation in students' reading and mathematics achievement (and achievement gains) in grades 1 through 11 than student socioeconomic status. An additional, smaller contribution to student achievement was made by lower pupil-teacher ratios and smaller schools in the elementary grades.

Of the teacher qualifications variables, the strongest relationship was found for scores on the state licensing examination, a test that measures both basic skills and teaching knowledge. The effects were so strong, and the variations in teacher expertise so great, that after controlling for socioeconomic status, the large disparities in achievement between black and white students were almost entirely accounted for by differences in the qualifications of their teachers. Ferguson also found that every additional dollar spent on more highly qualified teachers netted greater increases in student achievement than did less instructionally focused uses of school resources.

Another study (Strauss & Sawyer, 1986) found that North Carolina's teachers' average scores on the National Teacher Examinations (a licensing test which measures subject matter and teaching knowledge) had a strong influence on average school district test performance. Taking into account per-capita income, student race, district capital assets, student plans to attend college, and pupil/teacher ratios, teachers' test scores had a strikingly large effect on students' failure rates on the state competency examinations: a 1% increase in teacher quality (as measured by NTE scores) was associated with a 3 to 5% decline in the percentage of students failing the exam. The authors' conclusion is similar to Ferguson's:

> Of the inputs which are potentially policy-controllable (teacher quality, teacher numbers via the pupil-teacher ratio and capital stock), our analysis indicates quite clearly that improving the quality of teachers in the classroom will do more for students who are most educationally at risk, those prone to fail, than reducing the class size or improving the capital stock by any reasonable margin which would be available to policy makers (p. 47).

Ferguson and Helen Ladd (1996) conducted an analysis in Alabama similar to Ferguson's Texas study using a less extensive data set that included rougher proxies for teacher knowledge (master's degrees and ACT scores instead of teacher licensing

examination scores). They found somewhat smaller influences of these test scores, which are pre-college measures of general academic ability, compared to the licensing examinations in Texas, and somewhat larger influences of master's degrees. Together, teachers' academic ability, education, and experience, when combined with class sizes, accounted for 31.5% of the predicted difference in reading and mathematics student achievement gains between districts scoring in the top and bottom quartiles in mathematics, while 29.5% was explained by poverty, race, and parent education.

When student characteristics are held constant, the relationship of teachers' qualifications to student achievement is even more pronounced. A study of high- and low-achieving schools with demographically similar student populations in New York City found that differences in teacher qualifications (educational degrees, certification status, and experience) accounted for approximately 90% of the total variation in average school-level student achievement in reading and mathematics at all grade levels tested (Armour-Thomas et al., 1989).

A study of high school students' performance in mathematics and science using data from the National Educational Longitudinal Studies of 1988 (NELS) found that fully certified teachers have a statistically significant positive impact on student test scores relative to teachers who are not certified in their subject area, as do teachers who hold a degree in mathematics or mathematics education (Goldhaber & Brewer, 1999). Furthermore, in states with licensing examinations, newly trained teachers (those with probationary licenses granted to fully qualified new entrants) have a strong positive influence on student achievement. In an unusual finding, the study indicated that teachers with emergency certificates in science had higher-scoring students after other teacher education and student demographic variables were controlled. However, because there were only 23 such teachers in the sample of more than 2900 and more than 20 variables simultaneously tested in the equations, many of them highly correlated with certification status, it is difficult to know what to make of this finding. In cases like this, small cell sizes and multicollinearity problems often combine to produce sign changes and poor estimates of effects.

A more recent Texas study (Fuller, 1999) found that students in districts with greater proportions of licensed teachers were significantly more likely to pass the Texas state achievement tests, after controlling for student socioeconomic status, school wealth, and teacher experience. Teacher licensing was especially influential on the test performance of elementary students. In a recent school level analysis of mathematics test performance in California high schools, Fetler (1999) found a strong negative relationship between average student scores and the percentage of teachers on emergency certificates, as well as a smaller positive relationship between student scores and teacher experience levels, after controlling for student poverty rates.

These findings about the influences and relative contributions of teacher training and experience levels are reinforced by those of a recent review of 60 production function studies (Greenwald, Hedges, & Laine, 1996), which found that teacher education, ability, and experience, along with small schools and lower teacher-pupil ratios, are associated with increases in student achievement across schools and districts. In their estimate of the achievement gains associated with expenditure increments on various resources, spending on teacher education was found to be the most productive investment for schools, outstripping the effect of teacher experience and reduced pupil/teacher ratios.

Teacher Behaviors and Practices While these studies suggest that there are aspects of teaching effectiveness that may be related to teacher education, certification status, and experience, they do not reveal much about what it is about teachers' behaviors or abilities that makes the difference in how their students perform. Research on teachers' personality traits and behaviors has produced few consistent findings (Schalock, 1979; Druva & Anderson, 1983), with the exception of studies finding a recurring positive relationship between student learning and teachers' "flexibility," "creativity," or "adaptability" (Berliner & Tikunoff, 1976; Schalock, 1979; Walberg & Waxman, 1983). Successful teachers tend to be those who are able to use a range of teaching strategies and who use a range of interaction styles, rather than a single, rigid approach (Hamachek, 1969). This finding is

consistent with other research on effective teaching, which suggests that effective teachers adjust their teaching to fit the needs of different students and the demands of different instructional goals, topics, and methods (Doyle, 1985).

In addition to the ability to create and adapt instructional strategies, strong research support has linked student learning to variables such as teacher clarity, enthusiasm, task-oriented behavior, variability of lesson approaches, and student opportunity to learn criterion material. Teachers' abilities to structure material, ask higher order questions, use student ideas, and probe student comments have also been found to be important variables in what students learn (Rosenshine & Furst, 1973; Darling-Hammond, Wise, & Pease, 1983; Good & Brophy, 1986). No single instructional strategy has been found to be unvaryingly successful; instead, teachers who are able to use a broad repertoire of approaches skillfully (e.g., direct and indirect instruction, experience-based and skill-based approaches, lecture and small group work) are typically most successful. The use of different strategies occurs in the context of "active teaching" that is purposeful and diagnostic rather than random or laissez faire and that responds to students' needs as well as curriculum goals (Good, 1983).

Teacher education appears to influence the use of these practices. Teachers who have had formal preparation have been found to be better able to use teaching strategies that respond to students' needs and learning styles and that encourage higher order learning (Perkes, 1967-68; Hansen, 1988; Skipper & Quantz, 1987). Doyle (1986) hypothesizes that since the novel tasks required for problem-solving are more difficult to manage than the routine tasks associated with rote learning, lack of knowledge about how to manage an active, inquiry-oriented classroom can lead teachers to turn to passive tactics that "dumb down" the curriculum (see also Carter & Doyle, 1987), busying students with workbooks rather than complex tasks that require more skill to orchestrate (Cooper & Sherk, 1989).

It seems logical that teachers' abilities to handle the complex tasks of teaching for higher-level learning are likely to be associated, to varying extents, with each of the variables reviewed above: verbal ability, adaptability and creativity, subject matter knowledge, understanding of teaching and learning, specific teaching skills, and experience in the classroom, as well as interactions among these variables. In addition, considerations of fit between the teaching assignment and the teacher's knowledge and experience are likely to influence teachers' effectiveness (Little, 1999), as are conditions that support teachers' individual teaching and the additive effect of teaching across classrooms, such as class sizes and pupil loads, planning time, opportunities to plan and problem solve with colleagues, and curricular supports including appropriate materials and equipment (Darling-Hammond, 1997b).

Differences in State Policies Regarding Teaching

Despite logical presumptions and research evidence that student learning depends substantially on what teachers know and can do, states differ greatly in the extent to which they invest in teachers' learning as a key policy lever. At the front end of the career, there is wide variation in the standards to which entering teachers and teacher education institutions are held. Licensing standards are noticeably different from state to state, as are state commitments to enforcing these standards. Later access to professional development is also widely disparate.

In high-standards states like Wisconsin or Minnesota, for example, a prospective high school teacher must complete a bachelor's degree that includes a full major in the subject area to be taught plus coursework covering learning theory, child and adolescent development, subject matter teaching methods, curriculum, effective teaching strategies, uses of technology, classroom management, behavior and motivation, human relations, and the education of students with special needs. In the course of this work, the teacher must complete at least 18 weeks of student teaching in Wisconsin (at least a college semester in Minnesota) under the supervision of a cooperating teacher who meets minimum standards. In Minnesota, this experience must include work in a multicultural setting and with special needs students. If teachers are asked to teach outside the field of their major for part of the day, they must already be licensed with at least a minor in that field, and can receive a temporary license in the new field only briefly while completing a major. By

contrast, in Louisiana, prospective high school teachers can be licensed without even a minor in the field they will be teaching. The state does not require them to have studied curriculum, teaching strategies, classroom management, uses of technology, or the needs of special education students, and they can receive a license with only six weeks of student teaching (NASDTEC, 1997; Darling-Hammond, 1997a).

In addition to differences in the standards themselves, there are great differences in the extent to which they are enforced. Whereas some states do not allow districts to hire unqualified teachers, others routinely allow the hiring of candidates who have not met their standards, even when qualified teachers are available. In Wisconsin and eleven other states, for example, no new elementary or secondary teachers were hired without a license in their field in 1994. By contrast, in Louisiana, 31% of new entrants were unlicensed and another 15% were hired on substandard licenses. At least six other states allowed 20% or more of new public school teachers to be hired without a license in their field (Darling-Hammond, 1997a, Appendix A). Studies of teacher hiring show that even when there are an adequate number of qualified teachers in the labor market--which was the case nationally and in most states from the early 1980s through the mid-1990s--some districts hire unlicensed teachers because of cumbersome and poorly managed hiring procedures that discourage qualified entrants, perennially late hiring (e.g. waiting until late August or September to hire), patronage hiring, preferences for hiring lower salaried staff, and inequalities in salary schedules caused by state funding formulas and by local decisions to use budgets for purposes other than teacher salaries (see e.g. Haberman, 1995; Johanson and Gips, 1992; Pflaum and Abramson, 1990; National Commission on Teaching and America's Future; Wise, Darling-Hammond, and Berry, 1987).

More than 30 states allow teachers to be hired on temporary or emergency licenses without having completed preparation or having met other licensing requirements. During the late 1980s and early 1990s, at least 50,000 emergency or substandard licenses were issued annually by states (NCTAF, 1996). Nationally, in 1994, 27% of those who were new entrants into public school teaching held no license or a substandard license in their main teaching field (Darling-Hammond, 1997a). Even the rigor of these restricted licenses varies. States such as Minnesota will issue a restricted license only to a teacher who has already been fully prepared in a teaching field but who needs to complete additional coursework in order to enter from out-of-state or switch to a new field or teaching level. Such a license is only good for one year, while the necessary coursework is completed. Others, including Louisiana, will issue an emergency license to a person who does not even hold a bachelor's degree and will allow it to be renewed for several years while the candidate makes little progress toward becoming licensed.

It is certainly true that differences in student enrollment growth, coupled with teacher production rates and attrition, construct different levels of teacher demand that can affect the ease or difficulty of hiring within states. While incentives to enter and stay in teaching are affected by policies governing salaries, working conditions, and teacher education funding, student enrollments are less amenable to policy control. It is reasonable to ask whether these differences in operational teaching standards are mostly a function of demographic trends beyond the control of state policymakers. In examining state variations in hiring practices, however, it is clear that a number of high-growth states have enacted and maintained high standards for entry to teaching while many low-growth states have not. Policies appear to be at least as important as demographics in determining the qualifications of teachers hired and retained.

Because of these differences in licensing standards and enforcement, in 1994, more than 80% of high school teachers of academic courses in Wisconsin and Minnesota had fully met state certification requirements and had at least a college major in the field they teach. Four other states--Connecticut, Iowa, Montana, and North Dakota--reported similarly well-qualified teaching forces in that year. The comparable proportion of teachers with full state certification and a major in their field in Louisiana was only 64%. (An additional six states had fewer than two-thirds of their teachers similarly prepared.)

Interestingly, students in Minnesota and Wisconsin have typically scored at the top of the distribution on national assessments of reading and mathematics, along

with the four other states who share similarly well-qualified teachers. Together these states held six of the top ten spots in the national rankings in reading and mathematics in 1994 and 1996. Students in Louisiana have typically scored near the bottom of the NAEP distributions--no higher than 47th of 51 states in any of the assessments reported by 1996. The other six states with similar proportions of teachers holding a license and a major in their field all fall in the bottom quartile of states in the national rankings of average student achievement scores (Campbell et al., 1996; Darling-Hammond, 1997a, pp. 13, 26; Reese et al., 1997). Some have quipped that state-level student achievement in the U.S. can be best predicted by proximity to Canada--which in turn may be a proxy for variations among states in factors ranging from demographics (e.g., student poverty, parent education, and race) to political culture and spending on education. The distributions of scores described above could indeed partly support the "Canada hypothesis," which I test below.

States also differ greatly in the levels of funding they allocate to preservice and inservice teacher education, in the standards they apply to teacher education institutions and to schools, in the types and extent of professional learning opportunities and the incentives for professional study they make available to educators, and the extent to which they require or fund induction supports for beginning teachers. To illustrate these differences, in 1997 only three states required professional accreditation for schools of education and only nine funded induction programs that provided a structured program of mentoring for beginning teachers, including trained, state-funded mentors. Student teaching requirements ranged from 5 weeks in Massachusetts to 18 weeks in Wisconsin. As of 1994, the proportions of academic high school teachers teaching with both a license and a major in their field ranged from a low of 52% to a high of 85% across states. The proportions of mathematics teachers teaching with less than a minor in the field ranged from a low of 9% to a high of 56% (Darling-Hammond, 1997a, Appendices A and B). This means that a student in one state might have only one chance in ten of being taught by an out-of-field teacher, while a student in another state might have more than a 50% chance of being taught a subject by a teacher who is not adequately prepared in that subject.

In every category of possible investment in teachers' knowledge and in every area in which standards for teaching are set (e.g., licensing, accreditation, advanced certification, on- the-job evaluation), there are substantial differences in the policies and practices employed by states. States with some of the highest, most consistently enforced standards for teachers have tended to cluster in the upper Midwest (Minnesota, Wisconsin, Iowa, Nebraska, North Dakota, Missouri, Montana, Kansas). States with the lowest and least well-enforced standards have tended to include many in the southeast (Louisiana, Mississippi, Georgia, South Carolina) and in remote locations (Alaska, Hawaii). Some states have developed relatively ambitious standards for teaching but do not enforce them for large numbers of candidates (California, New York). Others have made major investments in preservice and inservice teacher development in recent years that have affected a substantial share of the teaching force (e.g., Connecticut, Kentucky, North Carolina, West Virginia). The possible outcomes of these cross-state differences are discussed below.

Trends in Student Achievement: Policy Hypotheses

In their book, *The Manufactured Crisis*, Berliner and Biddle (1995) noted that while U.S. secondary school students tend to score below the median in international assessments of mathematics and science, students in some states score as high as those in the top-ranked countries in the world while students in others score among the bottom-ranked. U.S. students also perform relatively better in some fields than others. For example, U.S. students have compared favorably with students in other countries in reading and at about the median in general science. However, in mathematics and physical science, U.S. students do much more poorly: In the most recent international assessments, 8th graders ranked 18th out of 25 countries that met the TIMSS guidelines in mathematics and 17th out of 25 countries in physics. Twelfth graders did even more poorly (Darling-Hammond, 1997a, pp. 28-29).

Although it may be purely coincidental, these differences in rankings are similar to the differences in teacher qualifications across these fields. Since the early 1980s, the U.S. has made major investments in teacher preparation in the area of reading. Not only are almost all elementary school teachers fully certified (more than 95%), an increasing number have been prepared in programs that have a strong emphasis on training to teach reading; there has also been a large increase in the number of reading specialists throughout the 1980s. In general science and biology, where U.S. middle and high school students scored at about the median on the most recent international assessments, there are relatively few uncertified or out-of-field secondary teachers (about 18% of the total). By contrast, in mathematics and physical science, where U.S. students fall well below the international norms, teacher qualifications are much weaker. In addition to the fact that most U.S. elementary teachers have had little background in mathematics, about 30% of U.S. mathematics teachers and 50% of physical science teachers at the high school level have been teaching with less than a minor in the field, many of them uncertified (Darling-Hammond, 1997, p.28 and Appendix Table 3). While these are only casual observations, other evidence point in similar directions.

Long-term Achievement Trends by State

Not only do U.S. students appear to perform least well in the fields in which U.S. teachers are least well prepared, the states that repeatedly lead the nation in student achievement in mathematics and reading have among the most highly qualified teachers in the country and have made longstanding investments in the quality of teaching (see Figures 1-3). The three long-time leaders--Minnesota, North Dakota, and Iowa--have all had a long history of professional teacher policy and are among the 12 states that have state professional standards boards which have enacted high standards for persons entering the teaching profession. They are recently joined at the top of the achievement distribution by Wisconsin, Maine, and Montana, states that have also enacted rigorous standards for teaching and that are among the few which rarely hire unqualified teachers on substandard licenses. Iowa, Minnesota, Montana, North Dakota, and Wisconsin have among the lowest rates of out-of-field teaching in the country and among the highest proportions of teachers holding both certification and a major in the field they teach. (Note 2) Maine joined these states in requiring certification plus a disciplinary major when it revised its licensing standards in 1988.

These states have also been leaders in redefining teacher education and licensing. Minnesota was the first state to develop performance-based standards for licensing teachers and approving schools of education during the mid-1980s and has developed a beginning teacher mentoring program in the years since (for details, see Darling-Hammond, Wise, & Klein, 1995). Wisconsin was one of the first states to require high school teachers to earn a major in their subject area in addition to completing extensive coursework in a teacher preparation program. Thus, teacher education in Wisconsin is typically a four-and-a- half to five year process. Maine, Wisconsin, Iowa, and Minnesota have all incorporated the rigorous new standards developed by the Interstate New Teacher Assessment and Support Consortium (INTASC) (Note 3) into their licensing standards and have encouraged universities to pilot performance-based assessments of teaching using these standards.

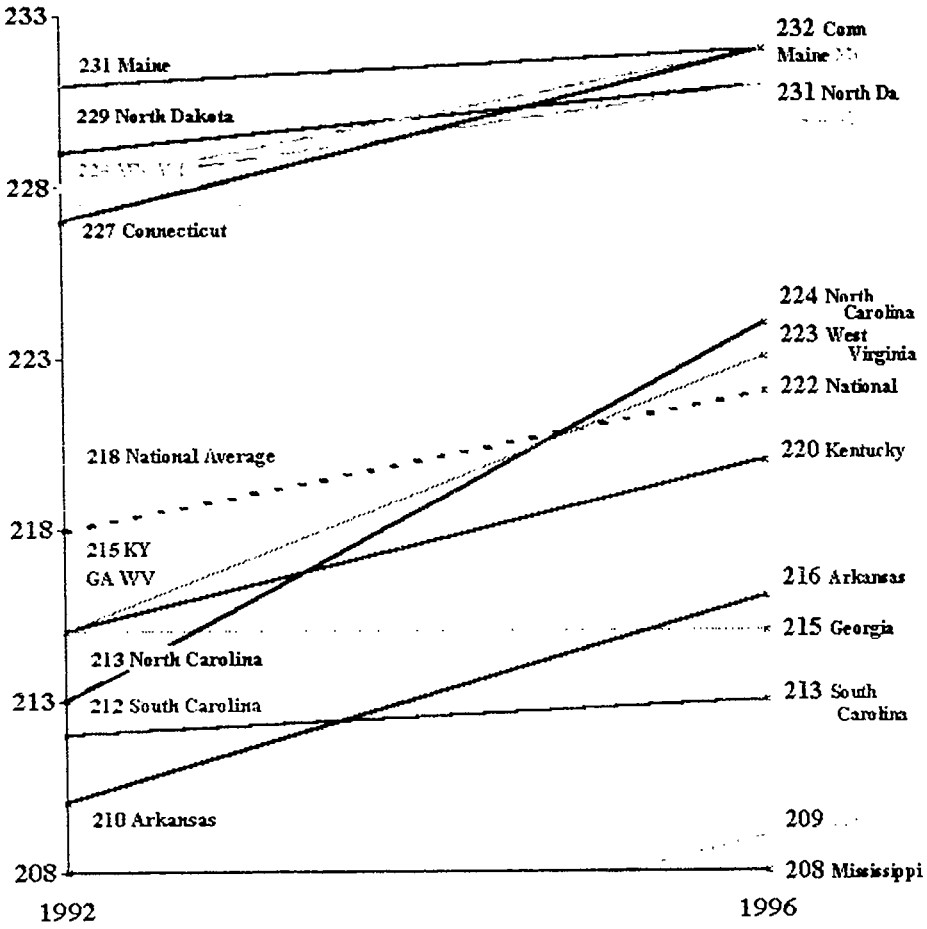


Figure 1. State Trends in Mathematics Achievement, Grade 4 (NAEP scores, 1992-1996)
Source: National Center for Education Statistics, *NAEP 1996 Mathematics Report Card for the Nation and the States*, Table 2.2, p. 28.

BEST COPY AVAILABLE

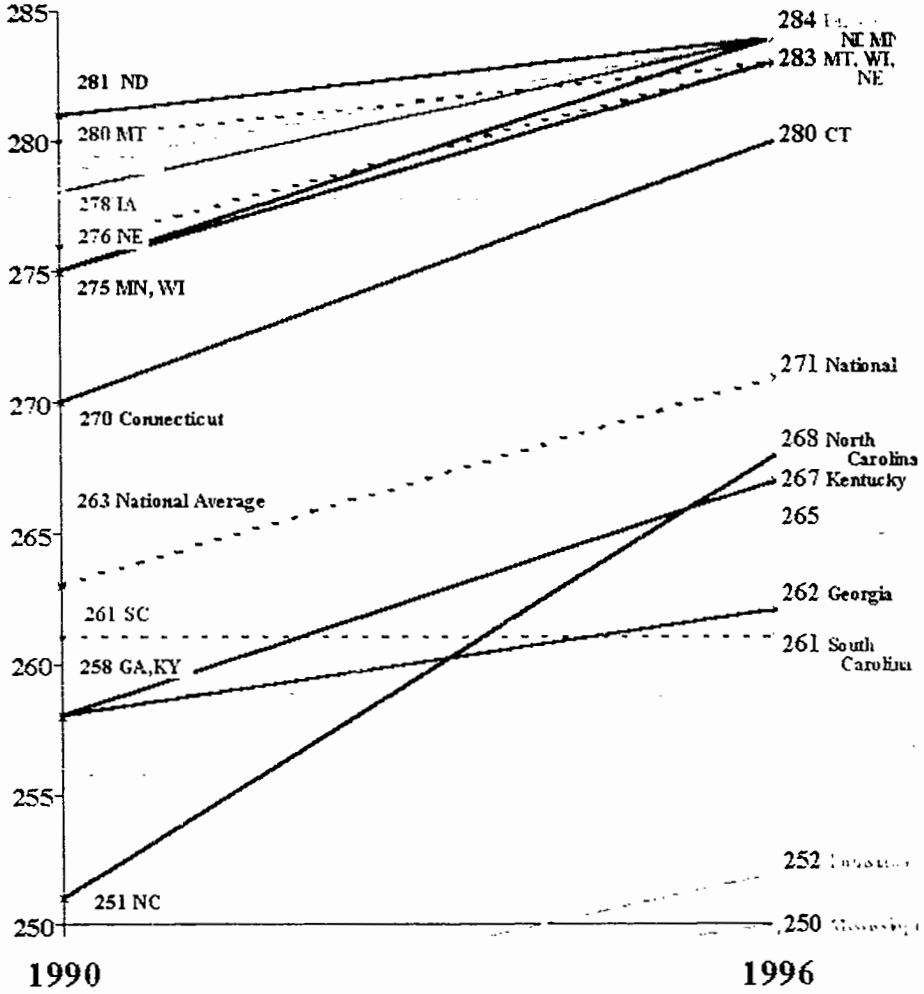
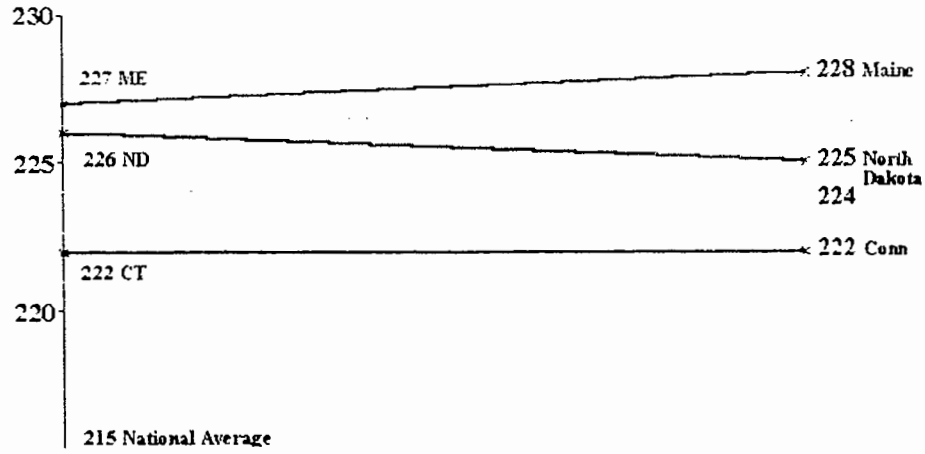


Figure 2.State Trends in Mathematics Achievement, Grade 4 (NAEP scores, 1990-1996)

Source: National Center for Education Statistics, *NAEP 1996 Mathematics Report Card for the Nation and the States*, Table 2.3, p. 30. (Note: Maine did not participate in 1990. Score is for 1992 assessment.)



BEST COPY AVAILABLE

Figure 3. State Trends in Reading Achievement, Grade 4 (NAEP scores, 1992-1994)
Source: National Center for Education Statistics, *NAEP Reading Report Card for the Nation and the States*, Table 2.3, p. 25.

One can still wonder whether policies are the source of these states' strong student outcomes or whether the "Canada effect" (general education spending combined with low rates of student poverty) is responsible. Among these six states, four spent below the per pupil national average in current expenditures in 1995, and the other two spent just above the average. All, however, spent a larger percentage of their expenditures on instruction than the national average. While these states did have a lower proportion of low-income students than the national average, none fell near the tail of the distribution. There were at least twelve states with lower proportions of low-income students who scored less well on the NAEP than any of these states. However, the relative contribution of student population characteristics and school inputs is an important one to pursue further. That question is raised again below.

State Achievement Gains

Another important question is whether investments in teaching could raise achievement in states that do not have a long history of this sort. Over the last decade of reform, a few states undertook major initiatives aimed at improving the quality of teaching. From a survey of state policies, we identified five states that enacted unusually comprehensive reforms of teaching during the late 1980s and 1990s: Connecticut and North Carolina enacted the most ambitious teacher legislation of any states nationally, followed by Arkansas, Kentucky, and West Virginia, which also initiated multi-faceted reforms of teacher preparation, licensing, professional development, and compensation, accompanied by substantial investments in teacher learning.

Of the 50 states, North Carolina and Connecticut undertook the most substantial and systemic investments in teaching during the mid-1980s. Both of these states, which share relatively large high-poverty student populations, coupled major statewide increases in teacher salaries and improvements in teacher salary equity with intensive recruitment efforts and initiatives to improve preservice teacher education, licensing, beginning teacher mentoring, and ongoing professional development. Since then, North Carolina has posted the largest student achievement gains in mathematics and reading of any state in the nation, now scoring well above the national average in 4th grade reading and mathematics, although it entered the 1990s near the bottom of the state rankings. Connecticut has also posted significant gains, becoming one of the top scoring states in the nation in mathematics and reading (ranked first at the 4th grade level in mathematics and reading and in the top five at the 8th grade level), despite an increase in the proportion of low-income and

limited English proficient students during that time.

North Carolina's reforms, launched with omnibus legislation in 1983, did many things simultaneously: (a) boosted salaries in the mid-1980s and again in the 1990s, (b) created a career development program that rewarded teachers for greater education and for achieving National Board Certification, (c) launched an aggressive fellowship program to recruit hundreds of able high school students into teacher preparation each year by entirely subsidizing their college education, (d) required schools of education to become professionally accredited by the National Council for the Accreditation of Teacher Education (NCATE), (e) increased licensing requirements for teachers and principals, (f) invested in improvements in teacher education curriculum, (g) created professional development academies and a North Carolina Center for the Advancement of Teaching, (h) developed teacher development networks like the National Writing Project and an analogous set of professional development initiatives in mathematics, (i) launched a beginning teacher mentoring program, and (j) introduced the most wide-ranging set of incentives in the nation for teachers to pursue National Board certification. North Carolina now boasts more Board-certified teachers than any other state. The state was recognized in the recent National Education Goals Panel report (NEGP, 1998) for having made among the greatest gains in teacher mentoring of beginning teachers as well as the greatest achievement gains for students.

These extensive investments in teaching occurred alongside sizable investments in early childhood education and general K-12 spending increases which lowered pupil/teacher ratios slightly. In the early 1990s, new curriculum standards were introduced and accompanied by an extensive program of professional development for teachers statewide. In 1993, the state enacted an assessment system linked to the curriculum standards and substantially aligned to the NAEP tests. This assessment program, which was implemented in 1994-95, occurred too late to account for most of the gains in achievement. Its effects would require several years to appear, but it may have had some modest influence on the gains after 1994.

A recent analysis of student achievement gains on the National Assessment of Educational Progress (Grissmer & Flanagan, 1998) attributed much of the NAEP score increase in North Carolina between 1990 and 1996 to the test-based accountability system. However, the new standards and assessments were not on-line until 1995, and the rewards and sanctions component of the accountability system was not enacted until 1997, so it was clearly not a factor in these trends. Grissmer and Flanagan also note the state's large-scale investments during the 1980s in early childhood education, reduced class sizes, teacher salary increases, teacher education upgrades, and extensive professional development. All of these factors could have influenced the achievement gains observed during this time period.

North Carolina's 1997 Educational Excellence Act furthered efforts to upgrade the quality of teacher preparation and teaching quality, pouring hundreds of millions of dollars into a new set of reforms. The Act created a professional standards board for teaching and required that all colleges of education create professional development school partnerships to provide the sites for year-long student teaching practicums. It also funded a more intensive beginning teacher mentoring program, further upgraded licensing standards, created pay incentives for teachers who pursue master's degrees and National Board certification, and authorized funds to raise teacher salaries to the national average. It will be useful to watch future trends in the state.

Connecticut's strategies were similar. The state's 1986 Educational Enhancement Act spent over $300 million to boost minimum beginning teacher salaries in an equalizing fashion that made it possible for low-wealth districts to compete in the market for qualified teachers. At the same time, the state raised licensing standards by requiring a major in the discipline to be taught plus extensive knowledge of teaching and learning as part of preparation; instituted performance-based examinations in subject matter and knowledge of teaching as a basis for receiving a license; created a state-funded mentoring program which supported trained mentors for beginning teachers in their first year on the job; and created a sophisticated assessment program using state-trained assessors to determine which first-year teachers could continue in teaching. An analysis of the

outcomes of this initiative found that it eliminated teacher shortages and emergency hiring, even in the cities, and created surpluses of teachers within three years of its passage (Connecticut State Department of Education, 1991).

Connecticut also required teachers to earn a master's degree in education for a continuing license and supported new, content- based professional development strategies in universities and school districts. In a National Education Goals Panel (1998) report highlighting Connecticut's strong performance and large gains in mathematics, state officials pointed to the salary increases and teacher education investments as central to their progress. These investments include an intensive professional development program in mathematics, science, and technology which, since 1983, has offered 4-week institutes with follow-up support to elementary, middle, and high school teachers.

The state has more recently invested in new curriculum frameworks and a statewide assessment system for students using extended performance tasks and constructed response items intended to measure higher order thinking and performance skills. Launched in 1995, this system, which is tied to statewide reporting of scores and substantial new professional development, may support future gains in student achievement. In addition, the state has further extended its performance-based teacher licensing system to incorporate the new INTASC standards and to develop portfolio assessments modeled on those of the National Board for Professional Teaching Standards (NBPTS). The new teacher assessments, which are tightly linked to the student standards, require beginning teachers to demonstrate that they can implement content-based teaching standards within their subject matter field and can analyze student work and learning. Finally, as part of ongoing teacher education reforms, the state agency is supporting the creation of professional development schools linked to local universities as sites for clinical training of entering teachers.

The Connecticut and North Carolina reforms both featured substantial investments in pre-service and in-service education for teachers linked to standards that incorporate much of the current knowledge base about teaching and learning (those of NBPTS, INTASC, and/or NCATE). While the reforms also included salary increases, the dollars were linked to improved quality via heightened licensing standards. Both states sought to increase not only the quality of preparation for teachers, but also the consistency with which they enforced their standards, sharply reducing the hiring of unlicensed and under-prepared staff.

Kentucky also realized substantial achievement gains during the 1990s, after undertaking perhaps the most extensive systemic education reforms of any state in the 1980s. These included major equalization of school funding along with large increases in teacher salaries and overall spending; changes in school organization, including multi-age primary grade classrooms; investments in early childhood education; the introduction of standards and curriculum frameworks, along with portfolios and performance assessments. Changes in teacher education and licensing accompanied these reforms, including the adoption of the INTASC licensing standards developed by a consortium of more than 30 states, the introduction of new licensing tests and teacher education requirements, incentives for colleges of education to meet national professional accreditation standards; and massive investments in professional development.

All of these efforts undoubtedly combined to produce the steep gains in achievement experienced in Kentucky. By 1994, data from the Schools and Staffing Surveys showed that Kentucky teachers were much better prepared in terms of their content and teaching coursework background than in 1988 and had experienced more extensive professional development than teachers in any other state (Darling-Hammond, 1997a). A recent survey of Kentucky teachers also found that more than 80% of beginners who graduated from Kentucky colleges of education felt well-prepared for virtually all aspects of their jobs (Kentucky Institute for Educational Research, 1997), in contrast to reports about teacher education from previous studies elsewhere. Although somewhat less ambitious in their reforms, Arkansas and West Virginia also raised teacher salaries and licensing requirements and required national accreditation of education schools during the late 1980s or early 1990s, while investing in more professional development for in-service teachers. These states also realized steeper gains in student achievement than the

national average.

In a recent report, Grissmer and Flanagan (1998) focused on Texas and North Carolina for their large gains in average student achievement. They attributed Texas' gains primarily to the state's accountability system, although they also mention its shifts of resources to more disadvantaged students through school finance equalization, class size reductions, and the creation of full day kindergarten. The school funding investments that occurred in the 1980s and were continued into the following decade may indeed have made some difference in Texas students' achievement in the 1990s. However, the state's new assessment and accountability system was not initiated until 1994 and not fully implemented until 1995-96, so it could not have accounted for gains between 1990 and 1996.

Texas was not included in the above analysis of state test score gains because it was not one of the states that made large comprehensive investments in teaching during the 1980s. (Texas did make some noteworthy investments in teacher salaries and professional development in the 1990s.) In addition, however, there are questions about the stability of scores in Texas and the extent to which the posted gains are real. First, Texas included fewer than 45% of its students with disabilities in the testing pool, a much smaller share than most states (NCES, 1997, Table D3). Excessive exclusions of low-scoring students from the testing pool can cause gain scores to appear much larger than they would otherwise be. In addition, recent studies in Texas have raised concerns that much of the ostensible gain registered by African American and Latino students has been a function of grade retentions and dropouts or pushouts, which have increased substantially in recent years. These practices also make average test scores look higher by eliminating lower scoring students from the testing pool (Haney, 1999; Kurtz, 1999; Mexican American Legal Defense and Education Fund, 1999). Assuming that some of the gains in Texas are not spurious, however, it is worth noting that, in addition to the equalization of funding and investments in kindergarten and reduced class sizes, Texas was among the few states recognized by the National Education Goals Panel (1998) for large gains since the early 1990s in the proportion of beginning teachers receiving mentoring from expert veterans. Texas has also had a growing number of 5-year teacher education programs in response to an earlier reform eliminating teacher education majors at the undergraduate level.

State reform strategies during the 1980s that did not include substantial efforts to improve the nature and quality of classroom work have shown little success in raising student achievement, especially if the reforms relied primarily on student testing rather than investments in teaching. For example, the first two states to organize their reforms around new student testing systems were Georgia, with its Quality Basic Education Act (QBE) of 1985, and South Carolina, with its Education Improvement Act of 1984. These states developed extensive testing systems coupled with rewards and sanctions for students, teachers, and schools. Although both states also mandated tests for teachers, they did not link these assessments to emerging knowledge about teaching or to new learning standards, nor did they invest in improving schools of education or ongoing professional development: Few districts in either state require teachers to hold a degree in the field to be taught and full state certification as a condition of hiring. As Figures 1-3 show, student achievement in mathematics has been flat in these states while achievement in reading has declined. Since 1996, Georgia has launched an ambitious series of reforms through its P-16 Council to upgrade the quality of teacher preparation and professional development and to raise licensing standards, as well as to recruit high ability students to teaching. Future analyses might examine whether these moves have made a difference.

There are competing hypotheses that could explain these across-state differences in achievement trajectories. One could speculate that student testing and curriculum changes are not in themselves powerful enough reforms to overcome the depressing effects on teaching quality of low standards for teacher education, licensing, and hiring, and the resulting large numbers of under-prepared teachers. On the other hand, one can argue that variables like student poverty and language background, rather than conditions that might influence the quality of teaching, are the determining factors in student achievement and that the critical differences between high- and low-achieving states are differences in their student populations.

It is interesting to compare the student achievement levels and trajectories for

some of these states in comparison to geographically proximate states with similar student populations that have taken very different approaches to teaching policy. While the comparisons in Table 1 are only suggestive, they demonstrate that student achievement cannot be assumed to be only or primarily a function of demographics. Although the states that have aggressively pursued investments in teacher knowledge and skills have equal or higher levels of student poverty than nearby states that pursued other, distinctively different reform strategies, their students now achieve at higher levels. Even though all of these states increased teacher salaries during the 1990s, those that insisted on higher standards for teacher education and licensing realized gains that were not realized by states that maintained or lowered their standards for entering teaching.

Table 1
State teacher salaries, student poverty, and student achievement
NAEP 4th grade mathematics scores, 1996

	NAEP Score, 1996	Gain from 1992	% of students in poverty	Teacher Salaries Minimum	Maximum
Connecticut	232	+ 5	18.6	$28,195	$56,189
New Jersey	227	+ 0	14.6	$28,424	$58,208
North Carolina	224	+11	18.4	$20,077	$38,733
Georgia	215	+ 0	18.5	$20,065	$42,134
West Virginia	223	+ 8	22.0	$21,466	$36,378
Virginia	223	+ 2	12.6	$23,098	$38,328

Data on student achievement and poverty status from NAEP 1996 Mathematics Report Card for the States, Washington, D.C.: U.S. Department of Education, 1997, pp. 28, 139. Data on teachers' salaries from NCES, America's Teachers: Profile of a Profession, 1993-94, Washington, D.C.: U.S. Department of Education, 1996, Table A6.2.

For example, with their industrialized urban areas and affluent suburbs, Connecticut and New Jersey are demographically and economically similar states, although Connecticut has noticeably higher rates of student poverty. Despite a more affluent student population, New Jersey's students did less well than those in Connecticut on the NAEP 4th grade mathematics assessments in 1996, and, in contrast to Connecticut's students, they have not improved in recent years. Whereas Connecticut raised teachers' salaries and equalized districts' abilities to pay for qualified teachers, New Jersey decreased its requirements for teacher preparation and licensing at the end of the 1980s, reducing the amount of education coursework for entry into teaching to a maximum of 18 undergraduate credit hours and encouraging the more extensive hiring of alternative certification candidates prepared in a short summer program. These less-prepared teachers are primarily hired in low-wealth city school districts that have had radically lower revenues and salary schedules than other parts of the state.

While New Jersey's average teachers' salaries are the highest in the country, even higher than Connecticut's, New Jersey's salary increases were not tied to improvements in the qualifications of teachers or to equalization in districts' ability to pay for qualified teachers. New Jersey also lacks the rigorous licensing examinations, requirements for a major in the field and a masters in education, and state-funded mentoring for beginning teachers that Connecticut enacted in 1986. Compared to Connecticut, New Jersey has much lower rates of beginning teachers receiving mentoring and induction, much lower proportions of districts insisting on rigorous hiring standards, much lower proportions of teachers receiving professional development, much lower rates of teachers holding full certification plus a major in the field, and much higher rates of out-of-field teaching in every subject matter field

(Appendix B, Tables 1-5, Darling-Hammond, 1997a).

In the same fashion, North Carolina's students now perform substantially better on the NAEP assessments than those in demographically similar Georgia, which North Carolina lagged behind in 1990. Although the states raised salaries during the 1980s and early '90s to comparable levels, Georgia did not raise standards for teacher preparation and licensing or invest heavily in teacher development at the same time. While North Carolina increased both the education and subject matter requirements for teacher preparation, introduced rigorous teacher examinations for licensing, and required national accreditation for all of its education schools during the 1980s, Georgia did little to increase expectations for either preservice or inservice preparation during those years. In addition to having had more extensive training to meet certification standards, North Carolina teachers are much more likely than their peers in Georgia to have had mentoring as beginning teachers and professional development opportunities as veterans.

And very poor West Virginia now ranks as well in elementary mathematics as its neighbor Virginia, whose students are much more affluent. Virginia, with its higher cost of living, pays its teachers more. However, West Virginia's efforts to raise salaries were accompanied by efforts to improve teacher education and licensing standards. All of West Virginia's teacher education programs must now meet national accreditation standards--a much higher set of requirements than those in Virginia, which lowered standards for education programs and licensing during the 1980s to among the lowest in the country. Like New Jersey, Virginia reduced the requirements for coursework on teaching and learning in undergraduate programs, while West Virginia raised its standards. West Virginia introduced an ambitious program of professional development even before it launched its new curriculum frameworks in the mid-1990s, and enacted a mentoring program for beginning teachers. Despite its relative wealth, Virginia hires many more unlicensed new teachers than West Virginia and its districts are less likely to insist on rigorous hiring standards.

These kinds of contrasts can be seen in many comparisons of geographically proximate, demographically similar states that have taken different approaches to the issue of teacher investments over the last decade. Policies that jointly raise salaries and standards may offer particularly high leverage on teaching quality. It is interesting to note that, like states that introduced testing without making investments in teaching, those that have raised salaries alone, without raising standards for preparation and licensing or investing in professional development, seem not to have realized the benefits of improved student outcomes. While interesting, these observations of individual state cases could be idiosyncratic. An important question is whether similar patterns exist when viewed from a national perspective.

A National View of Teacher Qualifications and Student Achievement

To examine further the relative contributions of teaching policies and student characteristics to student achievement, this analysis uses data on public school teacher qualifications and other school inputs available from the 1993-94 Schools and Staffing Surveys (SASS) and data on student achievement and student characteristics from the 1990, 1992, 1994, and 1996 assessments in reading and mathematics administered by the National Assessment of Educational Progress. These data are the basis for regression analyses of school resource variables on student achievement scores to examine whether teacher quality indicators, as well as other school inputs, are related to student achievement at the state level, after controlling for such student characteristics as poverty and language background.

The Database The 1993-94 SASS database includes linked surveys of 65,000 teachers (52,000 public and 13,000 private); 13,000 school principals (9,500 public and 3,500 private); and 5,600 school districts. SASS is designed to provide reliable estimates of the characteristics of schools and educators at the national and state levels. It also includes information from individual teachers, school principals, and districts about salaries and compensation policies, induction policies, school climate and context variables (e.g., time to work with other teachers, teacher involvement in decision-making), professional development support, teachers' views of teaching,

and their plans to remain in the profession. These analyses use the following data derived from the public school teachers' questionnaire: data on teachers' qualifications (teachers' degrees, majors, certification status), teaching assignments, and average class size. Also included in the analysis are data from the public school district questionnaire on district hiring policies (whether districts require, as a condition of hiring, full certification, graduation from an approved teacher education program, or a college major or minor in the field to be taught) and salary schedules (minimum and maximum salaries) as reported by district officials. Salary schedule data are more appropriate for gauging attractions to teaching than average salary data, which do not control for differential levels of experience and education across states. All of the SASS data were aggregated to the state level.

Teacher quality variables constructed from the SASS data include the proportion of "well-qualified teachers," defined as the proportion holding state certification and the equivalent of a major (either an undergraduate major or masters degree) in the field taught. For elementary teachers, the equivalent of a major is an elementary education degree for generalists who teach multiple subjects to the same group of students or a degree in the field taught for specialists (e.g. reading, mathematics or mathematics education, special education). The proportion of teachers who are "fully certified" includes teachers with standard or regular certification and new teachers on probationary certificates who have completed all requirements for a license except for the completion of the probationary period (usually 2 or 3 years of beginning teaching). The proportion of teachers who are "less than fully certified" includes teachers with no certificate and those with provisional, temporary, or emergency certification.

Additional data on each state, including policies regarding teacher education and licensing (number of weeks of student teaching required, presence of a professional standards board, percentage of teacher education institutions that are NCATE accredited), were collected directly from states and professional associations (see Darling-Hammond, 1997a, Appendix A). State school spending data (current per pupil expenditures) are from the Common Core of Data (NCES, 1995).

Data from the National Assessment of Educational Progress (NAEP) include state average achievement scores for students in mathematics at the 4th grade level in 1990 and 1996 and at the 8th grade level in 1992 and 1996, as well as data on state average achievement scores for students in reading at the 4th grade level in 1992 and 1994 (Campbell, Donahue, Reese, & Phillips, 1996) and student poverty rates (Reese, Miller, Mazzeo, & Dossey, 1997).

Limitations There are a number of limitations that pertain to the data set and the analyses. First, the NAEP data derive from tests that do not measure all of the valued outcomes of schooling held by parents, teachers, and schools. They cannot represent everything that schools do or should do. In addition, state scores and changes in average scores on these measures are sensitive to differences in the population of students taking the tests, including decisions about which students will be excluded from testing and differences across states in the extent to which populations are represented in school (as a function of school- age population characteristics, dropout rates and patterns, and other variables).

Finally, the level of aggregation necessarily influences the interpretations of results. Aggregating data to the state level produces different results than one would find if one looked at similar kinds of data at the individual student, teacher, school, or district level. The direction of the differences cannot be predicted with certainty (Ferguson and Ladd, 1996). While, on one hand, the possibility of greater variability or noise exists in disaggregated analyses, it is possible that omitted variables may bias the coefficients of school input variables upward when the data are aggregated to the district or state level (Hanushek, Rivkin, and Taylor, 1995). Although the results of more and less aggregated specifications can be consistent (for example, Ferguson and Ladd's (1996) Alabama analysis found comparable influences of teacher quality and class sizes on student achievement when measured at the student and the district levels), this may not always occur. In particular, the size of relationships found between variables measured at the state level cannot be assumed to represent the effect sizes one would find in a classroom level analysis. For the purposes of assessing broad policy influences at the state level, it is nonetheless

reasonable to examine state-level data as a gauge of major trends when other confirming and disconfirming evidence is available to supplement the analysis.

The Findings All analyses include public schools and teachers only. Although the sample includes all states participating in state NAEP and thus is not a representative sample from which one would draw population inferences, I report p-values as an aid to readers who wish to use them to interpret the relative sizes of relationships and the probabilities of a Type I error. Before constructing the multivariate analyses, initial bivariate correlations of school resource variables and student demographic variables with state average student test scores were conducted to examine the relationships among variables and to select variables for inclusion in the multivariate equations. These analyses confirmed several findings reported elsewhere:

- *Student characteristics such as poverty, non-English language status, and minority status are negatively correlated with student outcomes, and usually significantly so.* These student characteristics are also significantly and negatively correlated with the qualifications of teachers; that is the less socially advantaged the students, the less likely teachers are to hold full certification and a degree in their field and the more likely they are to have entered teaching without certification.
- *Student characteristics are generally not significantly correlated with state per-pupil spending or with teachers' salary schedules*, with the exception that salary schedules are higher in states with larger percentages of minority and LEP (limited English proficient) students. Salary levels show an insignificant, negative relationship with levels of student poverty.
- *Teacher quality characteristics such as certification status and degree in the field to be taught are very significantly and positively correlated with student outcomes.* Characteristics such as education level (percentage of teachers with master's degrees) show positive but less strong relationships with education outcomes.
- *Per pupil spending (measured as current expenditures) shows a significant positive relationship with student outcomes in 4th grade reading in both years, but no relationship with student outcomes in mathematics.* This may be because the spending measure incorporates resources spent not only on teacher salaries and professional development but also on class sizes and other resources that may especially support students in the early grades as they are learning to read. Although salaries and spending are strongly related to one another ($p < .01$), teacher salary levels, unadjusted for cost of living differences, are not correlated with student outcomes when aggregated to the state level.
- *Other school resources, such as pupil-teacher ratios, class sizes, and the proportion of all school staff who are teachers, show very weak and rarely significant relationships to student achievement when they are aggregated to the state level.*

Partial correlations confirm a strong, significant relationship of teacher quality variables to student achievement even after controlling for student poverty and for student language background (LEP status) in (see Table 2 and Figure 4). The most consistent highly significant predictor of student achievement in reading and mathematics in each year tested is the proportion of well-qualified teachers in a state: those with full certification and a major in the field they teach (r between .61 and .80, $p<.001$). The strongest, consistently negative predictors of student achievement, also significant in almost all cases, are the proportions of new teachers who are uncertified (r between -.40 and -.63, $p<.05$) and the proportions of teachers who hold less than a minor in the field they teach (r between -.33 and -.56, $p<.05$). General spending and salary variables, along with class sizes, are not significantly related to achievement once student characteristics are taken into account. It should be noted, however, that this analysis did not take into account cost-of-living differentials that may affect both salaries and spending levels; controlling for such differentials could produce a different set of results with respect to these variables.

BEST COPY AVAILABLE

Table 2
Partial Correlations (controlling for student poverty) between Selected Teacher Quality Variables and Student Achievement on the National Assessment of Educational Progress

	Grade 4 Math, 1992	Grade 4 Math, 1996	Grade 8 Math, 1990	Grade 8 Math, 1996	Grade 4 Reading, 1992	Grade 4 Reading, 1994
% of teachers well-qualified (with full certification and a major in their field)	.71***	.61***	.75***	.67***	.80***	.75***
% of teachers out of field (with less than a minor in the field they teach)	-.48**	-.44**	-.32	-.42**	-.56**	-.33*
% of all teachers fully certified	.36*	.20	.38*	.28	.57***	.41*
% of all teachers less than fully certified	-.36*	-.23	-.33*	-.28	-.55***	-.50*
% of new entrants to teaching who are uncertified (excluding transfers)	-.51**	-.39*	43**	-.38*	-.44**	-.47**
% of all newly hired teachers uncertified	-.40**	-.41**	-.53***	-.49**	-.59***	-.63***
Per pupil spending	.32	.28	.19	.29	.24	.27
Pupil: teacher ratio	.03	.22	.09	.12	.08	.08
Class size	-.03	.21	-.04	-.00	.08	.13

*p< 10 **p<.05 ***p< 01

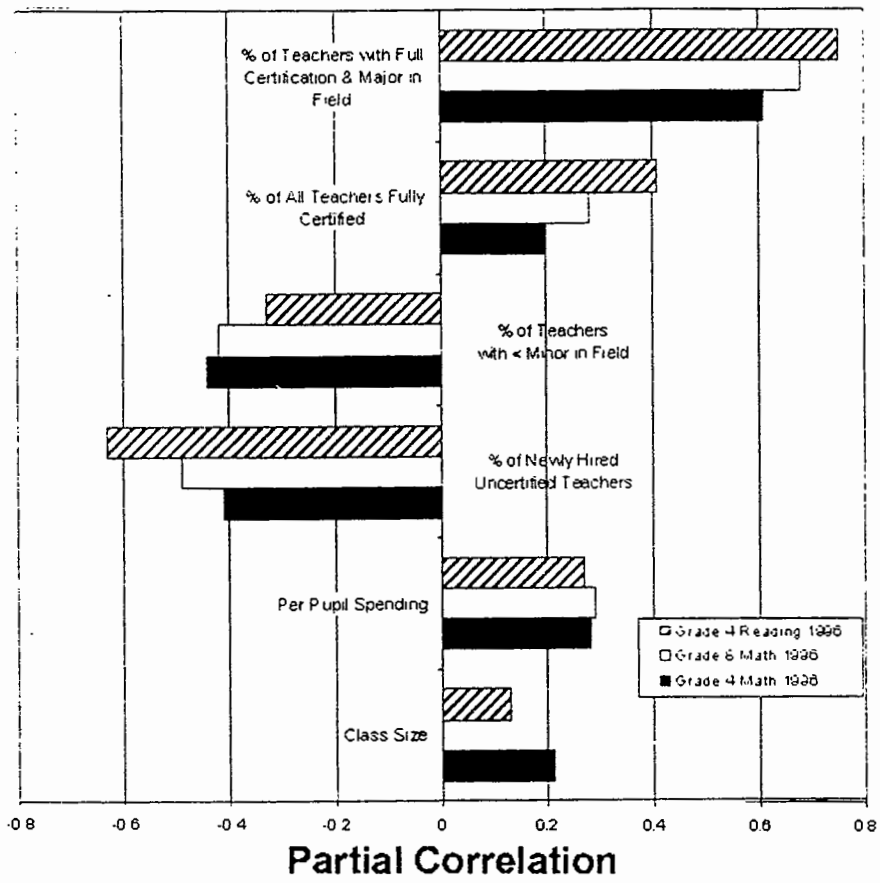


Figure 4. Partial Correlations (controlling for student poverty) between Selected Teacher Quality Variables and Student Achievement on the National Assessment of Educational Progress

Ordinary least squares regression analyses were performed to create the most parsimonious specification of a hyperplane of best fit with student achievement data. Because of the small sample size (n = 44 states participating in the state NAEP), the number of independent variables in each equation was minimized to preserve the necessary degrees of freedom (see Table 3). Variables were selected according to three criteria: to examine relationships often tested in other studies, to maximize explanatory power, and to avoid problems of multicollinearity. Teacher quality variables included the percentage of all teachers with full certification and a major in the field and the percentage of uncertified newly hired teachers, because these exhibit large influences on achievement, and the percentage of teachers with master's degrees, because this is a frequently examined teacher quality variable. Class size was also included because it is commonly found to influence achievement. Spending and salary variables were not included in the final estimations because they showed little relationship to student achievement in preliminary estimates. Because the percentage of minority students is highly correlated with both poverty rates (r=.55) and LEP status (r=.52), while poverty rates and LEP status are not as highly related to each other (r=.29), the equations were estimated with poverty rates and LEP status as key student characteristics to avoid multicollinearity.

Table 3
Influences of School Resources and Student Characteristics On State-Level NAEP Student Achievement Scores

BEST COPY AVAILABLE

Beta Coefficients (T values)	Math-4 1992	Math-4 1996	Math-8 1990	Math-8 1996	Reading-4 1992	Reading-4 1994
% Well-qualified Teachers (with full certification and a major in their field)	.857 (4.3)***	.818 (2.99)**	.869 (4.90)***	.79 (3.94)**	.824 (4.78)***	.636 (3.36)**
% with Masters Degrees	.075 (.59)	.159 (.91)	-.007 (-.06)	.157 (1.23)	.053 (.48)	.103 (.86)
% Unqualified Newly Hired Teachers (uncertified in their main assignment field)	.079 (.47)	.112 (.48)	-.058 (-.39)	-.034 (-.20)	-.092 (-.63)	-.199 (-1.2)
Class Size	-.077 (-.67)	.076 (.49)	-.081 (-.79)	-.032 (-.28)	-.111 (-1.13)	-.091 (-.83)
Poverty (% students with incomes below the poverty line	-.336 (-2.2)*	-.234 (-1.11)	-.211 (-1.5)	-.353 (-.2.3)*	-.080 (-.61)	-.166 (-1.14)
LEP (% students who are limited English proficient)	.276 (1.8)	.246 (1.2)	.286 (2.16)*	.391 (2.6)*	-.015 (-.11)	-.058 (-.41)
Multiple R	.91	.82	.9	.91	.93	.92
R-Square	.82	.67	.86	.82	.87	.84

*p<.05 **p<.01 ***p<.001

The equations explain between 67 and 87 percent of the total variance in student achievement, and the findings are robust across subjects and years. In all cases, the proportion of well- qualified teachers is by far the most important determinant of student achievement: it is highly significant in all equations for both subject areas in all years and at all grade levels. Other teacher quality variables contribute modestly to explaining student achievement. The proportion of teachers with master's degrees exerts a small, generally positive effect on achievement, while the proportion of uncertified new teachers exerts a small, generally negative effect. Together, these three teacher quality variables account for between 40 percent and 60 percent of the total variance in student achievement in reading and mathematics in each of the years and grade levels assessed, once student characteristics are taken into account.

Smaller class sizes are moderately associated with higher achievement in five of the six equations, with the largest effects visible in 4th grade reading. Student poverty rate exerts a negative influence on achievement, although it is not significant in four of the six equations. In mathematics, the proportion of LEP students exerts a positive effect on achievement after controlling for poverty status. In reading, LEP status exerts an insignificant negative effect on achievement when poverty is controlled.

Analysis of Policy Relationships

Clearly, in any analysis such as this, the variables that can be measured are only proxies for the actual conditions or traits that may matter to student learning. In this case, a large number of variables associated with teacher quality appear to bear a significant relationship to student achievement. These include various ways of measuring state certification status (the proportions of teachers with full certification, less than full certification, and no certification) and disciplinary preparation (e.g., a major or minor in the field to be taught). Given the differences in licensing standards and teacher education programs across states, these proxies are fairly crude ones; nonetheless, they seem to indicate that teachers' knowledge, skills, and preparation matter for student achievement. The findings are similar to those of several other studies described earlier (Ferguson, 1991; Ferguson and Ladd, 1996; Fetler, 1999; Fuller, 1999; Strauss and Sawyer, 1986) in finding much stronger influences on student achievement of variables measuring teacher knowledge and skills than of variables like teacher experience, class sizes, or pupil-teacher ratios, which are generally found to have noticeable but smaller effects on student achievement where data are aggregated to the school or district levels.

The strength of the "well-qualified teacher" variable may be partly due to the fact that it is a proxy for both strong disciplinary knowledge (a major in the field taught) and substantial knowledge of education (full certification). If the two kinds of knowledge are interdependent as suggested in much of the literature, it makes sense that this variable would be more powerful than either subject matter knowledge or teaching knowledge alone. It is also possible that this variable captures other features of the state policy environment including general investments in, and commitment to, education, as well as aspects of the regulatory system for education, such as the extent to which standards are rigorous and the extent to which they are enforced. Recall that some states require teachers to acquire a subject matter major as well as extensive education training in human development and learning and in the methods of teaching in their field, while other states require much less extensive preparation in the content area as well as teaching and learning. In addition, some states are vigilant in enforcing their certification standards while others are not.

Teaching Standards and Other Policy Strategies

Finally, there may be unmeasured correlations between the extent to which states enact and enforce high standards for teachers and the extent to which they have enacted other policies that are supportive of public schools. Although it does not appear that teaching standards are strongly related to investments regarding class sizes or to overall education spending, it is possible that there are other factors influencing student achievement which generally co-exist with teacher quality and which were unmeasured in these estimates. Since most of the states which ranked among the highest-scoring on the NAEP examinations are strong local control states that have traditionally not exerted much control over school decision making, there are relatively few policy areas in which they have been active. Perhaps the relative lack of policy intervention is itself a support for student learning, leaving educators free of regulations that might force greater attention to procedures than learning. Another possibility is the influence of these states' small school and district sizes, a factor that has been identified in much research as contributing to student learning (for reviews, see Green & Stevens, 1988; Howley, 1989). In another analysis, Feistritzer (1993) has pointed out that most of the top-scoring states on NAEP have very small average school sizes relative to national norms.

One area in which policies have not been positively correlated, however, is the extent to which states engaged in statewide student testing in the 1980s and the extent to which they enacted high standards for teachers. Among the 12 highest- scoring states in 8th grade mathematics in 1996 (10 of which had particularly high licensing standards in the form of subject matter and teaching coursework requirements), none had mandatory statewide testing programs in place during the 1980s or early 1990s. Only two of the top 12 states in 4th grade mathematics had statewide testing programs in place prior to 1995. By contrast, among the 12 lowest-scoring states (8 of which had particularly large rates of out-of-field and uncertified teachers), 10 had

BEST COPY AVAILABLE

extensive student testing programs in place prior to 1990, some of which were associated with highly specified state curricula and an extensive menu of rewards and sanctions.

There are several possible interpretations of the almost inverse relationship between statewide testing policies and both teaching standards and student performance: It may be that states with low student performance and less qualified teachers were more likely to seek education improvements through student testing strategies and curriculum controls. It may also be that states have tended toward different theories of reform, with some investing more in testing and others in teaching. It is possible that regional differences in education investments and centralization happen to be correlated with policies regarding both testing and teacher investments (with Southern states that tend to score lowest investing heavily in curriculum and testing controls, while Northeastern and North Central states invest more in teacher education and less in curriculum controls).

The lack of apparent relationship between testing programs and student achievement might be because, without other investments to improve teaching and schooling, tests alone do not transform learning. Another possibility is that the kinds of basic skills tests and curricula enacted in many states during the 1980s were at odds with the NAEP assessments which increasingly seek to measure higher-order skills and performance abilities. It may be worth noting that most of the high-scoring and fast-gaining states discussed earlier instituted curriculum and testing reforms in the mid-1990s that were linked to the national student standards that guide NAEP and were much more performance-oriented than the basic skills tests that predominated in state assessment systems of the 1980s. While there is little evidence yet of the effects of these assessment programs on student learning, policy analysts may want to watch to see whether the types of tests matter for broad student outcomes as well as whether and how the supports that do or do not accompany testing programs (professional development, funding equalization, investments in additional supports for students ranging from early childhood education to special services of various kinds) make a difference.

Policies that May Influence Teachers' Qualifications

Another set of questions has to do with whether there are particular policy strategies used by states or districts that are associated with the preparation and hiring of better qualified teachers. The SASS data set and additional data collected directly from states allowed us to examine several policies in this regard.

Teacher education accreditation National data from the National Association of State Directors of Teacher Education and Certification and from the National Council for the Accreditation of Teacher Education provided the percentage of teacher education institutions that were accredited by NCATE. NCATE-accreditation might lead to higher overall standards for teachers because NCATE standards revisions in 1988 and 1993 required higher admissions standards, evidence of greater subject matter preparation, and stronger rationales for the content of education coursework than those often emphasized by state approval systems.

Standard setting and enforcement mechanisms The state survey tracked the presence of a state professional standards board for teaching, analogous to the boards that govern other professions, which might enact and enforce higher standards. Since any policies for teacher education adopted by such a board would require several years to take broad effect, the enactment of a standards board prior to 1990 is the measure we used for examining influences on teacher qualifications in 1994.

District hiring standards SASS data provided the percentage of school districts in each state requiring each of the following as conditions for hiring: full state certification, graduation from an approved teacher education program, and a college major or minor in the field to be taught. There was wide variation across the states in the degree to which districts looked for evidence of these kinds of teacher qualifications as part of the hiring process.

Many more fine-grained variables, such as the content of licensing standards and the nature of teacher education programs, could not be tested with these data.

Nonetheless, the results suggest some interesting associations. As shown in Table 4, the strongest predictor of the percentage of well-qualified teachers (that is, teachers with both a major and full certification in their field) is the percentage of teacher education institutions in a state that meet national accreditation standards through NCATE ($p < .05$).

Table 4
Relationship Between Professional Accreditation And Teacher Qualifications

Variable (Beta coefficient)	% of well-qualified teachers	% of well-qualified English teachers	% of well-qualified Math teachers	% of math teachers out-of-field	% of English teachers out-of-field
% of colleges NCATE accredited	.42*	.49**	.36*	-.37*	-.37*

*p<.05 **p<.01

The proportion of NCATE-accredited institutions is also significantly and negatively correlated with the proportion of English and mathematics teachers who are "out-of-field" (i.e., have less than a minor in the field they teach). This may be because institutions that are NCATE-accredited must demonstrate that their students have the opportunity to acquire a base of content knowledge deemed acceptable by the subject matter associations that review applications as well as pedagogical knowledge in their field. Thus, these institutions may, as a group, have less variability than others in establishing reasonably high standards for disciplinary knowledge as well as knowledge of how to teach the discipline. It may also be that states in which professional accreditation is more widespread also happen to have other policies or practices in effect that support the preparation and hiring of well-qualified teachers.

As shown in Table 5, the extent to which districts maintain rigorous hiring standards (i.e., the percentage of districts requiring full certification, graduation from an approved teacher education program, and a college major or minor in the field to be taught) is a highly significant predictor ($p < .001$) of the proportions of teachers who are uncertified. It is also a strong predictor of the proportions of new and veteran teachers who are fully certified. Since teachers' certification status is also related to state demographics, these variables were regressed against hiring standards along with student poverty, percent minority, and percent LEP students. The relationship between hiring standards and teacher certification status continues to be highly significant after controlling for student poverty, race, and language status.

Table 5
Correlations between Teacher Qualifications and District Hiring Standards (Pearson r)

	District Hiring Standards (Percent of districts requiring full certification, graduation from an approved teacher education program, and a college major or minor in the field to be taught as a condition of hiring)
% of new teachers who are fully certified	.28**
% of all teachers who are fully certified	.33**
% of newly hired teachers who are uncertified	-.51***
% of all teachers who are uncertified	-.66***

*p< .05 **p< .01 ***p< .001

BEST COPY AVAILABLE

Table 6
Relationship between Teacher Qualifications and District Hiring Standards
(Controlling for Student Poverty, Minority Status, and Language Status)

Variable/ Beta Weight/ (t value)	% all teachers fully certified	% new teachers fully certified	% all teachers uncertified	% new teachers uncertified
District hiring standards****	.393 (2.51)*	.339 (2.16)*	-.636 (-4.73)***	-.502 (-3.19)**
Professional Standards Board			-.173 (-1.20)	-.080 (-.48)
% students in poverty	-.148 (-.64)	-.063 (-.27)	.172 (.94)	-.108 (-.51)
% students LEP	.226 (1.23)	.374 (2.02)	.105 (.63)	.045 (.23)
% students minority	.125 (.58)	-.112 (-.43)	-.352 (-1.66)	-.105 (-.42)

*p<.05 **p<.01 ***p<.001

****Percent of districts requiring, as a condition of hiring, full certification, graduation from an approved teacher education program, and a college major or minor in the field to be taught

This suggests that enforcing standards is both a state and local job. In a quasi-profession like teaching, there is a complex interplay between the standards adopted by states and the ways in which local schools and districts manage their hiring processes, sometimes in accord with and sometimes in violation of state standards. A minority of states enforce their teacher licensing standards in the inviolable fashion with which standards for doctors, lawyers, architects, and other professionals are enforced. These other professions use professional standards boards established by each state as standard-setting and enforcement bodies. Depending on the degree of authority and autonomy used as defining characteristics, 12 to 18 states have established such boards for teaching.

As shown in Table 7, the presence of a professional standards board prior to 1990 proves to be significantly related to district hiring standards, a relationship that holds up after controlling for student characteristics. In addition, as Table 8 indicates, the presence of a standards board is significantly associated with the proportions of certified and uncertified teachers. This relationship may work through the influence such a board exerts over district decisions about hiring qualified personnel, as suggested above. Districts often hire unqualified teachers even though fully prepared teachers are available if state agencies do not prevent them from doing so. This can occur as a function of cumbersome hiring procedures, patronage, lack of recruitment effort or incentives, or efforts to reduce salary costs (NCTAF, 1996). Depending upon how they are structured, some standards boards may have more authority and or more commitment to prevent the hiring of unqualified teachers than some state agencies do. In agency interviews, for example, a staff member of a highly effective state standards board described how the board examines the candidate qualifications as well as the district's advertising, selection, and hiring practices and applicant pool in any case where a district requests permission to hire staff on an emergency or temporary license. Very few requests for hiring of unqualified personnel are ultimately granted, and district hiring practices are often revised and improved in the process of the review. In other states, agency officials described routine, blanket approvals of district requests for emergency hiring even in situations where districts had just laid off large numbers of qualified teachers or had qualified applicants in the

applicant pool. These officials generally felt they did not have the resources or the authority to investigate or stem practices they felt were illegal and widespread.

Table 7
Correlations (Pearson r) of Presence of a Professional Standards Board with District Hiring Standards and Teacher Qualifications

% of districts requiring graduation from an approved teacher education program	.25*
% of districts requiring a college major or minor in the field to be taught	.23*
% of districts requiring full certification, graduation from an approved program, and a college major or minor	.30**
% uncertified teachers	-.27**
% fully certified teachers	.21*
% fully certified new teachers	.21*
# of weeks required for student teaching	.25*

*p<.05 **p<.01

Table 8
Relationship between Professional Standards Board Presence and District Hiring Standards

	District hiring standards
Professional Standards Board	.411 (2.49)**
% students in poverty	.132 (.58)
% LEP students	-.429 (-2.20)*
% minority students	.067 (.26)

*p< 05 **p< 01

These relationships between the presence of standards boards and teacher education or hiring practices, although statistically significant, are quite modest (correlations in the .2 to .3 range), suggesting that many other variables are at play here as well. It is certainly true that some states enact and enforce high standards for teaching without the presence of standards boards, while some standards boards do not pursue their mission with the same vigor as others. Where they exist, however, such bodies often appear to bring greater consistency of effort and attention to the issues of preparation and qualifications.

Conclusions and Implications

This analysis triangulates data from surveys of state policies, case study analyses of state policymaking, and quantitative examination of the distribution of

BEST COPY AVAILABLE

state achievement scores and resources, taking student characteristics into account. Some findings are particularly noteworthy. First, while student demographic characteristics are strongly related to student outcomes at the state level, they are less influential in predicting achievement levels than variables assessing the quality of the teaching force. Second, when aggregated at the state level, teacher quality variables appear to be more strongly related to student achievement than class sizes, overall spending levels, teacher salaries (at least when unadjusted for cost of living differentials), or such factors as the statewide proportion of staff who are teachers.

Among variables assessing teacher "quality," the percentage of teachers with full certification and a major in the field is a more powerful predictor of student achievement than teachers' education levels (e.g., master's degrees). This finding concurs with those of other studies cited earlier. It is not surprising that masters degrees would be relatively weaker measures of teacher knowledge, given the wide range of content they can include, ranging from specialist degrees in reading or special education that are directly related to teaching to fields like administration and others that have little to do with teaching. Other measures of certification status (e.g., the percent of teachers uncertified, the percent with full certification) are also strong correlates of student achievement. Finally, certain policy strategies associated with standard-setting at the state and local level--NCATE-accreditation of teacher education institutions, district hiring standards, and, to a lesser extent, state professional standards boards--appear to be related to teacher qualifications in the field.

While the triangulation of data from several sources lends some confidence to these findings, they should be viewed with caution. Like all studies that draw inferences from broad state trends and correlational data, there are many variables in play at any given time and many possible explanations for any phenomenon observed. While this article presents a range of competing explanations for student achievement trends (student background, curriculum and testing policies, school funding and equalization, school and class sizes), it could not fully test all of these explanations. This remains for other researchers to pursue. In addition, other data and other methodologies could shed further light on these questions. Adding information about parent education levels might make a difference in the measurement of student background; adding data about school and district size (from the Common Core of Data) and curriculum and testing approaches (from the NAEP background surveys) might shed greater light on school factors that matter; and adjusting salary and spending data for cost of living differentials might allow a better evaluation of fiscal influences.

By including estimates of the proportions of staff who are underqualified (and who tend to cluster in less advantaged schools and districts), this study's estimates tapped some of the local variability in resources made available to children. However, because state data on average class sizes and other school resources ignore wide variations in teaching and learning conditions that may be very important at the district, school, and classroom levels, these estimates cannot fully capture the effects of such variables. Average class sizes, for example, vary relatively little across states but vary substantially within states and districts. Thus, effects of this variable are much more likely to be perceived with more disaggregated data. By merging district, school, and teacher files, the SASS data can allow for the use of Hierarchical Linear Modeling techniques, which would be a useful tool for further exploring relationships between teaching and schooling variables at the school, district, and state levels.

Nonetheless, the findings of this study, in conjunction with a number of other studies in recent years, suggest that states interested in improving student achievement may be well-advised to attend, at least in part, to the preparation and qualifications of the teachers they hire and retain in the profession. It stands to reason that student learning should be enhanced by the efforts of teachers who are more knowledgeable in their field and are skillful at teaching it to others. Substantial evidence from prior reform efforts indicates that changes in course taking, curriculum content, testing, or textbooks make little difference if teachers do not know how to use these tools well and how to diagnose their students' learning needs (for a review, see Darling-Hammond, 1997b).

Like other studies cited earlier, this research indicates that the effects of well-prepared teachers on student achievement can be stronger than the influences of

student background factors, such as poverty, language background, and minority status. And while smaller class sizes appear to contribute to student learning, particularly in fields like elementary reading, the gains occasioned by smaller classes are most likely to be realized, as they were in the Tennessee experiment, when they are accompanied by the hiring of well-qualified teachers. The large- scale hiring of unqualified teachers, as was the case in California's recent class size reduction initiative, would likely offset any achievement gains that could be realized by smaller class sizes.

Another implication of this study is that states may impact the qualifications of the teachers through policies that influence the hiring standards of school districts (e.g., incentives and sanctions from the state level that encourage the hiring of well-qualified individuals), the accreditation of teacher education institutions (e.g., encouragement or requirements for the use of NCATE standards or others of equivalent rigor), and the bodies that establish and enforce teaching standards (e.g. establishment of professional standards boards or assurance of adequate capacity and authority for state agencies to uphold high standards for teaching).

Although this study used fairly crude measures of teacher knowledge and skills such as certification status, college major, and master's degrees, policymakers should be aware that there are much more fine-grained distinctions to be made among types of state certification standards, teacher education programs, professional development offerings, and education requirements that make a difference to the teachers' abilities and their students' outcomes. Reforms underway to create more thoughtful licensing systems, more productive teacher education programs, and more effective professional development strategies are producing evidence of the stronger effects on teaching and learning of approaches that strengthen teachers' abilities to teach diverse learners with a keen diagnostic eye and a wide repertoire of strategies supporting mastery of challenging content (for a review, see NCTAF, 1996; Darling-Hammond 1997a). Over the next decade, federal, state, and local policymakers interested in helping students meet higher learning standards may want to consider how investments in teacher quality, along with other reforms, can assist them in achieving their goals.

Notes

1. This research was funded in part by the Office of Educational Research and Improvement (OERI) of the U.S. Department of Education through the Center for the Study of Teaching and Policy, which is housed at the University of Washington and includes Stanford University, Teachers College, Columbia University, and the University of Michigan. The research was initiated while the author was a fellow at the Center for Advanced Study in the Behavioral Sciences with the support of the Spencer Foundation. The views represented in this article are those of the author alone, and do not represent those of any sponsor.
2. National Center for Education Statistics, *Schools and Staffing Survey, 1993-94: State by State Data*, Washington, DC: U.S. Department of Education, 1996, Table 3.5. Additional tabulations performed by the National Commission on Teaching and America's Future.
3. The INTASC standards, developed by a consortium of more than 30 states and professional associations under the auspices of the Council of Chief State School Officers, are based on knowledge of effective learning and teaching and on the student learning standards developed by professional associations such as the National Council of Teachers of Mathematics. The INTASC standards for beginning teacher licensing are compatible with the more advanced standards of the National Board for Professional Teaching Standards, which define and assess accomplished teaching among veteran teachers.

References

Adams, R.D., S. Hutchinson, and C. Martray (1980). A developmental study of teacher concerns across time. Paper presented at the American Educational Research Associaton Annual Meeting. Boston, Mass.

Andrew, M. & Schwab, R.L. (1995). Has reform in teacher education influenced teacher performance? An outcome assessment of graduates of eleven teacher education programs. *Action in Teacher Education, 17*, 43-53.

Andrews, J.W., Blackmon, C.R., and Mackey, J.A. (1980). Preservice performance and the National Teacher Examinations. *Phi Delta Kappan, 61*(5), pp. 358-359.

Armour-Thomas, E., Clay, C., Domanico, R., Bruno, K., & Allen, B. (1989). An outlier study of elementary and middle schools in New York City: Final report. NY: New York City Board of Education.

Ashton, P. & Crocker, L. (1987, May-June). Systematic study of planned variations: The essential focus of teacher education reform. *Journal of Teacher Education, 38*, 2-8.

Ayers, J.B., and Qualls, G.S. (Nov/Dec 1979). Concurrent and predictive validity of the National Teacher Examinations. *Journal of Educational Research, 73* (2), pp.86-92.

Barnes, S., Salmon, J. and Wale, W. (1989), "Alternative Teacher Certification in Texas." Presented at the annual meeting of the American Educational Research Association, March. ERIC Document No. 307316.

Begle, E.G. (1979). Critical Variables in Mathematics Education. Washington, D.C.: Mathematical Association of American and National Council of Teachers of Mathematics.

Begle, E.G. and Geeslin, W. (1972). Teacher effectiveness in mathematics instruction. National Longitudinal Study of Mathematical Abilities Reports No. 28. Washington, D.C. Mathematical Association of America and National Council of Teachers of Mathematics.

Bents, Mary, and Richard Bents (1990). Perceptions of Good Teaching Among Novice, Advanced Beginner and Expert Teachers. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Berliner, D.C. & Biddle, B.J. (1995). *The manufactured crisis: Myth, fraud, and the attack on America's public schools*. Reading, MA: Addison-Wesley.

Berliner, D.C. & Tikunoff, W. J. (1976). The California Beginning Teacher Study. *Journal of Teacher Education. 27* . 24-30.

Bowles, S., and Levin, H.M. (1968). The determinants of scholastic achievement—An appraisal of some recent evidence. *Journal of Human Resources, 3*, 3-24.

Brown, C.A., Smith, M.S., and Stein, M.K. (1995). Linking teacher support to enhanced classroom instruction. Paper presented at the annual meeting of the American Educational Research Association. New York, New York.

Byrne, C.J. (1983). Teacher knowledge and teacher effectiveness: A literature review, theoretical analysis and discussion of research strategy. Paper presented at the meeting of the Northwestern Educational Research Association, Ellenville, NY.

Campbell, J.R., Donahue, P.L., Reese, C.M., Phillips, G.W. (1996). NAEP 1994 reading report card for the nation and the states. Washington, D.C.: U.S. Department of Education.

Carroll, J.B. (1975). *The teaching of French as a foreign language in eight countries* NY: John Wiley and Sons.

Carter, K., and Doyle, W. (1987). Teachers' knowledge structures and comprehension processes. Pp. 147-160 in J. Calderhead (Ed.), *Exploring Teacher Thinking*. London: Cassell.

Connecticut State Department of Education Division of Research, Evaluation, and Assessment (1991). Research Bulletin, School Year 1990-91, No. 1 (Hartford, CT: Bureau of Research and Teacher Assessment).

Cooper, E. & Sherk, J. (1989). Addressing urban school reform: Issues and alliances. *Journal of Negro Education, 58*, 3, 315-331.

Cohen, D.K. and Hill, H. (1997). Instructional Policy and Classroom Performance: The Mathematics Reform in California. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., York, R.L. (1966). Equality of educational opportunity. Washington, DC: U.S. Government Printing Office.

Council on School Performance (1997). Teachers with advanced degrees advance student learning. Atlanta: Council for School Performance, Georgia State University.

Darling-Hammond, L. (1992). Teaching and knowledge: Policy issues posed by alternative certification for teachers. *Peabody Journal of Education, 67*, 3, 123-154.

Darling-Hammond, L. (1997a). Doing What matters most: Investing in quality teaching. NY: National Commission on Teaching and America's Future.

Darling-Hammond, L. (1997b). *The right to learn: A blueprint for creating schools that work*. San Francisco: Jossey Bass.

Darling-Hammond, L.; Hudson, L.; and Sheila Kirby (1989). Redesigning Teacher Education: Opening the Door for New Recruits to Science and Mathematics Teaching. Santa Monica:The RAND Corporation.

Darling-Hammond, L., Wise, A.E. & Klein, S. (1995). *A license to teach: Building a profession for 21st century schools*. Boulder: Westview Press.

Darling-Hammond, L., Wise, A.E. and Pease S.R. (1983). Teacher evaluation in the organizational context: a review of the literature. *Review of educational research, 53*, 285-237.

Denton, J.J., and L.J. Lacina (1984). Quantity of professional education coursework linked with process measures of student teaching. *Teacher Education and Practice*, 39-64.

Denton, J.J., and Peters, W.H. (1988). Program assessment report: Curriculum evaluation of a non-traditional program for certifying teachers. Texas A &M University, College Station, TX.

Doyle, W. (1985). Recent research on classroom management: Implications for teacher preparation. *Journal of Teacher Education, 36*, 3, 31-35.

Doyle, W. (1986). Content representation in teachers' definitions of academic work. *Journal of Curriculum Studies, 18*, 365- 379.

Druva, C.A., & Anderson, R.D. (1983). Science teacher characteristics by teacher behavior and by student outcome: A meta-analysis of research. *Journal of Research in Science Teaching, 20*,5, 467-479.

Erekson, T.L. and Barr, L. (1985). Alternative credentialing: Lessons from vocational education. *Journal of Teacher Education, 36*, 3, 16-19.

Evertson, C., Hawley, W., & Zlotnik, M. (1985). Making a difference in educational quality through teacher education. *Journal of Teacher Education, 36*,3, 2-12.

Feiman-Nemser, Sharon, and Michelle B. Parker (1990). Making Subject Matter Part of the Conversation or Helping Beginning Teachers Learn to Teach. East Lansing, MI: National Center for Research on Teacher Education.

Feistritzer, C.E. (1993). Report Card on American Education: A State-by-State Analysis, 1972-73 to 1992-93. Washington, D.C.: National Center on Education Information.

Fetler, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives, 7*, 9 [Entire issue]. (Available online at http://epaa.asu.edu/epaa/v7n9.html)

Ferguson, R.F. (1991, Summer). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation, 28*,2, 465-498.

Ferguson, R.F. and Ladd, H.F. (1996). How and Why Money Matters: An Analysis of Alabama Schools. Pp. 265-298 in Helen Ladd (Ed.) *Holding Schools Accountable*. Washington, D.C.: Brookings Institution.

Ferguson, P. and Womack, S.T. (1993). The impact of subject matter and education coursework on teaching performance. *Journal of Teacher Education, 44*, 1, 55-63.

Fuller, E. J. (1999). Does teacher certification matter? A comparison of TAAS performance in 1997 between schools with low and high percentages of certified teachers. Austin: Charles A. Dana Center, University of Texas at Austin.

Glass, G.V; Cahen, L.S.; Smith, M.L.; and Filby, N.N. (1982). *School class size: Research and Policy*. Beverly Hills, CA: SAGE Publications.

Glassberg, S. (1980). A view of the beginning teacher from a developmental perspective. Paper presented at the American Educational Research Association Annual Meeting. Boston, MA.

Goebel, Stephen D., Karl Ronacher, and Kathryn S. Sanchez (1989). An Evaluation of HISD's Alternative Certification Program of the Academic Year: 1988-1989. Houston: Houston Independent School District Department of Research and Evaluation. ERIC Document No. 322103.

Goldhaber, M. and Brewer, D. (1999). Does teacher certification matter? High school certification status and student achievement. Unpublished manuscript.

Gomez, D.L., and Grobe, R.P. (1990). Three Years of Alternative Certification in Dallas: Where are We? Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Good, T.L. (1983). Recent classroom research: Implications for teacher education. In D.C. Smith (ed.), *Essential knowledge for beginning educators*. Washington, D.C.: American Association of Colleges for Teacher Education.

Good, T.L. and J. E. Brophy. (1986). *Educational Psychology, 3rd Edition*. NY: Longman.

Grady, M.P., Collins, P. & Grady, E.L. (1991). Teach for America 1991 Summer Institute Evaluation Report. Unpublished manuscript.

Greenwald, R., Hedges, L.V., and Laine, R.D. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*, 361-396.

Grissmer, D. and Flanagan, A. (1998). Exploring Rapid Achievement Gains in North Carolina and Texas. Washington, D.C.: National Education Goals Panel.

Grossman, P. L. (1989). Learning to teach without teacher education. *Teachers College Record, 91*, 2, 191-208.

Guyton, E. and Farokhi, E. (1987). Relationships among academic performance, basic skills, subject matter knowledge and teaching skills of teacher education graduates. *Journal of Teacher Education, 38*. (Sept-Oct.), 37-42.

Haberman, M. (1995, June). Selecting 'star' teachers for children. *Phi Delta Kappan, 76*, 777-781.

Haney, W. (1999). Supplementary report on Texas Assessment of Academic Skills Exit Test (TAAS-X). Boston: Center for the Study of Testing, Evaluation, and Educational Policy.

Haney, W., Madaus, G., & Kreitzer, A. (1987). Charms talismanic: testing teachers for the improvement of american education. Pp. 169-238 in E.Z. Rothkopf (Ed.) *Review of Research in Education*, Vol. 14. Washington, D.C.: American Educational Research Association.

Hansen, J. B. (1988). The Relationship of Skills and Classroom Climate of Trained and Untrained Teachers of Gifted Students. Unpublished doctoral dissertation, Purdue University.

Hanushek, E.A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *The American Economic Review, 61*(2), 280-288.

Hanushek, E.A., Rivkin, S.G., and Taylor, L.L. (1995). Aggregation bias and the estimated effects of school resources. Working paper 397. University of Rochester, Center for Economic Research.

Hawk, P.; Coble, C.R.; and Swanson, M. (1985). Certification: It Does Matter. *Journal of Teacher Education, 36* , 3, 13-15.

Hellfritzch, A.G. (1945). A factor analysis of teacher abilities. *Journal of Experimental Education, 14*, 166- 169.

Hice, J.E.L. (1970). The Relationship between Teacher Characteristics and First-Grade Achievement. *Dissertation Abstracts International, 25*, 1, 190.

Johanson, G.A. and Gips, C.J. (1992). The hiring preferences of secondary school principals. *The High School Journal, 76*, (Oct./Nov.), 1-16.

Kentucky Institute for Education Research (1997). The preparation of teachers for Kentucky Schools: A survey of new teachers. Frankfort, KY: Kentucky Institute for Education Research

Klitgaard, R.E. and Hall, G.R. (1974). Are there unusually effective schools? *Journal of Human Resources, 10*, 3, 90-106.

Knoblock, G.A. (1986). Continuing professional education for teachers and its relationship to teacher effectiveness. Unpublished dissertation. Michigan State University. *Dissertation Abstracts International, 46*, 2, 3325A, University microfilms no. AAC8529729.

Kurtz, M. (1999). Schools, teens battle barrier of 9th grade. *American-Statesman*,

May 23, 1999.

LaDuke, D.V. (1945). The measurement of teaching ability. *Journal of Experimental Education, 14*, 75-100.

Lenk, Harriet Anne (1989). A Case Study: The Induction of Two Alternate Route Social Studies Teachers. Unpublished doctoral dissertation. Teachers College, Columbia University.

LuPone, L.J. (1961). A comparison of provisionally certified and permanently certified elementary school teachers in selected school districts in New York State. *Journal of Educational Research, 55*, 1, 53-63.

Lutz, Frank W., and Jerry B. Hutton (1989). Alternative teacher certification: Its policy implications for classroom and personnel practice. *Educational Evaluation and Policy Analysis, 11*, 3, 237-254.

McNeil, J.D. (1974). Who gets better results with young children -experienced teachers or novices? *Elementary School Journal, 74*, 447-451.

Mitchell, N. (1987). Interim Evaluation Report of the Alternative Certification Program (REA87-027-2). Dallas, TX: DISD Department of Planning, Evaluation, and Testing.

Monk, D. H. (1994). Subject matter preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13*, 2, 125-145.

Monk, D. H. and King, J.A. (1994). Multilevel teacher resource effects in pupil performance in secondary mathematics and science: The case of teacher subject matter preparation. Pp. 29- 58 in R.G. Ehrenberg (Ed.), *Choices and consequences: Contemporary policy issues in education*. Ithaca, NY: ILR Press.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children, 5*, 2, 113- 127.

Murnane, R.J. (1985, June). Do effective teachers have common characteristics: Interpreting the quantitative research evidence. Paper presented at the National Research Council Conference on Teacher Quality in Science and Mathematics, Washington, D.C.

Murnane, R.J., and Phillips, B.R. (1981). Learning by doing, vintage, and selection: Three pieces of the puzzle relating teaching experience and teaching performance. *Economics of Education Review, 1fR*, 4, 691-693.

National Center for Education Statistics (1994). Data compendium for the NAEP 1992 reading assessment of the nation and the states: 1992 NAEP trial state assessment. Washington, D.C.: U.S. Department of Education.

National Center for Education Statistics (n.d.) NAEP 1992, 1994 national reading assessments – Data almanac – Grade 4: Teacher questionnaire weighted percentages and composite proficiency means (Public school). (Available online at http://www.Nces.ed.gov/nationsreportcard/y25alm/almanac.shtml)

National Association of State Directors of Teacher Education and Certification (NASDTEC)(1997). Manual on certification and preparation of educational personnel in the United States and Canada, 1997-98.

National Center for Education Statistics (NCES) (1995). Digest of Education Statistics, 1995. Washington, D.C.: U.S. Department of Education.

BEST COPY AVAILABLE

National Commission on Teaching and America's Future (NCTAF) (1996). What Matters Most: Teaching for America's Future. New York: Author.

Natriello, G., Zumwalt, K., Hansen, A., and Frisch, A. (1990). Characteristics of Entering Teachers in New Jersey. Revised version of a paper presented at the 1988 Annual Meeting of the American Educational Research Association.

Olsen, D. G. (1985). The quality of prospective teachers: Education vs. noneducation graduates, *Journal of Teacher Education, 36*, 5, 56-59.

Penick, J.E., and Yager, R.E. (1983). The search for excellence in science education. *Phi Delta Kappan, 64*, 8, 621-623.

Perkes, V.A. (1967-1968). Junior high school science teacher preparation, teaching behavior, and student achievement. *Journal of Research in Science Teaching, 6*, 4, 121-126.

Pflaum, S.W. and Abramson, T. (1990, March). Teacher assignment, hiring, and preparation: Minority teachers in New York City. *The Urban Review, 22*, 17-31.

Quirk, T.J., Witten, B.J., and Weinberg, S.F. (1973). Review of studies of concurrent and predictive validity of the National Teacher Examinations. *Review of Educational Research, 43*, 89-114.

Reese, C.M., Miller, K.E., Mazzeo, J., Dossey, J.A. (1997). NAEP 1996 Mathematics Report Card for the Nation and the States. Washington, D.C.: National Center for Education Statistics, U.S. Department of Education.

Rosenholtz, S. J. (1986). The organizational context of teaching. In *Learning to Teach*. University of Illinois at Champaign-Urbana.

Rosenshine, B. and Furst, N.F. (1973). The use of direct observation to study teaching. In R.M.W. Travers (ed.), *Handbook of Research on Teaching (2nd Edition)*. Chicago: Rand McNally.

Rostker, L.E. (1945). The measurement of teaching ability. *Journal of Experimental Education, 14*, 5-51.

Rottenberg, Claire J. and David C. Berliner (1990). Expert and Novice Teachers' Conceptions of Common Classroom Activities. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Roupp, R.; Travers, J.; Glantz, F.; and Coelen, C. (1979). Children at the Center: Summary Findings and their Implications. Cambridge, MA: Abt Associates.

Sanders, S.L., Skonie-Hardin, S.D., and Phelps, W.H. (1994, November). The effects of teacher educational attainment on student educational attainment in four regions of Virginia: Implications for administrators. Paper presented at the Annual Meeting of the Mid-South Educational Research Association.

Sanders, W.L. & Rivers, J.C. (1996). Cumulative and residual effects of teachers on future student academic achievement. Knoxville: University of Tennessee Value-Added Research and Assessment Center.

Schalock, D. (1979). Research on teacher selection. In D.C. Berliner (Ed.), *Review of research in education, Vol. 7*, Washington, D.C.: American Educational Research Association.

Skinner, W.A. (1947). An investigation of factors useful in predicting teaching ability. University of Manchester. Master of Education thesis.

BEST COPY AVAILABLE

Skipper, C. E. and Quantz, R. (1987). Changes in educational attitudes of education and arts and science students during four years of college, *Journal of Teacher Education, 38*, May- June, 39-44.

Soar, R.S., Medley, D.M., and Coker, H. (1983). Teacher evaluation: A critique of currently used methods. *Phi Delta Kappan, 65*, 4, 239-246.

Stoddart, T. (1992). An alternate route to teacher certification: Preliminary findings from the Los Angeles Unified School District Intern Program. *Peabody Journal of Education,67*, 3.

Strauss, R. P. and Sawyer, E.A. (1986). Some New Evidence on Teacher and Student Competencies. *Economics of Education Review, 5*, 1, 41-48.

Summers, A.A., and Wolfe, B.L. (1975, February). Which School Resources Help Learning? Efficiency and Equality in Philadelphia Public Schools. Philadelphia, PA: ED 102 716

Taylor, J.K. and R. Dale (1971). A Survey of Teachers in the First Year of Service. Bristol: University of Bristol, Institute of Education.

Texas Education Agency (1993). Teach for America Visiting Team Report. Austin: Texas State Board of Education Meeting Minutes, Appendix B.

Vernon, P.E. (1965). Personality Factors in Teacher Trainee Selection. *British Journal of Educational Psychology, 35*, 140-149.

Walberg, H.J., Waxman, H.C. (1983). Teaching, learning, and the management of instruction. In D.C. Smith (Ed.), Essential knowledge for beginning educators. Washington, DC: American Association of Colleges for Teacher Education and ERIC Clearinghouse on Teacher Education.

Wiley, D. & Yoon, B. (1995). Teacher reports of opportunity to learn: Analyses of the 1993 California Learning Assessment System. *Educational Evaluation and Policy Analysis, 17*, 3, 355-370.

Wise, A.E.; Darling-Hammond, L.; and Berry, B. (1987). Effective Teacher Selection, From Recruitment to Retention. Santa Monica, CA: RAND Corporation.

Wright, S.P.; Horn, S.P.; and Sanders, W.L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 57-67.

About the Author

Linda Darling-Hammond
School of Education
Stanford University

Email: ldh@leland.stanford.edu

Linda Darling-Hammond is Charles E. Ducommun Professor of Education at Stanford University and executive director of the National Commission on Teaching and America's Future. Her research, policy, and teaching focus on teacher education and teaching quality, school restructuring, and educational equity. Among other writings, she is author of *The Right to Learn*, which received the Outstanding Book Award from the American Educational Research Association in 1998.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina – Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois– UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografía e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

archives | abstracts | editors | board | submit | comment | subscribe | search

EPAA • enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 1699 times since January 2, 2000

Education Policy Analysis Archives

Volume 8 Number 2 January 2, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

America Y2K: The Obsolescence of Educational Reforms

Sherman Dorn
University of South Florida

Abstract
The passing of the deadline for fulfillment of the national education goals in the United States (the beginning of 2000) reflects the frequently hyperbolic statements of objectives and the manic pace of school reform efforts over the past two decades. The domination by schools of child and family life has combined with a longstanding reliance on schools to solve social problems to make school reform a politically opportune as well as visible issue. Thus, even if the phrasing of national education goals in the U.S. changes to reflect the passing of the nominal deadline, those pressures will remain.

Yesterday, observers of educational reform in the United States woke up to the policy equivalent of the Y2K problem: what does a nation do when a set of official goals has become obsolete with the passage of time? A summit of the nation's governors and then-President George Bush in 1989 declared the first six national education goals as part of an "America 2000" strategy for highlighting key targets. (See the National Education Goals Panel website for more information.). "*By the year 2000*," each of the (now) eight goals has asserted, the nation would have kindergartners ready to learn, 90 percent graduation, solid academic achievement (including "first in the world" achievement in science and math), a literate adult population, safe and drug-free schools, superb professional development for teachers, and committed parental involvement in schools. As those who are reading this article on computers (Y2K-compliant or not) can attest, we have reached the deadline for every goal. Yet we have apparently not reached the goals. Overall, of the 28 key indicators chosen by the National Education Goals Panel, 16 have shown either no improvement or declines. The most concrete goal, 90 percent graduation, was within striking distance in 1990 but has eluded our collective grasp: 86 percent of 18-24 year olds had high school diplomas or alternative credentials in 1990, while

BEST COPY AVAILABLE

85 percent had done so in 1998 (National Education Goals Panel,1999). Faced with the nominal obsolescence of specific national education goals in the U. S., perhaps we should rename the America 2000 strategy the America Y2K problem, for the goals self-destructed at midnight.

Curiously enough, the general conclusion of the most recent report of the National Education Goals Panel entirely eschews matters of outcomes:

> We believe that the National Education Goals have moved America forward and, on balance, encouraged greater progress in education. We are clearer about what appropriate Goals are and how to measure progress toward them at the national and state levels. There is no doubt that the National Education Goals have encouraged a broad spectrum of educators, parents, students, business and community leaders, policymakers, and the public to work toward their attainment. Reporting progress toward the Goals has provided valuable information to states and inspired them to reach higher. Can we do better? Of course we can. But we are convinced that our gains have been greater because we have had National Education Goals to guide our efforts. Ten years of progress have shown us that the Goals are working. (National Education Goals Panel, 1999, p. 6)

The singular discussion of process above seems to contradict the whole notion of evaluating policy using concrete outcomes. One may wonder whether such a conclusion constitutes denial. After all, with substantial evidence that a national effort to reform education has not met its putative goals, is such a paragraph mere hedging in the face of the panel's own data? I believe such a criticism is unfair, for two reasons. First, one should measure a policy discussion not only by the realities one can observe on the ground but also in the agenda it sets for the future. Whether one agrees with the specific goals or the notion of a national education agenda, the summit in 1989 did help frame the policy debate that has ensued. Second, the deadline itself was primarily an instrument of political rhetoric, in the eyes of its creators a useful goad for change. The focus on process in the report is a pedestrian rather than a weighty irony, in this instance. The more substantive criticism of federal policy should aim at the content and means of reform.

Still, the deadline reflects what the rest of the world often sees as prototypically optimistic boasting of the United States. Such optimism has some side effects, as Potter (1954) described almost half a century ago. We in the U.S. often feel pressured by the assumption of affluence to individual and collective acts of hype and disappointment. The New Year (whether one believes we are in a new millennium yet or not) should prompt some reflection on the workings of such an approach to social change. The failure to meet the national education goals was the result of a common dynamic in school reform. The problem with the national education goals was not that they set virtually unreachable goals but that they were not unusual in attempting to push change by setting impossible standards.

A brief survey of recent educational reform efforts in almost any city or state illustrates the impatience in modern reform dynamics. Chicago witnessed first the radical decentralization of control over schools in the 1988 reform legislation and then recentralization in the hands of Mayor Richard Daley in the years since 1995. Florida and California are two examples of rapid-fire reforms at the state level. In the last quarter-century, Florida schools have been the target of minimum competency tests, increased seat-time requirements for graduation, mandatory standardized testing for students, teacher competency tests, the removal of state mandates for universal standardized tests and their replacement with partly performance-based testing in several (but not all) grades, site-based management of schools, alternative credentialing procedures for teachers, the reinstallation of both criterion- and norm-referenced testing in the majority of grades, the public grading of schools on an A-F basis, and vouchers. California schools have witnessed many of these efforts as well as an aborted experiment in performance-based assessment for the whole state and a highly politicized battle over methods of teaching reading.

Larry Cuban argued that much of the educational reform dynamic begins with the unreasonable demands we have placed on schools to accomplish social reform in

the U.S (Cuban, 1990). Historians can trace back almost two hundred years a chain of statements assuming the power of formal schooling to eliminate or ameliorate poverty, and the first legal decrees requiring education in British North America (albeit mandating family rather than formal schooling) were to promote morality in the seventeenth century. To the extent that we keep expecting schools to solve all our social problems, we are overestimating their power. Cuban's argument about how social reformers have used schools to avoid resolving broader political conflicts helps explain much of the rhetoric of school reform over the past twenty years. *A Nation at Risk* (1983) blamed schools for economic woes in the midst of a broad trend towards deindustrialization that we now call "economic globalization" (Harrison & Bluestone, 1988; National Commission on Excellence in Education, 1983; *The Nation* (Dec. 6, 1999 issue)). The protesters at the Seattle meeting of the World Trade Organization argued that key politicians around the world were hiding the social dislocation and other problems of international capital liquidity behind the platitudes of free trade. In the meantime, one of the alleged bromides for such dislocation in the United States has been, predictably, educational reform. Certainly no one could argue with "world-class" achievement for any child. But are there any world-class standards for family subsistence (on which, not incidentally, one must base a poor child's education)?

One must acknowledge, however, that the rhetoric of school reform is not merely a shadow-game. It has such political power because it resonates at some level with parents' and other citizens' experiences. Parents may not know much about the debates over globalization, but most want their children to be able to get and keep jobs as adults, and they may well perceive the quality of an education, or at least an educational credential, as important to that goal. Some of those parents and their neighbors purchased their homes in part on the reputation of local schools. In addition, parents do not have the luxury of waiting five to ten years for deeper school reform to affect their children; in the life of a child and her family, a year is a very long time.

Part of this impatience with and targeting of schools also comes from the expansion of schools' role within the daily routines of families. One hundred years ago, formal schooling was one of many ways that a child spent time. Far more seventeen-year-olds worked than studied in high schools. Even for younger children, attendance was sparse compared to the present. (That some children are regularly truant in contemporary schools is an exception that proves the rule; a century ago, attendance was less regular for most students.) Today, by contrast, children's and parents' lives in the United States revolve around the school schedule. Schooling has become an institution that dominates time and consciousness, affecting our assumptions about what is important. One response to such dominating organizations is to target those key institutions for inspection, concern, and responsibility for solving broader problems. Thus, voters are willing to credit politicians with concern about schools, apparently legitimating expectatins that no school reform effort could meet.

Many observers have commented on the practical problems of trying to reform schools dramatically in a short time Sarason, 1990; Tyack and Cuban, 1995), and I do not wish to revisit those issues here. Rather, my assertion is that several factors, some longstanding in North American culture and others more recent, have encouraged and helped legitimate the obsession with speedy statewide and nationwide school reform. The foreseeable obsolescence of the national educational goals thus represents the culmination of the reform dynamic, not the exception. One may wonder, then, what shall be the fate of the outdated goals? Extensive sociological writings exist on how organizations change their goals. The first work commonly cited, Michels' *Political Parties* (1915/1959), describes what he called the "iron law of oligarchy," the way that the need to create a political apparatus to affect legislation shifted the emphasis of party organizations from the original ideals onto party maintenance and thus made those political structures conservative. The ensuing literature on organizational goals expanded this notion of shifted goals from goal displacement (such as the evolving goals of political organizations) to goal abandonment or, alternatively, goal succession with the achievement of explicit goals (Blau, 1956). The classic example of goal succession in the United States is the March of Dimes, originally organized to ameliorate the suffering of polio

victims. Its leaders later sponsored the mass field tests of the Salk polio vaccine and realized with the success of the vaccine that it had worked itself out of a job. The national board quickly found another (in the field of birth defects) (Sills, 1958). The literature on the history of goals in organizations suggests that the internal needs of organizations help shape the specific future for written goals is automatic.

The major difference between the problems of organizational goals and the national education goals is that the education goals were the putative objectives not of a specific institution but of an entire country. The dynamics of a single organization are simply not an issue in educational politics or public policy in general. Nonetheless, one can draw the lesson from organizational sociology that a larger version of institutional dynamics, specifically how people have built their lives around the existence of routines, strongly influences what happens to explicit objectives. The course of political goals (and here I mean nothing pejorative by calling them political) depends on partisan struggle and also on how the structure of people's experiences (in this case, the organization and practices of schooling) help define what people see as important. To be specific, schools have evolved a complex set of goals that have a complicated, interdependent relationship with how individuals become active in educational politics. In the nineteenth century, Katznelson and Weir (1985) have argued, public education became tied to the franchise as both universal white male franchise and free elementary schooling spread through the United States. Since then, those active in educational politics have become involved in many ways depending on their interests and whether they define schooling as a matter of concern for them as residents of a neighborhood, as workers in an economy, (more recently) as consumers of various markets, or in some other way tied to some aspect of their identities. Schools have accrued these purposes and associated identities as they have become well-established in the United States, and these agglomerated interests are unlikely to disappear.

One caveat to this general argument about the intransigence of speedy reform is important. The new theme of choice in educational politics over the past ten or fifteen years in the United States is likely to complicate the reformulation of educational reform, possibly at the expense of achievement goals (See note below.) Not all parents believe that measurable achievement is the most important purpose of schooling, and arguments in favor of parents' power over schooling is likely to undermine arguments in power of the state's interest in improving test scores and other measures of achievement. What is less likely is for the notion of choice in schooling (whether public or private) to affect the momentum of high-stakes reforms. The shape of those reforms may change, but until schools become far less important to the everyday lives and concerns of families, the reasons for political opportunity in education reform will remain. Voters will remain concerned about formal education for a variety of reasons, and officeholders and candidates will demand reform as a way of establishing political credentials.

One can thus predict, with some accuracy, that the national education goals will undergo some amendment in the near future, but in a way to keep some implicit pressure on schools and public policy to change. I suspect that the National Education Goals Panel will not simply replace "2000" with "2010" or some such formulation that will invite ridicule. Instead, a more vague phrasing is likely to appear, suggesting the imperative nature of change without specifying another deadline. The essential dynamic will remain, though, of demands for change that occasionally shift in emphasis. The "waves" of reform will keep pounding on our political shores. A recent report on deaths caused by medical errors in the United States provides an unusual and sad reason for comparing educational and medical systems in this imperative for action: for once, observers of school reform can tell medical reformers what to expect from attempted systemic change. The paper, by the Institute of Medicine's Committee on Quality of Health Care of America, estimated that medical mistakes cause more than 40,000 deaths annually in the U.S. It recommended a vigorous accountability system to report all medical mistakes, a center for patient safety to set safety goals and monitor progress towards them, and a reassessment of such progress at the end of five years, by which time the committee hopes such deaths would fall by half (Corrigan, Kohn, & Donaldson, 1999). One can examine this report as an example of attempted reform and analyze the factors that may affect its success. Cutting mortality from any cause in half within five years is

desirable, but this result would require the type of fundamental change in health care that the creation of a center is unlikely to stimulate. If, as the report indicates, overworked staff members in poorly-funded and -supplied institutes are more likely to make mistakes than others, then the stingy characteristics of the managed care system in the U.S. are likely to thwart much of the power of reporting, tracking, and analysis of a center on the ultimate medical accountability—life. In this respect, the report on fatal medical mistakes is eerily similar to attempts to improve education through statistics-gathering and accountability mechanisms.

Such a comparison, however comforting it may be to cynical observers of school reform, is not likely to be a revelation to scholars of public health. Medical historians and sociologists are well aware of problems with technocratic approaches to public health concerns. For example, assumptions about the ability to conquer sexually-transmitted diseases by antibiotics have, in retrospect, hidden much of the moralizing aspects of the anti-venereal disease campaigns early in the century (Brandt, 1985). Few on the committee are likely to underestimate the difficulties involved in such broad goals. Instead, perhaps a more useful way of looking at the report is to see it as an example of an ambitious set of goals and deadlines that are impossible to meet. In that regard, the goal of halving mortality from medical mistakes is akin to the establishment of national goals for education. All are certainly worthy ideals in an abstract sense. Yet what is driving the putative timetable for reform is not feasibility but the vulnerability many citizens feel in connection with both schools and hospitals. One consequence of setting such goals is having at some point to re-evaluate their attainment and, ultimately, legitimacy. Whether the United States will have such an open political debate on the national education goals or the appropriate pace of reform is unknown.

Note

Jurgen Herbst, professor emeritus from the University of Wisconsin-Madison, is currently researching a comparison of school choice history in the United States and central Europe, and his work is likely to suggest, as Claire Smrekar's does, the diversity of private purposes for education in the context of choice.

References

Blau, Peter M. *Bureaucracy in Modern Society*. New York: Random House, 1956.

Brandt, Allan. *No Magic Bullet: A Social History of Venereal Disease in the United States since 1880*. New York: Oxford University Press, 1985.

Corrigan, Janet, Linda Kohn, and Molla Donaldson, eds. *To Err Is Human: Building a Safer Health System*. Washington, D.C.: National Academy Press, 1999.

Cuban, Larry. "Reforming Again, Again, and Again." *Educational Researcher* 19, 1 (1990): 3-13.

Harrison, Bennett, and Barry Bluestone. *The Great U-Turn: Corporate Restructuring and the Polarizing of America*. New York: Basic Books, 1988.

Katznelson, Ira, and Margaret Weir. *Schooling for All: Class, Race, and the Decline of the Democratic Ideal*. Berkeley: University of California Press, 1985.

Michels, Robert. *Political Parties: A Sociological Study of the Oligarchical Tendencies of Modern Democracy*. Trans. Eden and Cedar Paul. New York: Dover Publications, 1915/1959.

National Education Goals Panel. *The National Education Goals Report: Building a Nation of Learners, 1999*. Washington, D.C.: U.S. Government Printing Office, 1999. Retrieved December 2, 1999, from the World Wide Web: http: www.negp.gov

Potter, David M. *People of Plenty: Economic Abundance and the American Character*. Chicago: University of Chicago Press, 1954.

Sarason, Seymour B. *The Predictable Failure of Educational Reform: Can We Change Course before it's too Late?* San Francisco: Jossey-Bass, 1990.

Sills, David L. *The Volunteers: Means and Ends in a National Organization*. New York: The Free Press, 1958

Smrekar, Claire. *The Impact of School Choice and Community: In the Interest of Families and Schools*. Albany, N.Y.: State University of New York Press, 1996.

Tyack, David, and Larry Cuban. *Tinkering toward Utopia: A Century of Public School Reform*. Cambridge, Mass.: Harvard University, 1995.

About the Author

Sherman Dorn

Email: dorn@typhoon.coedu.usf.edu

Sherman Dorn is Assistant Professor in the Department of Psychological and Social Foundations at the University of South Florida. He received his Ph.D. in history at the University of Pennsylvania in 1992 based on his work on the history of dropout policies. He is currently looking at the history of special education in Nashville, Tennessee, from 1940 to 1990.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina -- Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (Mexico)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

EPAA • enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 1111 times since January 5, 2000

Education Policy Analysis Archives

Volume 8 Number 3 January 5, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Social Science Research Findings and Educational Policy Dilemmas: Some Additional Distinctions

Steven I. Miller
Loyola University, Chicago

Marcel Fredericks
Loyola University, Chicago

Abstract

The article attempts to raise several distinctions regarding the presumed relationship of social science research findings to social policy making. The distinctions are made using Glymour's critique of the *Bell Curve*. An argument is made that (1) social science models and research findings are largely irrelevant to the actual concerns of policy makers and (2) what is relevant, but overlooked by Glymour, is how ideological factors mediate the process. The forms that ideological mediation may take are indicated.

Although there have been a variety of attempts to understand how social science research does or does not affect the "voices" of those being studied (Harding, 1993; Longino, 1993), we wish to revisit the issue from another angle. What has been overlooked in even the most ambitious constructivists' forays (Fuller, 1988) into dominant epistemologies is *why* such research findings are, generally, so overwhelmingly ineffective in social policy formulation. That is, we wish to consider some of the deeply implicit notions of the "research act" (Denzin, 1989) itself; those that contribute to either the tacit acceptance of such knowledge production or generate vociferous attacks (Lakatos, 1978) of various sorts. More specifically, our argument is that social policy makers assume an atypical "gatekeepers" role where, in this case, they must attempt to appropriate, translate, and filter social science research findings to relevant publics; however, the very act of doing so is most likely doomed to fail. Those who are then to "benefit" from the

BEST COPY AVAILABLE

social policies, informed and enlightened by social science findings, are the very ones whose voice often cannot be heard.

The issue is, to use Quine's (1969) overworked phrase, one of an "indeterminancy of translation." It is not that a translation is impossible, however, but rather that some thing is lost *in* the translation. What is lost is the subject of our analysis, including an attempt to show—again borrowing from Quine (1960)—that there is indeed a "fact of the matter" about all of this, but an unexpected one. We will attempt to show how the "translation" issue works by using the recent analysis of the well known philosopher of science, Clark Glymour, to account for the relationship of social science research, to social policy, to social practice. Specifically in his provocative article, "What went wrong? Reflections on Science by Observation and *The Bell Curve* (1998:1-32), Glymour recognizes the issues of evidence and policy relevant to both the philosophy of science and social science and how they overlap into the ambiguous realm of public policy- making. However, the need for additional analysis lies not only in the fact that Glymour has not fully explored a series of mostly implicit, but very significant, assumptions that are involved in social policy making, but also to illustrate that the nexus of scientific thinking and the formulation of social policy often support ideologically-based belief systems that selectively utilize "scientific" findings. Our aim will be to illustrate how even a well-known philosopher such as Glymour falls victim to the very trap he is trying to expose and avoid.

To begin with, Glymour's critique of the methodological (and in a deeper sense, ontological) issues he raises concerning the analysis of *The Bell Curve* (1994) are arguably some of the best made to date. The social sciences, Glymour argues, have been plagued by the alleged importance of uncovering the causal mechanisms underlying social behavior and practices. This is not a new problem. What is important, as he points out, is the inability of the social sciences to acknowledge that these implicit causal structures are highly complex, and being so, how they can produce contradictory conclusions within a given research domain. The complexity of these causal structures is often overlooked by social scientists because of implicit beliefs concerning the validity of the methodological techniques themselves (Campbell, 1987). For instance, if a social scientist can employ such relatively powerful quantitative techniques as multiple regression, discriminate analysis, and factor analysis, there are usually two corresponding beliefs that seem to come into play: (1) that such techniques take precedence over "philosophical" beliefs concerning the nature of (and presumed importance of) causality, and (2) the use of such techniques, irrespective of their ability—or lack of—to uncover true causal structures, still *improves* the claims that can be made about social behavior over-and-above what could be said in their absence. Again such debates, as Glymour correctly points out, mistake the importance of clear causal thinking with the technical application of methods.

He states the issue (p. 1):

> Social statistics promised something less than a method of inquiry that is reliable in every possible circumstance, but something more than sheer ignorance; it promised methods that, under explicit and often plausible circumstance, converge to the truth, whatever that may be, methods whose liability to error in the short run can be quantified and measured.

Glymour further correctly points out (pp. 2-3) that social scientists are still under the sway of a certain form of positivism that is suspicious of causal analysis itself. For him, there is a solution: "Clear representation by directed graphs of causal hypotheses and their statistical implications, in train with rigorous investigation of search procedures, have been developed in the last decade in a thinly populated intersection of computer science, statistics and philosophy" (p. 3). However, even this solution, potentially elegant as it is, in our view, will not provide the needed framework for rational social policy making. We will try to address why this is so in the sections that follows.

I.

To put the issue rather crudely, for those engaged in the policy making

process what Glymour envisions, "just doesn't matter!" What we mean by this is that in social policy making, at many levels and across a variety of contexts, the discovery and justification of elegant (or even elementary) causal processes is largely irrelevant to the decisions made by policy makers. Part of the problem, to begin with, is the fact that there is what we will call an "ontological bifurcation" between social scientists and policy makers (who are usually not social scientists). These two groups—at least based on our own experiences—simply view the "world" in different ways, and often in such fundamentally different ways, that although they *want* to communicate often they cannot because, ultimately, they are unable to do so. While the story of why this is so is rather complex, Fuller's attempt to explain it is relevant here. He wrote (1988), for example,

> Unfortunately, as our remarks were meant to suggest, *the crucial epistemological differences occur at the level of the different textual embodiments*, since a popularization of quantum mechanics offers the lay reader no more access to the work of the professional physicist than a state-of-the-art physics text offers the professional physicist access to the general cultural issues which interest the lay public. [His emphasis.](p. 272)

There are indeed different "textual embodiments" that are at the heart of the issues, but for us the policy maker-as-gatekeeper role is the crucial one to consider. This role serves as the principle "translator" one, mediating between the social scientist-as-researcher and the voices of specifically involved publics. In contrast with Fuller, however, we see the issue as primarily "ontological", although heavily conditioned by the epistemological. By this we mean, the issue of increased technique-sophistication, along with the causality issue, is *believed* to be necessary (and possibly sufficient) for an increasingly satisfactory and accurate "ontological-representation" of what social science research findings *can* do. We are suggesting, on the other hand, that the very belief in what social science can do for social policy making is at the center of differing views of (social) reality between these two groups, leaving aside the affected publics. One initial way of capturing the difference is to begin with a few "themes" about evidence that figure into the debate but are often not explicitly indicated as such. These themes are fundamentally about what constitutes "good" evidence for (eventually) the making of "good" policy, or about how differing textual embodiments come about.

Theme 1: "What is your evidence?"

From the policy maker's side of the ontological divide, the pressing issue is to be able to "take and use" the evidence of social science research, with methodological finesse(ness) be damned. Moreover, this is often the case for policy makers who are trained as social scientists. The issue of the evidence theme takes various forms. Perhaps, the most central one centers around the following distinction: "What evidence counts?" vs. "What counts as evidence?" The distinction is one with a difference, as we see it. Taking the latter one first, what counts *as* evidence includes a large class of possibilities, such as empirical and non-empirical (i.e., qualitative), historical, legal data, and so forth (Miller & Safer, 1993). Any of these *types* of evidence may be deemed to be relevant by the policy maker in terms of formulating, implementing or evaluating a given social policy. (Note 1) The issue is not trivial since how it is addressed, and by whom, can determine a wide range of decisions affecting peoples lives in terms of what voices they may or may not eventually have.

What is crucial to see, however, is how choices as to what does *not* count as evidence automatically entail *what* evidence counts. Thus, if we reject the use of, for example, ethnographic findings as evidence for a social policy issue, and our only other choice is some type of empirical evidence, then the process of elimination dictates the epistemological choice of what evidence counts. Here we may find a great deal of variation: experimental vs. correlational findings, for instance, and both further delineated by way of causal robustness. Moreover, each type of evidence may be further distinguished by such factors as "weight" and "number". Thus, the

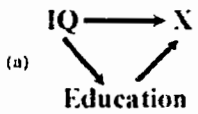
"weight of the evidence" may be a function of how "much" there is of it and how these concerns are counterbalanced by "internal" factors such as sampling strategies and numbers, parametric vs. non-parametric measures, the putative validity and reliability of measures used, their "normal distribution", and so on.

All of these considerations need to be, but seldom are, taken into consideration by the policy maker. Or, more precisely, even when they are their eventual impact on the policy making process is usually minimal.

Theme 2: "Do you have a causal model?", or "Does your data give rise to or support a pre-determined causal model?"

In many policy making scenarios, Theme 2 may or may not be related to *Theme 1*, and this from either side of the ontological divide. Social scientists who serve as (adjunct) policy makers in their role of "experts", based on our experience, seldom, if ever, *explicitly* engage in discussions of the causal robustness or the efficacy of their models. At best, such attempts are ad hoc; even where publication in empirical social science journals is concerned, the issue of "causality" is usually given the obligatory conceptual "nod" but then quickly forgotten. From the view of the non-social scientist policy maker the issue is moot, since it is usually so far divorced from what *needs* to be accomplished, it is perceived as irrelevant.

However, where a causal model could be specified with the precision argued for by Glymour, the implications for policy making are probably not as dramatic as he makes them out to be. Consider his two models (pp. 16-18, figures 12 and 13, respectively) as examples.



In (a), Herrnstein and Murray's (1994) model, IQ is the presumed cause of X (let's say some outcome variable), and while Education may "intervene" or "mediate" the IQ -- X relationship, something the social scientist would want to know, Glymour argues the "answer" to (a) may be mistaken because of the inability to account for the possibility of "U" in case (b). The "U" (e.g., "latent factors", other unknown "variables") may themselves be correlated with X and Education and hence give a false picture of what is presumed in (a).

Now, both (a) and (b) are examples of models that "count". Let's also assume that (b) is somehow fully specified and with "U" accounted for the role of Education is either enhanced or drastically reduced (i.e., in terms of explained variance). What is the social scientist-as- policy maker and policy-maker-non-social-scientist to make of this *for* policy purposes? The first may examine the total amount of variance explained (i.e., R^2), with or without the underlying causal structure, as not being *that* relevant. By this we mean, the social scientist *as* policy maker may: (1) judge (b) to be a "better" causal model because when "U" is taken into account the overall percentage of variance explained in X is "greater" than in (a), (2) maintain faith in (a) because the amount of unexplained variance (i.e., $1 - R^2$) has not been "sufficiently" reduced in model (b), or (3) perhaps "go with" (a) or (b) depending on what "U" is determined to be. If U is something like the mysterious "g-factor" for ability, as opposed to a more "straightforward" variable such as, hypothetically, "Parental Attitudes", the decision may be to stick with model (a) because it is putatively more amenable to policy making. On the other side, the non-social scientist policy maker (even given some understanding of the technical issues) still needs to know *what* to do—and (a) or (b) will not be very useful here. Why not?

One reason is that the policy maker (perhaps of either variety) is

engaged—although most likely implicitly—in the formulation of a *practical argument*; one, roughly, similar to Aristotle's (*DeMotu Animalium*, Ch. 7., *Nicomachean Ethics VU*, 3:1 47a; *VI*, 2:113a, *DeAnima III*, II:1143b. (cited in Green, 1980:xvi) where the conclusion of the argument is in the form of an "act", or here for the policy maker, "Do X." In such a case, even a well formed argument with "true" premises is no guarantee that a policy maker will take such an argument seriously (Miller and Safer, 1993). For the policy maker, who happens to be a philosopher of social science, let us say, the situation is even more desperate. Even with a fully specified model of the kind argued for by Glymour, the philosopher-as-policy-maker will quickly recall the possibility of radical under-determination (Quine, 1960). Conversely, if the model is so fully specified, from a god's-eye point of view so that all possible (even incompatible) models are somehow integrated into a meta-model, the situation for making concrete ("Do X") policy decisions becomes exponentially worse because of the complexity (and, most likely, abstruseness) of the model. Ironically, if the super-model were to be "reduced" to a simple, parsimonious and elegant one, its "simplicity" would argue against its applicability to social policy concerns which now come to be viewed as "highly complex" and beyond the "simplicity" of the model.

The ideas above may be further related in a general way with Glymour's (1980) own notion of "bootstrapping." (Note 2) Even if we had a good, formal, and elegantly simple model (theory) of, say, the determinants of income inequality (see Miller, 1987:237-242 for arguments against the bootstrapping issue which, perhaps, *ought* to be the method-of-choice in showing how a causal-modeling framework is relevant to social policy-making). For instance, assume that the State Superintendent of Schools has evidence (in the form of standardized test scores used in the system) that there is a "strong" (e.g., r = .70) positive correlation between test scores and the SES of schools, i.e., SES and Achievement Test scores covary. From a bootstrapping perspective, we might suggest that any of the models, such as the ones noted above, could in conjunction with the evidence, be used to infer an hypothesis something like, "when controlling for IQ the relationship between SES and Achievement Test scores will be substantially reduced." Let us say this hypothesis is subsequently tested and IQ indeed does reduce the relationship between SES and test scores. This goes on in different ways and the theory is increasingly "confirmed"—in at least this sense of the elusive term (Achinstein, 1983). Bootstrapping would seem to be (if indeed it is increasingly supported) a desirable consequence for the policy maker; but in fact it is not.

II.

While desirable, an increasingly well confirmed theory is ordinarily of little pragmatic value for the policy maker. And this is not primarily due to the complexity or theoretical "simplicity" of the theory, nor to a lack of reliability searches, or problems of adequate statistical modeling, but rather to (1) the lack of a "logic" of policy implementation given the nature of the indicators in causal-modeling approaches themselves, (2) the lack of a clear "inference to the best explanation" model in which the issues raised previously—what counts as evidence and what evidence counts—become central, and (3) the lack of acknowledging the power of what we will call Ideological Proclivities in determining the "meaning(s)" of (1) and (2).

The major problem with using social science methods and modeling to *make* social policy is the failure to see that a type of "naturalistic fallacy" is involved, whereby the "is", in this case of *The Bell Curve*, as well as other attempts, is believed capable of being translated into the "ought" of policy making. To see this, some comments on the three points above. First, one of the most difficult issues policy makers confront is the implementation of indicators (as a part of *formulating* and *implementing* a policy) whose "status" may be epistemically sound but ontologically problematic. And, the problem is made worse as, paradoxically, we become more sophisticated in (as Glymour applauds) the use of such techniques as factor analysis which are used to reveal complex "underlying structures" or concepts. Thus, even with a non-problematic construct such as SES, the policy maker is confronted with the issue of how to implement its effects. That is, if SES is correlated with, say, IQ (a *problematic* construct), the policy maker must decide if

(a) the *construct* can be changed or altered in such a way that those who do not have "enough" of it can obtain "more" of it or (b) if new social arrangements have to be constructed wherein those who have "enough" or "too much" of it can be persuaded to "share" it with others (e.g., social policy issues such as desegregation of schools through "bussing") who have "less" of it, or those who have "enough" of it are kept away from those who do not because doing so (anticipating point three, ideology) is justified in some way. Now multiply this one variable case with the type of sophisticated causal modeling envisioned by Glymour and the problems increase accordingly.

The second issue related to the one just mentioned, is that of providing an "inference to the best *policy* decision "based on conventional notions of inference to the best explanation models (generally, Lipton, 1991). What is involved here is essentially the need for "rules" of inference which operate in two directions. The first involves the *creation* of a causal modeling theory which is the result of previous thinking and perhaps partial testing of the various "paths" in the model. The complete model is then tested further and claims about its efficacy *as* a model are put forth. In principle the model (or parts of it) can then be taken as the framework for developing a social policy, which then is tested. Both traditional "deductive" notions of theory use and Glymour's bootstrapping would fall under this approach. Now, even granting the "status" problems of the variables in the model as being capable of testing in some meaningful way, if such testing does take place the conclusions about whether the policy has "worked" are still problematic.

One problem of course is the adequacy of the testing procedures themselves, while another one is how the evidence stands in relation to the model and to the policy that is being evaluated. In another words, can the same evidence simultaneously constitute a best-inference explanation to both? In many cases, the answer to both is no. In the first instance, the way we often attempt to map the presumed *causal* relations of the model to the "real world" are contrived, or at best, constitute a partial mapping. As Glymour correctly points out, the way we "conditionalize" across different samples is crucial in what one's measures do or do not show. But the point we wish to emphasize is that such evidence, both in the "what evidence counts" and "what counts as evidence" senses, is not necessarily the evidence that counts *for* the policy. For example, the finding that SES and School Achievement do vary and are "explained" by IQ, let us say for the entire state of California, is more of a way of "confirming" this assumed relationship in the model than of formulating, implementing or evaluating a policy. That is, because of the nature of policy making as a form of practical argument ("Do X"), even a high correlation of model-specified variables is no guarantee of policy relevance in either the formulation, implementation, or evaluation phases of policy making. Yet such evidence may be strong confirming evidence for the model itself.

On the other hand, what counts as evidence might be given a broad definition for a given policy irrespective of any causal modeling considerations, or perhaps more accurately, *incidentally* of causal-model considerations. For example, the Superintendent of Schools in a state is aware that the "literature" is strongly supportive of a SES-IQ-School Achievement connection, and a similar pattern seems to be the case in her own school system. She formulates a specific policy in which she believes the only way to raise test scores (which are deemed "not acceptable") is to permit no one in teacher training programs with an IQ of less than 115; remove teachers who score below this; and significantly increase the salaries of present and future teachers who are or will be at this level. Additionally, what counts as evidence for the policy (in its formulation and implementation) may be a wide variety of "evidence" including previous empirical and non-empirical studies, reports, anecdotal descriptions, philosophical arguments, and so on. These same, or different, evidence sources may also be used to judge the "success" of the policy in its evaluation phase. In this scenario, which by the way actually often occurs, the inference-to-the-best-*policy* judgment is made on the basis of non-causal model based evidence as instances of the inference to the best explanation (read "explanation" as "successful" policy). While all of these variations on the social policy-causal modeling theme are relevant in varying degrees to the policy making process, the most relevant one in our view is that of implicit or explicit ideological preferences. How this issue works, and how even Glymour is not fully aware of its

power, will be described below. However, before this is addressed, some further brief reflections on the points above may be in order.

Although not addressed by him specifically, we have found some of the recent work by Searle (1988, 1995; also see Review Symposium on Searle, 1998) to be especially useful in situating the social science research-social policy issue. In his continuing analysis of intentionality, Searle (1983, 1998: 99-104) introduces the notion of "conditions of satisfaction," a phrase which refers to the possibilities of judging a large class of intentional states in terms of their propositional contents. Some intentional states such as beliefs and hypotheses can be judged as true or false according to what Searle refers to as their mind-to-world direction of fit. That is, these intentional states are supposed to reflect the way the world is in terms of an independently existing reality. On the other hand, intentional states such as desires and intentions have a different direction of fit: a world-to- mind direction. Here, the issue is one of trying to make the world correspond to what is believed about it (see also, Anscombe, 1959; Austin, 1962).

The interesting parallel to the policy making-social research issue is that the direction-of-fit problem is actually counterintuitive to what one would expect. If we look at Figure 1, Glymour and many social scientists would expect that the increased sophistication of, especially, causal modeling processes will increasingly yield a true mind-to-world fit [i.e., ***A***]. And, indeed, while this may prove to be the case in some ontologically- realist sense, it comes at the increased cost of having to *demonstrate* that the world (in the policy making sense) *is* such, and, hence, we end up with ***C***: trying to fit the world to (again, in terms of policy making) what we believe it should be like on the basis of what it is predicted to be.

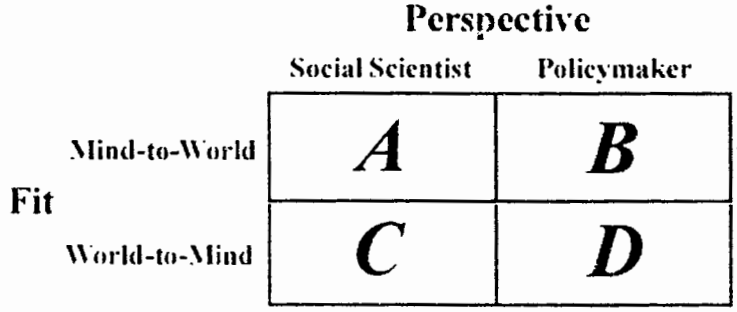


Figure 1

On the other hand, the policy maker want the world to be like (b), but in trying to apply ***A*** to it, she must argue for ***D***. Both groups start out as "realists", in at least a broad ontological sense, but end up as "idealists" in having to reconstruct the desired fit. What results is a type of "reversed intentionality" where beliefs become desires, and desires are fitted into the beliefs—a result where social policy which "fails" is not so much the fault of the model itself but, ironically, of its sophistication. The double irony is that a "simple" model, while "fitting" in both senses, may be rejected by both policy makers and social scientists for this very reason. There is. however, another factor that needs to be addressed and we turn to this now.

III.

Glymour's article opening is entitled, "What went wrong...?" In effect nothing went wrong! By this we mean the critical dimension in trying to understand the relationship between social science causal-modeling and social policy is how the "variable" of ideological preference enters into the equation. The importance of "U" (p. 18) in Glymour's critique is not in some covert empirical variable influencing our model making but rather how model-making is interpreted by way of ideological preferences and proclivities. It is this "variable" that ultimately accounts for our constructions of social reality (Searle, 1995).

The ideological factor is a world-to-mind problem of fit and does, of course,

BEST COPY AVAILABLE

go in both directions—those of social scientists as well as policy makers. Moreover, while the ideological frameworks of those above may be implicit or explicit, there is yet another "level" or group that comes into play here, namely those affected by the policy. What "voice" these individuals obtain from the policies that are usually imposed on them is a function of how well decisions affecting them are understood and the degree of political action garnered for or against the policy. Knowledge of how the ideological factor operates is further complicated by the fact that there are at least two methodological stances one may take to characterize this process—a variety of the mind-to-world problem. These possibilities are given in Figure 2.

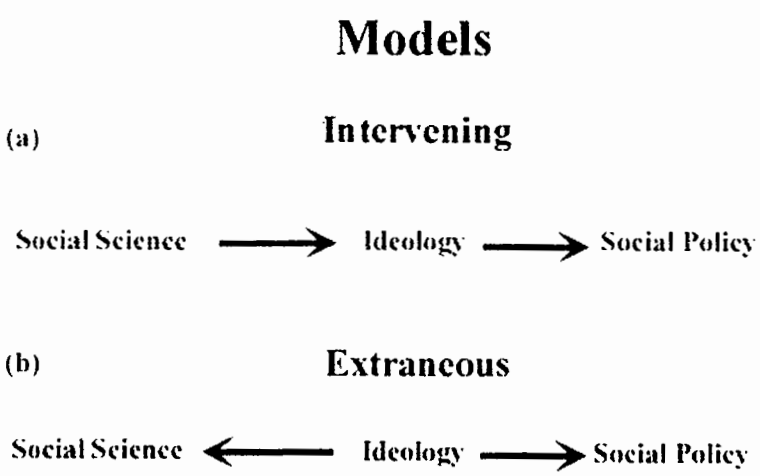


Figure 2

The categories of "intervening" and "extraneous" are meant to be used as they are in social research: an intervening variable as logically "fitting" between an independent and dependent variable, and extraneous, as a variable separately influencing the independent and dependent variables (Nachmias & Nachmias, 1981). For social research and policy, the intervening variable example suggests that an ideological stance is taken (by either social scientist, policy maker, on those directly affected) in such a way that one views it as being compatible with the social policy. That is, the ideology becomes the *justification* for the policy; it is a filter which translates the findings into *acceptable* policy decisions. Thus, if one believes, as in the *Bell Curve*, that there are empirical data which clearly support cognitive differences among racial and ethnic groups, that belief system "intervenes" nicely between the research findings (and approach) and the policy subsequently formulated. In the "extraneous variable" model, the ideological belief system, let us say of the policy maker, is different because it admits of the possibility that the policy maker may *reject* the research findings and yet maintain the efficacy of a particular policy formulation. For instance, if SES differences are correlated with performance on standardized tests, one may reject that they have a hereditary basis and yet may find such results compatible with a "welfare state liberalism" or "educational progressivism" social policy which would support a variety of educational interventions. Moreover, even if the research indicated that racial or ethnic differences remained after controlling for SES, one could still argue that the meaning of SES is "interpreted" differently by different groups. Thus "income", for example, may be "equal" between two groups, but one group utilizes income to invest in "cultural capital" than the other, and it is this factor that makes the difference in test scores; again, an interpretation ideological compatible with the categories above.

We are not suggesting, in some simplistic fashion, that ideological commitments or preferences are always working as "biasing-filters", but only that they are an often overlooked factor in *explaining* how social policies are formulated, implemented and evaluated given social science research findings. Additionally, the ideological proclivities of all directly or indirectly involved in policy making

produce a variety of conflations that are often overlooked in discussions of these issues. Thus, some feminist epistemologists (Tyson, 1998) see their particular agendas, and the social policies flowing from them, as being more (or only) compatible with "qualitative" research methods—what counts as evidence and what evidence counts is ideologically conditioned. In a similar way, entire ideological movements such as "constructivism" (Cobb, 1994, Von Glaserfeld, 1995), while not being overtly hostile to empirical methods, do come down on the side of "ethnographic" approaches.

How the ideological factor is prominent in Glymour's thinking can be made clear when he states (p. 28):

> Sensibly read, much of the data of *The Bell Curve*, as well as other data the book does not report, demands a revived and rational liberal welfare state, but instead the book ends with an incoherent, anti-egalitarian plea for the program of right-wing Republicans.

We now know where Glymour stands ideologically, although it is an open question if his political preferences were "caused" directly by the evidence, *his* reading of it, or irrespective of both. It is probably the middle option of the above. On the same page (p. 28) he berates *The Bell Curve's* assumptions that the decline of the two-parent family is a factor in such things as low school performance. He may be correct in this, but his citing of Murray (1984) to the effect that two parent families are in decline in industrialized societies, does not tell us how or why the Murray evidence conforms to his own causal-modeling structures. Does the evidence in Murray adequately account for all the problems he has cited? If so, some passing mention of it could have been made.

Continuing on (pp. 27-29), Glymour makes a huge leap from the fact that Herrnstein and Murray favor some form of privatized schooling to the "fact" that we will end up with "Ku Klux Klan schools, Aryan Nation Schools... and more schools of ignorance, separation, and hatred will bloom like some evil garden, subsidized by taxes" (p. 29). Before the quote here he uses the phrase, "The consequences are predictable." How poor Modus Ponens is still abused! Where is there any evidence that privatization has or will lead to such outcomes. There are several other instances in the remaining pages (pp. 29-30) of the article where Glymour does seem to be aware of what evidence counts or why it counts. For example,

- He favors neither more decentralization or privatization of schools but rather national standards, testing and funding.
- He favors schools that are always open for children from 1 to 17, that can serve as both centers of learning and safe havens, and says they are the "sane and comparatively economical way to create and sustain a civil society."
- He favors early intervention efforts as worthy and these *can* produce lasting effects (contrary Herrnstein and Murray's conclusions) if "teachers are paid reasonably." He also says not having his vision of infancy to young adulthood quality schooling will result in higher "opportunity costs" than the 100 billion per year cost he estimates.
- He believes "over credentialing" (carried out by colleges and universities) penalizes the potentially positive effects of various compensatory efforts (i.e., affirmative action programs).

Finally, Glymour gives us *his* complete policy vision (p. 30): "Here is an alternative vision, one I claim *better warranted by the phenomena* Herrnstein and Murray report: nationalized, serious, educational standards, tax supported day and night care, a living minimum wage, capital invested in systems that enable almost anyone with reasonable training to do a job well." He then concludes if policies advocated by such conservatives as Gingrich and Gramm are instituted, we will end up pretty much a nation like Honduras!

In brief, the "policy" recommendations Glymour is advocating are not substantiated explicitly by any evidence that would count in their favor. And if there were such evidence, he does not tell us of its adequacy in causal-modeling terms.

Ironically, Glymour's strong support for national standards is very close to what Hirsch (1996) has recently, and somewhat persuasively, argued for—although we would not equate Hirsch with being politically liberal. But the most telling phrase, we believe, in all of this is the emphasized passage above; namely that from the same data presented by Herrnstein and Murray, Glymour draws quite different conclusions—certainly an interesting variant on the under-determination thesis.

Finally, so that we may not be misunderstood, we agree with almost all (except the Honduras slam!) that Glymour is advocating. We are just saying that you can't get there in the way the Glymour thinks you can. The "is" of causal-modeling processes in the social sciences will not translate in the "Do X" of policy making. If Glymour does not believe this, he ought to consider running for a local school board.

Notes

1. One may notice that the policy-making process involves at least these three stages. Each may have an independent or sequential relation to the issue of social science research findings as evidence.
2. Bootstrapping refers to the complexity of trying to adequately determine what evidence and what type of evidence properly applies to the testing of theories. The "bootstrapping" means that the evidence is first connected with the theory and both, then, are used to deduce the hypotheses of the theory. The general issue is *how* theories are to be confirmed. Here, how *do* social science theories result in social policy?

References

Achinstein, P. (Ed.) (1983). *The concept of evidence*. Oxford: Oxford University Press.

Austin, J.L. (1962). *Sense and Sensibilia*. New York: Oxford University Press.

Campbell, D. (1987). "Guidelines for Monitoring the Scientific Competence of Prevention Intervention Centers," *Knowledge* 8(3): 389-430.

Cobb, P. (1994). "Where Is The Mind? Constructivist and Sociocultural Perspectives on Mathematical Development," *Educational Researcher* (October): 13-20.

Denzin, N.K. (1989). *The Research Act: A Theoretical Introduction to Sociological Methods*, 3rd edition. New Jersey: Prentice-Hall.

Fuller, S. (1988). *Social Epistemology*. Bloomington, IN: Indiana University Press.

Glymour, Clark. (1980). *Theory and Evidence*. Princeton: Princeton University Press.

Glymour, Clark. (1998). "What Went Wrong? Reflections on Science by Observation and *The Bell Curve*." *Philosophy of Science* 65(1):1-32.

Green, Thomas. (1980). *Predicting the Behavior of The Educational System*. Syracuse, NY: Syracuse University Press.

Harding, S. (1993). "Rethinking Standpoint Epistemology: "What is Strong Objectivity?" In L. Alcoff and E. Potter (Eds.), *Feminist Epistemologies* (pp. 49-82). New York: Routledge.

Herrnstein, Richard J., and Charles Murray. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Hirsch, E.D., Jr. (1996). *The Schools We Need and Why We Don't Have Them*. New York: Doubleday.

 BEST COPY AVAILABLE

Lakatos, I. (1978). "History of Science and Its Rational Reconstructions." In I. Lakatos, *The Methodology of Scientific Research Programs* (pp. 102-138). Cambridge: Cambridge University Press).

Lipton, Peter. (1991). *Inference to the Best Explanation*. New York: Routledge.

Longino, H. (1993). "Subjects, Power and Knowledge: Description and Prescription in Feminine Philosophies of Science." In L. Alcoff and E. Potter (Eds.), *Feminist Epistemologies* (pp. 101-120). New York: Routledge.

Miller, Richard W. (1987). *Fact and Method: Explanation, Confirmation and Reality in the Natural and Social Sciences*. Princeton: Princeton University Press.

Miller, Steven I., and L. Arthur Safer. (1993). "Evidence, Ethics and Social Policy Dilemmas," *Education Policy Analysis Archives* 1(9) [Entire issue].

Murray, Charles. (1984). *Losing Ground: American Social Policy 1950-1980*. New York: Basic Books.

Quine, W.V.O. (1969). *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Quine, W.V.O. (1960). *Word and Object*. Cambridge, MA: MIT Press.

"Review Symposium on Searle," *Philosopyy of the Social Sciences* (1998) 28 (1 March): 102-121.

Searle, J.R. (1998). *Mind, Language and Society: Philosophy In the Real World*. New York: Basic Books.

Searle, John R. (1995). *The Construction of Social Reality*. New York: Free Press.

Tyson, C.A. (1998). "A Response to 'Coloring Epistemologies': Are Our Qualitative Research Epistemologies Racially Biased?," *Educational Researcher* (December): 21-22.

Van Glaserfeld, E. (1995). *Radical Constructivism: A Way of Knowing and Learning*. Washington, DC: The Falmer Press.

About the Authors

Steven I.Miller
Professor,Department of Leadership,Foundations and Policy Studies
Loyola University, Chicago

Dr.Miller specializes in policy studies as well as the philosophy of social science.He teaches courses in qualitative research methods,philosophy of education and social theory applied to education.

Marcel Fredericks
Professor,Department of Sociology and Anthropology
Loyola University, Chicago

Dr.Fredericks specializes in Medical Sociology,Social Theory and qualitative research methods. He is Director of the Office of Medical Sociology.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http: epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu.. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina -- Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois - UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatistica
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

EPAA • enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 1743 times since January 7, 2000

Education Policy Analysis Archives

Volume 8 Number 4 January 7, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES.**
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Technology and School Reform: A View from Both Sides of the Tracks

Mark Warschauer
America-Mideast Educational & Training Services
Cairo, Egypt

Abstract

A discourse of reform claims that schools must be transformed to take full advantage of computers, while a competing discourse of inequality warns that technology-enhanced reform is taking place only in wealthy schools, dooming poor and minority students to the wrong side of a digital divide. A qualitative study at an elite private school and an impoverished public school explored the relationship between technology, reform, and equality. The reforms introduced at the two schools appeared similar, but underlying differences in resources and expectations served to reinforce patterns by which the two schools channel students into different social futures.

BEST COPY AVAILABLE

As educators cope with the task of integrating information technology into the schools, two main discourses have appeared: the discourse of reform and the discourse of inequality. The discourse of reform suggests that schools must transform themselves in order to make effective use of computers. As an educator in Hawai'i (Note 1) commented,

> The analogy that I have to give is that there is television and there is radio and there is in person. And you would never take a radio program and try to put it on television and expect it to work without modifying for the media. And what we've done is we've taken education curriculum that is a person-to-person curriculum and tried to put it on this medium called the Internet and that doesn't work. And so one of the things we're doing...is trying to work with teachers and with students to say, "What is the appropriate use of the Internet" you know, if it's not to just recreate school as we think school is, how do you do it?

The discourse of reform draws on research from both education (e.g., Cuban, 1986; Sandholtz, Ringstaff, & Dwyer, 1997; Warschauer, 1998, 1999) and industry (e.g., Kling & Zmuidzinas, 1994; Zuboff, 1988) demonstrating that the infusion of new technologies produces little results if underlying relations do not change. The root of the problem is seen in the mismatch between industrial models of schooling and post- industrial organization of society (Cummins & Sayers, 1990; Hodas, 1993; Lemke, 1998); the solution is seen not just in the diffusion of technology in the schools, but rather through creating new models of interactive, autonomous, student-centered learning which allow students to use technology in a process of critical collaborative inquiry (Cummins & Sayers, 1995). As Sandholtz, Ringstaff, and Dwyer (1997) explain, "the benefits of technology integration are best realized when learning is not just the process of transferring facts from one person to another, but when the teacher's goal is to empower students as thinkers and problem solvers" (p. 176).

Though the model of a learner-centered environment is not new, it is believed that technology provides the impetus which will finally allow this dream to be realized. According to one optimistic (but not atypical) prediction, the introduction of more computers in the schools will help bring about eight major shifts in education, including changes from "whole class to small group instruction," "from lecture and recitation to coaching", "from a competitive to a cooperative social structure", and "from all students learning the same things to different students learning different things" (Starr, 1996, n.p.)

While the discourse of reform is hopeful, the discourse of inequality is troubling. From this perspective, increased use of technology in the schools is bound to heighten distinctions among students based on class, language, and race. As a teacher in Hawai'i explained,

> The problem that I see with this change is it's going to create two classes of schools: those schools that can afford the technology and those schools cannot afford the technology. And the rich schools will get richer and we're going to create a greater divergence between our best educated students and our poorest educated students. You cannot change it now. It's out of the box, and it's just going to get bigger and bigger and bigger.

The discourse of inequality draws on its own body of research demonstrating that low-income and minority students either have less access to new technologies or are more likely to use them for rote learning activities rather than for cognitively demanding activity (Market Data Retrieval, 1997; Novak, Hoffman, & Project 2000 Vanderbilt University, 1998; Wenglinsky, 1998) Inequality falls in at least three areas:

- Home access: Wealthy families are seven times as likely as poor families to

BEST COPY AVAILABLE

own a home computer, and white families are more than twice as likely as Black families to own one. The percentage gap in both of these areas increased from 1994-1997. Black and Hispanic families trail (non-Hispanic) white families in computer ownership by a substantial margin even within the same income groups(Novak, et al., 1998).

- School access: More than 78% of public schools in low- poverty communities had Internet access in 1997 compared to less than 59% of public schools in communities with high poverty rates. And public schools with over 50% minority enrollments had an average of 8.4 students per computer, while schools with fewer than 5% minority enrollment had 6.6 computers per student (Market Data Retrieval, 1997).
- Use within schools: African-American students and Hispanic students are more likely to use computers for drill and practice, whereas white and Asian students are more likely to use them for simulations or applications; the same differences appear between poor students and wealthier students (Wenglinsky, 1998).

Putting the discourses of reform and inequality together, two scenarios emerge. The dream scenario is that the information age will help bring about the kinds of educational change that reformers have pushed for all century, with schools becoming sites of critical collaborative inquiry and autonomous constructivist learning as individuals and groups work with new technologies to solve authentic problems under the guidance of a facilitative teacher (see, for example, Lemke, 1998). The nightmare scenario is that this type of educational transformation will occur only in elite private schools and in some upper-middle class suburbs, with the urban and rural poor attending schools that either lack computers or use them in the most traditional and ineffective ways.

The truth of course will probably lie somewhere in-between. Not all wealthy schools will use computers well, and not all poorer schools will use them badly. Nevertheless, there are a number of factors that make the nightmare scenario all too likely, including the depth of already-existing inequality in U.S. schools (Kozol, 1991), the heightening economic polarization in the U.S. in recent years (Mishel, Bernstein, & Schmitt, 1996), and a hundred-year history in which learner-centered reforms have almost always been implemented more readily among privileged students than among poor ones (Cuban, 1993).

But just because one master narrative might ring truer does not mean that it is true. As Bryson and de Castell (1998) point out, the "normativizing" (p. 76) of any one particular account of educational technology as *the* account imposes premature closure on what may be accomplished, thus discounting and restricting the human agency which can actually bring about transformative educational results. Classroom research, and particularly qualitative research which attempts to understand classroom practices from the perspective of the participants, can help bridge the gap between story and reality.

To further explore the relationship between technology, reform, and equality, I carried out a qualitative study in two schools in the state of Hawai'i from 1997-1998. The first, Leina High, is a public school in one of the poorest neighborhoods of O'ahu. The second, Kaunani, is one of the most elite college preparatory schools in the nation. However, this study was not meant to be a simple comparison of "rich good school vs. poor bad school". Both Leina and Kaunani have reputations for excellent use of new technologies, and that is why I selected these two schools for investigation. Through the study, I was hoping to learn more about good uses of new technology in radically different sociocultural circumstances as a way of discovering both the possibilities of reform as well as some of its limitations.

I conducted the study using an interpretive qualitative approach based on classroom observations, interviews, and analysis of texts. I chose the two schools based on interviews and informal discussions with school district administrators and teachers as to their opinions of the best schools in O'ahu in integrating technology and instruction. From the suggestions offered, I chose these two schools based on their distinct socioeconomic populations. I then visited the two schools on approximately a weekly basis over a six-month period in the 1997-1998 school year.

During my visits, I interviewed school administrators, technology coordinators, counselors, department chairs, classroom teachers, and students on their thoughts regarding integration of technology in education. In the majority of cases I tape recorded and transcribed the interviews. In situations where spontaneous discussions arose that were not possible to record, I took notes during or immediately after the discussions. From my discussions with administrators, department chairs, and teachers, I sought the names of teachers who had a reputation for outstanding use of information technology in their teaching. I observed these teachers' classes during my visits to the schools. During these observations, I interacted with students and spoke to them about their experiences. I sometimes helped students while they were working at computers. I took notes during my observations, or, if I was busy helping students, immediately thereafter. Finally, I was provided by teachers and administrators with school reports and documents, and also had access to papers, reports, newsletters, and World Wide Web sites produced by students.

In the remainder of this article, I will share what I learned at these two schools, and then explore the similarities and differences of the reform process.

Leina High

Leina High is a sprawling school of low bungalows in a semi- rural corner of O'ahu. The neighboring community of Leina is one of the few remaining areas on O'ahu with a large percentage of Native Hawaiians. It is also one of the most economically depressed areas in the state. Fewer than 10% of the adults living in the area have completed bachelor's degrees, and per capita income in the area is less than $10,000 per year.

Leina High's character is shaped by that of the neighboring community. Half the students are Native Hawaiians and many of the rest are Samoan and Filipino immigrants. Most qualify for free or reduced- cost lunch programs. Some live in homeless encampments on nearby beaches. Twice as many students are performing below grade level as is the national norm, and only one-sixth as many are performing above grade level. Of those who are able to graduate, the majority seek work, join the military, or study part-time at nearby community colleges. Only 11% of seniors claim that they plan to enter directly into a four-year college or university; no statistics are available on how many actually do. Information in this and the preceding paragraph was provided in a personal interview with the school principal (November 13, 1997) and in school documents which she provided.

To better meet the challenges the school faces, Leina administrators have launched an aggressive reform campaign in recent years. At the centerpiece of the reform plan is a school-to-work plan to better prepare students for success in Hawaii's competitive economy. As part of the planned reforms, students in the future will select a career pathway such as arts and communications, business and management, health services, human services, or natural resources, and then take a number of related courses in that particular pathway while also participating in extra- curricular activities such as visits to local workplaces.

Another important goal of reform at Leina is for better integration of technology into the school's programs. The school's technology committee has laid out an ambitious five-year plan to ensure that the school's infrastructure will allow teachers and students to access a wide variety of technologies, that teachers will have the training to competently integrate technology into their curricula, and that students will have multiple opportunities to become technology literate for their chosen career pathways. Based on these plans, Leina High won an award for having the best technology vision in its school district.

From my visits I could see that implementation of the plan was clearly in its early stages. Though the library had assembled a fair amount of electronic resources, in several visits I never saw more than one or two students using them. Outside the library, computers were relatively scarce, with a total of some 200 computers for Leina's 2200 students. And only a few buildings on campus were wired for the Internet. Susan Bello, the school's educational technology coordinator, explained to me why the wiring was going slowly:

> Due to lack of funds, we had to get volunteers to dig the ditches to lay the cable. So we've had teachers, parents, community members out

> helping dig. But it's been really slow going since the buildings are spread out, and there's only a few inches of soil before you get to solid coral.

Other problems have to do with the existing infrastructure of the buildings. The classrooms, which were built in 1957 and have not been rewired, are unable to handle the power and electricity requirements of modern computer equipment.

In spite of these challenges, a number of teachers at Leina are making efforts to integrate computers into their teaching, and some have had great successes. When speaking to Susan and other teachers and administrators at Leina, I was pointed to three programs which had made strides in this area: the communications program, marine sciences program, and Hawaiian studies program.

Communications

The communications program at Leina dates back tp 1994 when two social studies teachers teamed up to teach an introductory mass media course, focusing on both video production and computer multimedia production. This single course has since expanded into an ambitious program of more than 400 students integrating video production, radio production, Web site design, computer animation, journalism, and yearbook production. The majority of the students in the program take an introduction course co-taught by two teachers and a teaching assistant; students in the course choose to specialize in either video production, radio production, or Web site design. More advanced students take courses in video or multimedia journalism and work to produce video and Web documentaries, multimedia computer animations, and a television news program shown on a local cable station. The program has won numerous state, national, and international awards, including a top price in an international Internet fair for a student-produced World Wide Web site on the Leina Coast, providing multimedia information on the region's history and ecosystem.

During my own visits to the mass media class, students were working on developing Web pages for Leina sports teams and clubs. More advanced students were working independently on more sophisticated Web sites (including a written report and video of a recent surfing competition) and developing complex computer animation. Students were working in a highly independent fashion, with the teacher providing individual or small group support and guidance.

Marine sciences

Another innovate program which has attempted to make use of new computer technologies is in marine sciences. Students in the interdisciplinary marine sciences class engage in collaborative project work related to different aspects of the subject, including growing and selling their own commercial seaweed, and preparing for and participating in sailing voyages around Hawai'i. Computer work centers around producing a newsletter about their projects, based on their own collaborative writing and editing as well as research they conduct on the World Wide Web. Students work in terms to discuss and select stories. They then write an outline and at least three drafts of their article, with it peer reviewed by a student editor. Students receive extra points of their work is published in the newsletter.

The teacher, May Wong, explained how the rationale behind the newsletter:

> My big thing is I want the students to be computer literate. Cause I really feel that's real important in today's word. So I require that all the students come in either before school, after school or during recess to get computer time. And every newsletter that's once a month. They have to have at least three times to use the computer outside of class time. Now, they cannot use the computer during class time and get this. And they can do it for anything. They can come here during English class and say, "Can I type an English paper, and they'll still get computer time?" Cause my big thing, are they comfortable, are they literate on the computer. I don't care if they're doing my work or not.

During my own visits while students were working on the newsletter, they worked to make plans in groups, word process their papers, or seek information from the Web about current events. They were just beginning to use the Web, and their searches were quite cursory, reflecting a quick desire to grab a likely story for the newsletter rather than an informed search, analysis, or critique of online information.

Hawaiian Studies

A third program that is starting to make use of new technologies is Hawaiian Studies, an interdisciplinary program incorporating Hawaiian language and culture, anthropology and history, and physical agriscience. Students in the program also engage in fieldwork, including a weekly visit to a Hawaiian cultural center where they help plant traditional Hawaiian crops such as taro. Use of new technologies in the program has been mostly dedicated to student documentation of the program and its projects. This includes a student- produced newsletter using desk-top publishing and student-produced videos and Web pages on the Hawaiian studies program. The teacher is planning on getting the students involved in an international environmental data-sharing Internet project, but students had not yet begun the project during the time of my visits.

Unlike the Marine sciences program, which has a dozen computers, the Hawaiian studies classroom only has two, one of which is in disrepair. From my visits it appeared that work on the computer was largely controlled by a small group of students who were most comfortable with it. These students help produce the newsletter and Web page and will enter the data in the future Internet project.

Overall, relatively few students were using computers at Leina. Though the library had a new computer laboratory available for classes or individuals, the computers were rarely in use during my visits there. There were no other drop-in laboratories for students at the school, and there were relatively few computers in classroom. A few teachers, as reported above, are starting to integrate computers into the classroom for production of newsletters and informational Web sites, and some of the students in the media program are learning sophisticated multimedia production techniques.

Kaunani School

Many people would consider Kaunani (K-12) School to be the polar opposite of Leina. Kaunani is one of the most expensive private schools in Hawai'i and one of the top-ranked college preparatory schools in the United States. Approximately 97% of its graduates go on directly to four-year colleges and universities, with many going to elite private colleges on the U.S. mainland.

Kaunani has strict admissions policies, requiring a battery of tests for all applicants. In addition to paying some $10,000 per year, potential Kaunani students (even applicants to kindergarten) must test two full years above grade level. The ethnic mix is also quite different at Kaunani than Leina; most Kaunani students are of European, Japanese, or Chinese ancestry, with relatively few Hawaiians, Samoans, or Filipinos.

Though Kaunani already has the reputation as the best school in the state, it is working to improve in a number of areas. According to a recent five-year plan, Kaunani seeks to strengthen its emphasis on critical thinking skills; collaborative and autonomous learning; global education; and ethics, spirituality, and community service.

Like Leina, Kaunani is placing great emphasis on technology, but Kaunani has much greater financial means to implement its plans. While Leina has a technology coordinator for the school, working in the back of the library, Kaunani has an entire department devoted to this effort, with a coordinator, a large staff, and its own multi-room building. Kaunani has been able to wire the entire school (using union labor, not volunteers) and has some 1000 computers available for its 3,700 students (a ration of 3.7 students to computer as compared to 11.0 students to computer at Leina). Most impressive of all though are plans for a new $64 million

BEST COPY AVAILABLE

science and technology center, the construction of which is currently underway. The center will include a large lecture hall with multimedia presentation capacity and one Internet connection for every two seats; numerous laboratory and classrooms fully equipped with networked computers and other technological equipment; a math science resource center for students; a science workshop for hands-on interactive demonstrations and themed exhibits; and high-tech faculty conference rooms and work rooms to promote interdisciplinary teacher collaboration.

Use of computer technologies for teaching, though also at a relatively early stage, is more common at Kaunani. I will examine briefly four programs in which computers are being used: high school English and social studies, high school foreign language, high school science, and elementary school science.

English and Social Studies

English and social studies teachers are trying to use computers to help their students develop literacies in new media as well as to use the online world for academic collaboration and research. One English teacher taught a special online writing course during summer. Students in participated in the course while also engaging in summer travel (one student was on holiday in the Netherlands). Activities included computer-mediated discussions of readings, the posting of student essays on the Web, and the development of an online writing center with links to and reviews of sites related to writing and technology. The same teacher is planning a new regular course which will integrate global education and ethics by having Kaunani students connect with students in other countries to analyze and reflect on ethical themes in world literature.

A social studies and literature teacher are jointly teaching an interdisciplinary course on American studies in which all students have been assigned laptop computers for the school year. Students use the laptops to take notes in class, to write their papers, to discuss topics via e-mail, and to develop and show multimedia presentations on their research.

Foreign Language

Foreign language teachers at Kaunani have been at the forefront of using new technologies for global interaction and education. For example, several of the Japanese teachers at the school have integrated e-mail and the Internet into their teaching. One Japanese teacher is having her students produce a Japanese-language radio program for a local station. To help prepare the program, the students are working in teams to survey Japanese correspondents via e-mail. They then, using both e-mail and live video-conferencing, further discuss with their Japanese correspondents the topics and content of their radio scripts. The teacher is planning a project next year where students will select several Japanese characters on display at a local cultural center. They will then research the historical meaning of characters and combine that with current interpretations based on e-mail interviews with students in Japan. The goal is to compare the language and culture of contemporary Japanese society with that of the Japanese who came to Hawai'i 100 years ago.

Science

Computers are being used extensively in honors physics and advanced placement (A.P.) biology programs. (Approximately half of Kaunani students take honors and/or A.P. classes). In physics class, students perform computer-based simulations of motion experiments one day, and then the next day they perform the actual experiments in laboratories of sophisticated equipment (e.g., frictionless air tubes). The computer-based simulations allow them to try out a broader range of hypotheses related to motion and collision of multiple objects traveling in multiple directions at multiple velocities. In biology class, the students use special hand-held devices for probing the temperature, acidity, absorption spectra and other features of plant life in the classroom and in nearby ponds. Students then download data from these devices to personal computers, where special software allows them to graph and compare data in order in order to interpret it.

Elementary School Science

Use of computers for science begins at elementary school at Kaunani. Fifth grade students learn to write computer programs for a Logo-Lego system. Unlike earlier Logo systems, in which these programs were used to manipulate a drawing of a turtle on the screen, this new Logo-Lego system can be physically connected by wires to the students' own constructions made up of plastic Lego building blocks. Students thus first build small cars and traffic lights, and the use the computer programs they write to make the cars go and stop per the change of traffic signal.

Overall, there was a substantial presence of computers and computing at Kaunani. There were several large wired computer laboratories available to classes or individual students on a drop-in basis, and the use of these labs was quite heavy. In the labs and on their home computers, students frequently searched the Internet to get information for school papers. Several teachers had begun to integrate computers into their academic programs in areas related to writing, foreign language collaboration, and scientific research and analysis.

Common Elements of Reform

As seen from these above examples, there are many common elements of successful classroom use of technology which are evident at both Leina and Kaunani. These elements, which I will briefly discuss, include interdisciplinary and team teaching, collaborative/apprenticeship learning, flexible scheduling, and support for teacher initiative and involvement.

Interdisciplinary and Team Teaching

Almost all the cases of excellent technology use that I observed in these classes are attempting in some way to break out of traditional classroom disciplines. In some cases this involves an individual teacher designing a project with many disciplines in mind, such as the elementary school teacher planning a Logo-Lego project which incorporates math, physics, computer programming, and engineering concepts for elementary school students; or a Japanese teacher planning a lesson which incorporates language, culture, and history. In other cases, the courses themselves are interdisciplinary by design, such as the marine sciences course at Kaunani. And in many cases, teachers have found ways to form partnership or team teaching relationships with those from other disciplines. For example, the computer component of A.P. biology was set up through cooperation with a mathematics teacher; in the future, the two teachers plan to establish a paired A.P. biology and A.P. calculus course. The video production and computer production teachers at Leina have joined for a combined Mass Media course, and they coordinate together with the teachers in business, journalism, and yearbook production. Similarly, these interdisciplinary programs coordinate with each other at a meta level, with students from the Hawaiian studies or Marine sciences programs who are also in the mass media program working on projects which combine their interests (e.g., a Web site or video about marine sciences).

Collaborative Apprenticeship Learning

In addition to breaking down traditional boundaries among disciplines and among teachers, successful technology-enhanced programs at both Leina and Kaunani are also breaking down traditional teacher- student roles. Virtually all the computer projects I saw at either school were based on social constructivist principles of learning, with students working in groups to define and carry out projects. For example, in the Web production program at Kaunani was organized more like a semester- long workshop than a traditional teacher-centered class. Students came and went immediately to their computers, which were spread out in clusters around the class. The teacher occasionally offered explicit instruction to the whole class, but students paid (or didn't pay) attention based on their own particular interest in the topic of discussion. Students worked in teams and were encouraged to

pursue areas of their own interest, with some students focusing on researching and writing texts, others focused on advanced Web production techniques, and others focused on artistic areas such as multimedia animation. Students sought help as they needed it from each other or the teacher. Grades in the course were based either on the students' or on occasional performance assessments, in which students were required to create Web pages with certain features. The teacher acted as a coach and guide, bringing in new instructional videos and books for students to use, giving them individual or small group guidance on their work, letting them know (and helping them prepare for) upcoming competitions, inviting students to accompany him to either attend advanced workshops or give basic and intermediate workshops to others, and providing students moral support and encouragement. For example, he would frequently remind them of the successful national awards won by previous students, and would also tell them that Leina High is "the Kaunani of Web design," just like people might say that their city is the Paris of Asia, or Africa, or the Middle East. In essence the teacher is a master Web page designer who is working hard to continuously upgrade his knowledge of the most sophisticated new technologies, ranging from "VRML" (Virtual Reality Modeling Language) to "Claymation" (clay animation). Students are his apprentices; they begin by working under his guidance on simple projects such as the design of a Web page about a sports team at Kaunani. Those who show a serious interest continue to more substantial efforts, such as the previously mentioned virtual tour of the Leina coast.

The biology course at Kaunani indicated a similar collaborative apprenticeship approach. In this course, students were apprenticing to be biologists rather than Web designers. Though portions of the course were devoted to lecture, other portions were devoted to engagement in the practice of biological research using computer technology as a tool in the same way a scientist might. Students worked in groups to carry out and interpret their experiences, achieving results, which according to the teacher, were potentially publishable in scientific journals. The teacher wandered around the classroom and guided the students in everything from the gathering of data to its interpretation to the formation of overall conclusions.

Flexible Scheduling

At both schools, an interdisciplinary approach and collaborative apprenticeship learning were facilitated by flexible scheduling—of a somewhat simple form at Leina, and a more complex form at Kaunani.

At Leina, Mondays, Tuesdays, and Fridays were organized according to a traditional six-period high school program. However, Wednesdays and Thursdays were based on double periods, with students having three two-hour classes on Wednesday (first, third, and fifth periods) and three two-hour classes on Thursday (second, fourth, and sixth periods). These double-periods were essential for carrying out the kind of in-depth project that apprenticeship learning often involves. Students in video production wandered campus to carry out filming and interviewing. Students in marine sciences tended to their seaweed. Students in Hawaiian Studies combined two two-hour slots and worked at the nearby Hawaiian cultural center.

At Kaunani, the reorganization of scheduling has been more dramatic. School is organized according to six-day cycles, rather than five-day weeks (e.g., cycle 1 is M-T-W-Th-F-M, cycle 2 is T-W-Th-F-M-T). Teachers are assigned a certain number of contact hours per day, which they can divide up however they please. For example, English teachers are assigned 85 student-contact hours a day. They can teach, if they want, five one-hour classes of 17 students, or one one-hour lecture of 85 students, or some combination. Most teachers put together a schedule which includes a combination of larger lectures, smaller discussion groups, and possibly small but lengthier laboratory sessions. This approach, while obviously much more complex and difficult to set up, is even more advantageous than the Leina setting for implementing technology-enhanced project work, as teachers can create the combination of laboratory, discussion, lecture, or other sessions that are most appropriate for the type of course they are teaching. For example, the American Studies course met twice per cycle for one-hour classes of 27 students, twice per cycle in one-hour discussion seminars of 13 or 14 students, and once per

cycle for a two-hour sessions of 60 students for lectures or films. The biology and physics classes both combine longer sessions of smaller groups in the labs and computer rooms, with larger shorter lectures.

Teacher Initiative and Involvement

As Larry Cuban (1986) has documented, new and supposedly revolutionary technologies have been imposed from above for a century, with poor results. Central district and school administrators have a history of urging or demanding use of radio, television, film, and now computers, with little involvement from classroom teachers in making school-wide decisions about technological implementation.

Both Leina High and Kaunani School have avoided this problem. On the contrary, both schools seem to be exemplary in involving teachers in shaping the direction of the school, and in particular encouraging their initiatives regarding technology.

Leina High is a designated School Community Based Management (SCBM) site and thus receives extra support from the Hawai'i Department of Education for teacher and community involvement in decision-making, including the potential of receiving special waiver days (in which students are dismissed from school for teacher planning). Leina has used these days to the maximum over the last three years to involve teachers in developing the five-year plan for the school. Teachers I spoke with were quite familiar with the details of the plan, and couched their own teaching goals and visions in accord with the plan's language.

Teachers at Leina have also been quite involved in shaping policies regarding technology. The technology plan has arisen through grassroots teacher involvement, and teachers have been given release time to work out its implementation. In addition, grassroots teacher initiatives are respected and appreciated, especially when they involve crossing disciplinary boundaries. As the principal told me,

> We've been encouraging teachers to informally hook up with each other. Do interdisciplinary projects. Do things together. Get out of your own four walls or your own content area and try doing something different with a teacher from another department. So we've been encouraging this kind of behavior among the staff...And so technology, with [the media] program, they've been deliberately expanding and trying to encompass more areas into what they do. And with the Hawaiian Studies program the technology really just supports what they're doing in terms of having the kids learn about agriculture. From agriculture all the way to architecture and archaeology. And then with the marine sciences program also they're doing a , they're now integrating what they're doing in marine sciences with social studies. History as well as modern day Hawaii. So that's the direction. The direction is toward integration and towards creating career pathways and so we expect to see more people jumping in and doing that kind of thing.

Recently the teachers in the communications program were pulled out of their classes for four straight days to plan the future of their program, and the role of technology within it, while substitute teachers taught their classes. The media teacher complimented the role of the principal:

> I credit her the most as far as our successes. She is real action oriented. She's visionary. And she's very, very supportive of what we do. She's been very supportive. She's given us the leeway. And I think as a result of her support we've been successful. We've been able to try and move things. Cause without a principal that says, sure, try a recording studio, or, sure, try a radio station, sure, you want a digital camera; - I needed money to get a digital camera - she doesn't really understand what it is but she understands that we want to stay on top of the new technology.

The support for teacher involvement and initiative at Kaunani is equally impressive. The school just thoroughly reviewed its policies and goals as part of a review by the Western Association of Schools and Colleges. All faculty and staff participated in meetings to help clarify the school's purpose, as well as hundreds of students, parents, and alumni. Teachers are also given substantial support to integrate new technologies, including release time from the college for innovative practices, special funds for purchase of equipment, and support for taking of classes. The social studies teacher making use of laptop computers is doing so with a school grant (both for equipment and release time) and is also taking a course on distance education with funding from the school. And a special interdisciplinary committee of the faculty is meeting on a regular basis to discuss uses of the Internet and distance education, again with release time from the school for these purposes. Teachers who engage in such projects are also expected to produce reports for the rest of the faculty based on their experiences.

Different Resources, Different Expectations

As seen above, there were many substantial areas of overlap between the reform process in these two diverse schools. At the same time, though, there are also important areas of difference. I will group them into two general areas, related to resources and expectations.

Resources

When looking at resources at the two schools, it is important to start from the differential access to technology that students have at home. At Kaunani, in one social studies class I surveyed every single student had a home computer, and the majority had 2, 3 or 4 computers at home with one or more Internet accounts. My informal polling of students indicated that it was rare of find a student at Leina who had home access to a modern computer—most either lacked a computer or had part-time access to a very old machine. As a librarian at Leina explained to me,

> We have to provide technology because they don't have it at home. The only exposure to technology they have is at school. Most don't even have push-button phones, or indeed any workable phone line at all. Often when we call their phones are out of order or disconnected. People are struggling at home to pay their phone bills.

Unfortunately, this differential access between Kaunani and Leina students is further multiplied at school. Classes at Leina are held in dilapidated bungalows with poor infrastructure to support modern technologies. The Hawaiian Studies class, for example, has a dial-up connection to the Internet as the building lacks the electrical facilities to support a hard-wired connection. Leina's Web production teacher—one of the most honored teachers in the state, with awards of recognition from the Mayor, Governor, House of Representatives, and State Senate—has only eight computers in his classroom, so students must double or triple up on a machine. In contrast, Kaunani already has a fully wired school and a high computer-student ratio, and it is in the midst of building one of the most modern and well-equipped school science and technology centers in the country. Dozens of high-paid construction workers labor away day-by- day at Kaunani, while technological improvements Leina depends in part on the sweat of unpaid volunteers.

Differences extend to the support given for teachers as well. Leina High does its best with limited resources, but it has only so much to offer. Teachers who want extra funding have to write grant proposals on their own time. Kaunani has its own financial support staff on campus which seeks grants for the school; the money is then made available to teachers for the asking. And while teachers at Leina teach six classes a day of up to 35 students, Kaunani teachers face an average of 85-100 students a day (based on 17-20 students per period for five periods) in a schedule totally at their own control. Smaller class sizes and fewer classes mean that teachers

can spend more time preparing for their classes, including thinking about how to integrate technology, and can devote more personal attention to individual students as they uses computers in the classroom.

Expectations

The second major difference has to do with the goals, visions, and expectations of the schools. While the processes of reform are in many ways similar in the two schools—include interdisciplinary and team teaching, collaborative/apprenticeship learning, flexible scheduling, and high levels of teacher initiative and involvement—the goals toward which the reform is geared differ dramatically in the two schools.

Leina High's reform process, including the uses of technology, is geared toward better preparing students for the workforce. Teachers work to help students develop the types of technological literacy and human relations skills that might be needed in the workplace, without great emphasis on academic content. To illustrate how this takes place, I will briefly examine two programs, the communications program and the marine sciences program. The strong majority of students in the communications program take either radio production or video production. In both of these classes, most students focus principally on learning technical skills, such as how to videotape or how to edit a radio program. Likewise, for the minority of students who take Web production, most of the work is focused on the technical aspects of Web page production. (In contrast, at Kaunani, students also are involved in producing the school's Web pages, but they do this as part-time paid work, rather than as part of their academic course load).

In the marine sciences program, much of the work the students do has little relationship to science. The teacher spends a good deal of time with the students discussing the meaning of inspirational quotations, or reading stories from the popular book, *Chicken Soup for the Soul*, and even had students write their own stories for a classroom version of the book (Portuguese Soup for the Soul). Work at the computers serves a similar communitarian purpose; the newsletter the students produce has little hard scientific information in it and instead focuses on the students personal experiences (e.g., "Students sail on the voyage of a lifetime," "Dear Journal". (The Hawaiian studies newsletter also featured similar personal stories, introducing the teachers and students, discussing attendance policy, and announcing a calendar of upcoming events.)

Both the communications teacher and the marine sciences teacher both explained to me their hidden curriculum—the purpose behind what they do in the classroom. Carla, the communications teacher explained that:

> We have to make it relevant, because when they leave us, we want to be able to say that they not only, you know maybe as we're teaching teamwork, cooperation, respect for themselves and others. We just so happen to be teaching that through video production. Through computers. Through radio. And when they leave that, when they leave us we want them to learn how important it is to have teamwork, cooperation, and respect for themselves and others and property. Because no matter what they do, right, whether it's in a job or a relationship, they have to have that. And hopefully at least that they're taking with them. And they have some kind of a skill that's going to be able to get them a job. Whether it be media or anything, you never know what they're going to grab on.

The marine sciences teacher explained to me her very similar approach, also stressing respect and cooperation:.

> There's four things that I expect the students to learn.... Number one I expect them to learn respect. How to be respectful. Number two, responsibility. Number three, to work cooperatively in a team situation. And number four is to be seekers of information. If I can teach you those

> four things by the time you graduate I will feel like I've done my job. And I said, you notice, there's nothing to do with science. 'Cause to me the science portion will come as a part of being responsible and useful seekers of information....As far as I'm concerned they cannot learn the science and they cannot learn the material if they're doing all of the above.

She later explained to me why she didn't feel it was important to emphasize scientific concepts in her marine sciences program:

> I've been doing this for about six, seven years now, seven years of so. And the really interesting thing is about two or three years ago this whole school-to-work thing came out. And they went to big companies, and they asked these employers, they said, when our students graduate from high school what do you want them to know? And the employers all came up and said, We don't care what they studied, we want a student who's respectful, who's responsible, who can work together with other people and want to learn and want to learn, we can train them. We don't care. We don't need them to be honors students and all that. We can train them on the job. Give us kids who know how to be respectful, responsible, team players. And so it's right in line with what we've been doing and I feel really good about that...cause this is what employers want.

For both teachers, the central element is not the content, but the attitudes that students learn from the class. And the attitudes which are most important are the respect and cooperation—how to be a good team member—which employers value.

Both of these teachers are trying to further strengthen the school-to- work component of the program in another way too, by integrating a strong business component into their teaching. The marine sciences teacher is seeking to develop a team-teaching relationship with a business teacher, so that the students can better market their seaweed. She also hopes to have students track the progress of local stocks on the World Wide Web as part of their education for future marketing, sales, and investment.

The communications program has already brought a business teacher to help teach sales, marketing, and accounting. As she explained:

> We want to be looked at as a production company. So say, when you come into this class you're not coming to class, you're coming to work. Each of you have a job to do. And we want to start, because this type of class takes so much money to, repair and maintenance, and, you know we want to get air conditioners and this and that. We need to start raising our own money. So we want to start selling our services. For the last four years we've been doing it for free. And we still want to do it for free to a certain extent. Especially to the community as community service. But, we also need to start generating our own income. So we want to start selling videos. If somebody wants, let's say, a wedding video, we want to be able to do that. Somebody wants a little documentary on their project, or, for the radio people want, they want a little 15 second radio, commercial on their business, or you know if they want to come and record themselves we want to be able to generate funding through that. Our kids have done numerous Web sites. And you know that costs money if you go on the outside. But we've been doing it for free. And we can make money doing them. We won't charge them an arm and a leg, but we'll charge them something. And the kids need to know how do you go about doing that. How do you market? How do you sell your product? What do you sell. What do you charge? You know, business fundamentals.

I did not have a chance to interview or observe all the teachers in the school. Yet these teachers have been identified by colleagues, students, and the administration as the most exemplary that Leina has to offer. And from my observations of their classes and interviews with students in their classes, that is not surprising. They are both highly engaged, committed professionals who devote untold hours and boundless energy to providing new opportunities for their students. And their energy is devoted to reshaping students' attitudes, and providing the skills and acumen, to better compete in the job market, with the use of new technologies serving these purposes as well.

In some cases, the Leina teachers themselves are anxious to raise standards but find the challenge overwhelming. Last semester, for example, May was teaching a combined beginning-advanced class with some 45 students. Her original plan was to have the 30 beginning students work on introductory projects, while the 15 advanced students (all in their second or third year of video production) worked on more challenging news programs and documentaries. But coordinating different levels of so many students in the same semester became overwhelming, especially with limited amounts of video equipment. Most of her time and energy thus was devoted to the beginning students, and she was not able to get the advanced students working on the projects until much later in the school year.

At Kaunani, expectations, policies, and teaching and learning conditions differ dramatically. The school is designed to produce the academic and professional leaders of tomorrow. Discussions of school reform are framed by the goal of helping students meet the requirements and expectations of the most prestigious universities. As for technology, teachers seek to use it for academic rather than communitarian purposes (for an interesting discussion of the differential impact of a communitarian climate and an academic climate, see Phillips, 1997) This is seen, for example, in the Japanese classrooms, where students' use long-distance exchange for analysis of complex cultural and linguistic issues. Or in the biology classes, where students use computers to perform the same types of analysis and research that a university research might perform, rather than to produce a newsletter (and where the teacher is teaming with a calculus teacher, not a business teacher). The biology teacher at Kaunani explained his own rationale for using computers, which is quite different from the perspective of the science teacher at Leina:

> We've been working over the years on our biology program, particularly our advanced biology program, to give students the type of experience that they need to prepare them for college work...I had been a research scientist at Berkeley and Stanford as a graduate student. So I have a very strong background in research, which I loved. And I try to share that love of research with my students. And since I was pretty much lab oriented and biochemistry oriented I did what I knew and tried to implement those kinds of experiments. When the advanced placement biology program became formalized they gave us a lab. And at first that was very frustrating but we gradually were able to do all the labs they asked us to do and still implement our own program and add to it and to the best of our ability we maintained a strong program that we feel prepares students for college level work. And it became obvious as we, over the last ten years, the computers were becoming one of the most important scientific tools available. And, so we wanted to implement the computers into the program. And the way we did this was we brought two computers of own, our own personal computers from home, we purchased the software ourselves and we demonstrated to the administration of this school that we could use the computers in the classroom in a productive and effective way. Once we'd proved that we could use them they were willing to fund it. And so we had cooperation from the parent/faculty association and the administration. And they funded our computer program. And we realized that this was an important scientific direction for our students to go.

Perhaps most interesting to me was my observation of the fifth-grade students work on the Logo-Lego project and my discussions with the teacher of that project. Similarly to the teachers at Leina, this teacher told me his own "hidden curriculum" behind the teaching.

> I'm teaching a lot of other things besides math and science. Probably the most important think is project management, making complex things happen in a certain amount of time. I'll say, O.K., based on these commands that we know how they control the machines, now do this in the next hour. And they have to work in teams. Or, I'll make an extension based on what they know, and then there are multiple solutions, so there's all different ways to do it. But they have to do it within an hour. Getting to operate under those conditions I think that's important.

It was noteworthy that both Leina and Kaunani teachers stressed the importance of participating in teams. But whereas Leina students were expected to learn things such as responsibility to the group and respect for other, Kaunani students—even those as young as in the fifth grade—were expected to learn how to manage complex systems.

Conclusions

This study examined the process of school reform and technology implementation in two diverse schools, an elite private school and a school in a low socio-economic status neighborhood. Interestingly, the process of reform in the two schools showed a good deal of similarity. Both schools encourage interdisciplinary and team teaching, collaborative/apprenticeship learning, flexible scheduling, and active teacher initiative and involvement in shaping the use of new technologies. In some ways, Kaunani's reforms in these areas were more dramatic, as seen in the total modular scheduling at Kaunani as compared to the double-period days at Leina. But Leina nevertheless implemented similar reforms within the school's more limited means. The study thus provides a positive example of how a low-SES school can engage in the types of reform that are seen as necessary to make effective use of technology (see for example Sandholtz, et al., 1997; Means, 1998) and which are believed to rarely occur outside of elite private schools or public schools in well-to-do suburbs (Cuban, 1993).

But in spite of the above, it is also the case that Kaunani continues to socialize students into academia, and Leina socializes students into the workforce, a difference made explicit by the emphasis on school-to-work at Leina. And the students from Leina, who enter high school far behind their Kaunani counterparts in technological literacy due in part to limited access to home computers, are likely to fall much further behind from the respective high school education students receive in the two schools. Kaunani students have more school computers at their access and are more likely to use them for scholarly experimentation and research than are students at Leina.

Studies have shown that students in low SES neighborhood schools frequently used computers for exercises and drills in basic skills (e.g., Wenglinsky, 1998). That is not what I observed here. Perhaps the era of "drill and kill" may fade away, at least in secondary schools, to be replaced in low -SES schools by the development of attractive but limited- content Web pages or newsletters.

Leina's best-regarded teachers are building award-winning programs which are inspiring students and actively engaging them in the learning process. They are turning many lives around, and their best students are winning national and international awards for their media projects. These teachers' hard work has indeed made Leina "the Kaunani of Web page design". The types of collaborative apprenticeship project-based teaching they are engaged in, together with other reforms such as team teaching and flexible scheduling, have contributed to these positive results, and are worthy of emulation by other schools.

But Kaunani school itself remains "the Kaunani" of mathematics, physics,

biology, history, literature, and foreign languages. And that in the end has a profound effect on the differing life opportunities for Leina and Kaunani students. To seriously diminish that difference, it will take more than team teaching or flexible scheduling or collaborative learning, but rather a challenge to the unequal allocation of resources and expectations to Leina High and Kaunani School and to the thousands of other Leinas and Kaunanis across the country.

In analyzing integration of technology into instruction, Cuban (1993) proclaimed that "Computer meets classroom: Classroom wins" (p. 185). The implication was that the traditional patterns of classroom organization are proving impermeable to change, even with the introduction of large numbers of computers into schools. This study suggests that even in those cases where the computer "beats" the classroom, it doesn't necessarily beat the system. Computers, Internet use, re-arranged classrooms, flexible schedules, and interactive instruction can all leave intact or even reinforce patterns by which schools channels students into different social futures.

This study thus provides support for both the discourse of reform and the discourse of inequality. Schools of diverse socio-economic circumstances can carry out the types of technology-enhanced reform that make education more interactive. But these reforms take place in a social context that will likely make education more unequal.

Note

The names of schools, administrators, teachers and students have all been either changed or deleted in this study.

References

Bryson, M., & Castell, d. S. (1998). Telling tales out of school: Modernist, critical, and postmodern "true stories" about educational computing. In H. Bromley & M. W. Apple (Eds.), *Education/technology/power* . Albany: State University of New York Press.

Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920*. New York: Teachers College Press.

Cuban, L. (1993). Computer meets classroom: Classroom wins. *Teachers College Record*, 95(2), 185-210.

Cuban, L. (1993). *How teachers taught: Constancy and change in American classrooms 1890-1980* (2nd ed.). New York: Longman.

Cummins, J., & Sayers, D. (1990). Education 2001: Learning networks and educational reform. *Computers in the Schools, 7*(1/2), 1-29.

Cummins, J., & Sayers, D. (1995). *Brave new schools: Challenging cultural illiteracy through global learning networks*. New York: St. Martin's Press.

Erickson, F. (1986). Qualitative methods in research on teaching. In M. Wittrock (Ed.), *Handbook of research on teaching* (pp. 119- 161). New York: Macmillan.

Hodas, S. (1993). Technology refusal and the organizational culture of schools. *Education Policy Analysis Archives, 1*(10). Retrieved November 29, 1997 from the World Wide Web: http://olam.ed.asu.edu/epaa/v1n10.html

Kling, R., & Zmuidzinas, M. (1994). *Technology, ideology and social transformation: The case of computerization and work organization*. Revue International de Sociologie, 2, 28-56.

Kozol, J. (1991). *Savage inequalities*. New York: Harper Collins.

Lemke, J. L. (1998). *Metamedia literacy: Transforming meanings and media.* In D. Reinking, M. McKenna, L. Labbo, & R. D. Kieffer (Eds.), Handbook of Literacy and Technology: Transformations in a Post-Typographic World (pp. 283-301). Hillsdale, NJ: Erlbaum.

Market Data Retrieval (1997). *Technology in education 1997: A comprehensive report on the state of technology in the K-12 market.* Shelton, CT: Market Data Retrieval, Inc.

Means, B. (1998, April). *Models and prospects for bringing technology-suported education reform to scale.* Paper presented at American Educational Research Association Annual Conference, San Diego.

Mishel, L., Bernstein, J., & Schmitt, J. (1996). *The State of Working America, 1996-97.* Washington, DC: Economic Policy Institute.

Novak, T. P., Hoffman, D. L., & Project 2000 Vanderbuilt University (1998). *Bridging the digital divide: The impact of race on computer access and Internet use* [Online Article]. Retrieved September 27, 1998 from the World Wide Web: http://www2000.ogsm.vanderbilt.edu/papers/race/science.html

Phillips, M. (1997). What makes schools effective? A comparison of the relationships of communitarian climate and academic climate to mathematics achievement and attendance during middle school. *American Educational Research Journal, 34*(4), 633-662.

Sandholtz, J. H., Ringstaff, C., & Dwyer, D. C. (1997). *Teaching with technology: Creating student-centered classrooms.* New York: Teachers College Press.

Starr, P. (1996). Computing our way to educational reform. *The American Prospect*(27), 50-60. Retrieved December 10, 1999 from the World Wide Web: http://epn.org/prospect/27/27star.html

Warschauer, M. (1998). Online learning in sociocultural context. *Anthropology & Education Quarterly, 29*(1), 68-88.

Warschauer, M. (1999). *Electronic literacies: Language, culture, and power in online education.* Mahwah, NJ: Lawrence Erlbaum Associates.

Wenglinsky, H. (1998). *Does it compute?: The relationship between educational technology and student achievement in mathematics.* Princeton, New Jersey: Educational Testing Service.

Zuboff, S. (1988). *In the age of the smart machine: The future of work and power.* New York: Basic Books.

About the Author

Mark Warschauer
Mark Warschauer (http://www.lll.hawaii.edu/markw), formerly a faculty researcher at the University of Hawai'i, is currently director of educational technology on a large US-funded development project in Egypt. He is the editor of *Language Learning & Technology* journal and the author of numerous works on technology in education, including, most recently, *Electronic Literacies: Language, Culture, and Power in Online Education* (Lawrence Erlbaum, 1999).

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy AnalysisArchives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (480-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

BEST COPY AVAILABLE

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

BEST COPY AVAILABLE

EPAA • enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 368 times since January 10, 2000

Education Policy Analysis Archives

Volume 8 Number 5 January 10, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Education Policy in Portugal: Changes and Perspectives

Jesus Maria Sousa
Universidade da Madeira
Portugal

Abstract
The Revolution of 25 April 1974 in Portugal put an end to a forty-eight year old dictatorship, opening the country to democracy. The purpose of this article is to describe education reform from the standpoint of a country that experienced a major political transition and had to start from the very beginning to devise an education policy. Rather than merely describing the organization of the Portuguese education system, I present a condensed analysis of Portuguese education policy, as I view it, making use of indicators of the nature of an education system proposed by D'Hainaut (1980).

BEST COPY AVAILABLE

The Revolution of 25 April 1974

Portugal is a small country with a total area of 91,985 square kilometers located in the extreme west of Europe and with two archipelagos in the Atlantic Ocean, Azores and Madeira, which are politically autonomous regions. The resident population is 9.853 million; only one language is spoken throughout the country, Portuguese. The Revolution of 25 April 1974, in Portugal put an end to a forty-eight year old dictatorship, dominated by a political police force, the so-called PIDE. After Salazar's death in 1968, the new prime- minister Marcello Caetano attempted the gradual opening up of the regime (the Marcellist Spring), but the dictatorship had grown so corrupt that a revolution broke out in the early morning hours of 25 April 1974. Zeca Afonso's banned protest song "Grandola, Vila Morena" was broadcast on Portuguese radio as a secret signal to a group of rebel officers to move against the regime. The army, tired of the bloody and useless war in remote colonies in Africa, led the Revolution. Most of the leading military officers of MFA (Armed Forces Movement) were involved in left-wing activities. The Revolution was quite peaceful. It was called the Carnations Revolution because carnations were in bloom at that time of the year and were placed in the guns of the soldiers. The forces of the "ancien régime" surrendered with little resistance.

The national euphoria did not last long. In spite of the coherent "three D's" political program, which promised Democracy, Decolonization and Development, the MFA was not a unified body. Some officers wanted a liberal democratic state while others sought radical social transformations. In the subsequent two-year period, there were six provisional governments, two presidents, a failed right-wing coup attempt, a failed left-wing coup attempt, three elections, seizures of land and housing, bombings and strikes, while the country was flooded by millions of Portuguese settlers escaping from ex-colonies at war. Yet, surprisingly and contrary to the expectations of most observers, national political leaders committed to a democratic system laid down by the Constitution of the Portuguese Republic were approved by the Constituent Assembly on 2 April 1976.

According to the Constitution, Portugal is a democratic state based on the rule of law, the sovereignty of people, the pluralism of democratic expression and respect for fundamental rights and freedoms for all citizens. This democratic political organisation is based upon the principle of separation and interdependence of the sovereign bodies: The President of the Republic, the Assembly of the Republic, the Government and the Courts.

Education Policy in Portugal

Having just celebrated the silver anniversary of democracy in Portugal, I wish to share some information from the standpoint of a country that experienced a political transition and had to start from the very beginning to articulate an education policy. The Constitution approved in 1976 proclaimed that everyone had the right to education based on a foundation of equal opportunities to both access to and success at school. Being responsible for the democratization of education, the state was not entitled to orient education and culture to any particular philosophical, aesthetic, political or religious ideology. Education was also expected to minimize economic, social and cultural differences, stimulate democratic participation in a free society and promote mutual understanding, tolerance and a spirit of community. These general principles aimed at creating a "new" education were eagerly embraced by a changing society. Nevertheless, the Education System Act, which established the general framework for the reorganization of the Portuguese education system, had to wait twelve years to be discussed in the Assembly of the Republic. The Law (Law 46/86) developing those principles written on the Constitution hasn't arrived so quickly as we could expect. However, it was the result of a large participation of the political parties. Five parties presented each a project of the Law, having all been voted favorably in general by all parliamentary groups. After a long debate of 175 hours along 30 meetings within the specialized committee, our Magna Carta of Education got an expressive approval in the Plenary of the Assembly of the Republic.

Considering that education policy is the translation of a series of political intentions, our Education System Act is one of the most important sources for this

analysis. Where could one find a more explicit statement of intentions? In other official documents? In politicians' speeches? According to D'Hainaut (1980), there are two ways of getting at the education policy of a country: either through a content analysis of intentions, or an analysis of the reality, the latter being more complicated. Analysis of intentions without reality or vice versa leaves the picture incomplete. Following D'Hainaut, I propose to concentrate on five indicators (among many possibilities), which reflect the values, the moral, political and cultural philosophy, that's to say, the fundamental choices faced in developing Portugal's education policy: Focus on the Individual vs the Group; Past, Present or Future Orientation; the Role of Political Ideologies; Access; Homogeneity.

1. Focus on the Individual vs the Group

The first question to be asked concerns whether the education policy of Portugal gives priority to the individual or groups of individuals. Does society as a whole matter more than the individual? Or is the policy designed for the interests of particular pressure groups, one social class more than the others, an economic lobby, a political party or a religious group? Or is there a balance between the interests of each individual and the whole society? Or is the struggle among social classes and the tension between the individual and society being ignored?

In spite of acknowledging the contribution of individual action to the development of society, the Education System Act shows a preoccupation with the individual. Over and over, it claims "the right to be different, out of respect for personalities and different ways of life, as well as consideration for and valuing of different fields of knowledge and culture." ["...o direito à diferença, mercê do respeito pelas personalidades e pelos projectos individuais da existência, bem como da consideração e valorização dos diferentes saberes e culturas."] But reality does not exactly accord with the Law. How to develop the individual's capacities? Are our schools provided with a variety of resources? Are they prepared to provide pupils different options in subject matter? Are there individual *curricula*? Do we contemplate an individual process of evaluation of pupils? Contrary to the intentions embodied in the Education System Act, the reality of Portuguese education is closer to neglect of individual differences.

2. Past, Present or Future Orientation

Is the Portuguese education system looking to the past, to that "golden age", when everything was perfect? Is it focused on a past where one can find the "best" models for behavior, the national heroes? Is our priority the preservation of old traditions? Or are we interested in facing the present as we live it, in solving the problems as they appear to us at the moment? And what attention is given to the future? And what kind of future is envisioned? A future that conforms to our plans and expectations, or an unpredictable future to which we must learn to adapt?

The Education System Act asserts that the education system has to "contribute to the defense of the national identity and to the strengthening of allegiance to the nation's historic origins, through development of awareness of the cultural patrimony of the Portuguese people." But the same text goes on to say that this must be accomplished "in the frame of the universalist European tradition and the growing interdependence and necessary solidarity among all the people of the world." ["...contribuir para a defesa da identidade nacional e para o reforço da fidelidade à matriz histórica, através da consciencialização relativamente ao património cultural do povo português" (art.3.a.). "...no quadro da tradição universalista europeia e da crescente interdependência e necessária solidariedade entre todos os povos do mundo." (art.3.a.).] "We are proudly alone!" Salazar said when Portugal was being pressured by the nations of the world to grant independence to its colonies. Facing increasing globalization, Portugal is now implementing programs that look beyond its borders: a) International exchanges (students and teachers are encouraged to participate in European exchange programs); b) access to world-wide repositories of information (primary schools have started to become linked to the internet); and c) emphasis on foreign language instruction (there are now instances of English teaching in primary schools). Portuguese education policy is oriented to the future

more than to the past or the present. The schools are no longer focused on a "glorious" distant past, memorizing the dynasties, and the kings and queens.

3. Political Dynamics

D'Hainaut's third analytic indicator has to do with political dynamics, the nature and the intensity of the changes the political forces want to introduce into the education system. Do they seek a conservative, progressive or revolutionary system? For which political system are we preparing our pupils to be participants? Or are they not being prepared for political participation at all? Are they being prepared for a totalitarian, a democratic or an anarchist regime? And when "democracy" is spoken of, is it the popular democracies of the past Soviet regime or the contemporary Chinese regime? Or is reference made only to western democracies, either presidential or parliamentry? The Education System Act speaks of democratization of society and teaching that guarantees "the right to a just and effective equality of opportunity for access to and success in school." Education is expected to "promote the development of a democratic and pluralistic spirit, that respects others and their ideas, and is open to dialogue and a free exchange of opinions." Education is also expected to "form citizens capable of judging with a critical and creative spirit the social milieu of which they are part and to strive for its progressive transformation." ["...o direito a uma justa e efectiva igualdade de oportunidades no acesso e sucesso escolares." "...promover o desenvolvimento do espírito democrático e pluralista, respeitador dos outros e das suas ideias, aberto ao diálogo e à livre troca de opiniões" (art.2.5.). "...formar cidadãos capazes de julgarem com espírito crítico e criativo o meio social em que se integram e de se empenharem na sua transformação progressiva." (art.2.5.).]

Nevertheless, students' participation in school life has decreased significantly, in spite of the existence of academic associations in higher education and also in secondary schools. Perhaps, contemporary issues simply do not galvanize them to action as did those in the past when the end of war in the African colonies was a popular student cause. Students appear to be more pragmatic now. The slogan "Not one more soldier to Africa" has been replaced by "No more fees!"

4. Openness and Effectiveness of Education

The fourth indicator proposed by D'Hainaut has to do with the openness and effectiveness of education. All political intentions are in accord in this respect, referring to the fact that all Portuguese people should have the right to education and culture. But the reality of attaining this goal is seen in the schooling rates, illiteracy rates, length of compulsory education, and the like. Salazar used to say the democratization of education would go against "natural inequalities," the legitimated and necessary hierarchy of values and persons in an well-ordered society. "It's necessary to put an end to the legal overproduction of intellectual forces" the Ministry of Education said. (Monteiro, A. R. 1975. 144). "Illiteracy in Portugal is not recent and nor did it prevent our literature from becoming one of the richest in past centuries" Salazar proclaimed. (Monteiro, A. R. 1975. 145-146).

Compulsory education in Portugal after the Revolution took the form of a program of Basic Education, which lasts nine years, divided into three consecutive cycles: a) First cycle, which lasts for four years (6 to 10 years old); b) Second cycle, which lasts for two years (10 to 12 years old); c) Third cycle, which lasts for three years (12 to 15 years old). Basic Education is free of charge: pupils don't need to pay any entrance or enrollment fees and they all have school insurance. General support, such as school meals, transports, books and materials are provided only to the most needy pupils.

Pre-school education is still optional, in spite of being part of the state education system. The number of places available is less than the number of applicants. Secondary education is also not compulsory. It is organized in a single cycle covering the 10th, 11th and 12th years of schooling and aims to consolidate and deepen the knowledge acquired in basic education to prepare young people both for further studies and for employment. Access to the university or polytechnic colleges is determined by the well-known *numerus clausus*. A combination of

secondary grades and performance on a national test is used to decide entrance to higher education. Talents and interests are simply ignored or subordinated to the need to balance supply and demand for occupations. It often happens that a student who dreamed of becoming a doctor is trained as a science teacher instead. And what possibilities for access to education exist for older, non-traditional students? "Lifelong learning" has entered the vocabulary of politicians. But what has been done other than traditional education? Has anyone begun to experiment with continuous education, sabbaticals, the adult literacy, and the like? Portugal has a long way to go to achieve a meaningful education system for non-traditional students.

But openness and effectiveness of education is not only measured by criteria of access to a particular level of schooling. How many of those who enroll ever graduate? And how long does it take to complete each level of schooling? And what about early school-leaving and school failure? Little is known about any of these features of the education system.

5. Homogeneity of Education

By the "homogeneity of the education system"—D'Hainaut's fifth indicator—we mean whether the same quality education is available for all people. In fact, education is very often stratified according to the age, sex and social origin of the persons to be educated. In my opinion, the Portuguese education system measures up well in this respect. The Portuguese Education System Act was acutely aware of these considerations when it recommended the goal of providing "a school system with a second opportunity for those who did not take adavantage of opportunities at the appropriate age." (art.3.i.) or when it promised "to assure equality of opportunities for both sexes. . ." (art.3.j.) or when it referred to "cultural promotion." ["...uma escolaridade de segunda oportunidade aos que dela não usufruíram na idade própria..." (art.3.i.) The access of women to education is a fact now, contrary to the situation in the past. In the last decade, women have entered some predominantly male professions, such as those related to law, medicine and university teaching. The creation of new universities and polytechnic colleges has also promoted social mobility for disadvantaged groups.

Geography can also affect the equality of schooling. The Education System Act acknowledged that Portugal's "unevenness of regional and local development should be corrected, which should enhance in all regions of the country equal access to the benefits of education, culture, and science." ["...assimetrias de desenvolvimento regional e local a serem corrigidas, devendo incrementar em todas as regiões do País a igualdade no acesso aos benefícios da educação, da cultura e da ciência". (art.3.h.)]. Ten years ago, a Portuguese resident of Madeira had less chances of having a higher degree than a Portuguese citizen living on the mainland. The creation of the University of Madeira (the youngest Portuguese University) made real the political intention of correcting such geographic inequities. Another dimension of the homogeneity of education is the curriculum itself. Shall it be the same for all people, or shall it be diversified according to each person's aptitudes, interests, social needs and talents? Shall it be the same for all Portugal, or is there a place for regional variations according to regional needs? Little has been done in this regard. The nation's curriculum is still heavily centralized. Before the Revolution, one spoke of one uniform curriculum from Minho (a northern region from Portugal) to Timor. One curriculum remains too much the reality today.

Conclusion

Rather than merely describing the organisation of the Portuguese education system, I have instead presented my interpretation of the system built by the new political regime. By contrasting intentions and reality, we learn at least three things about how policy shapes the education system:

1. Education policy has two rarely coincident dimensions: an official and a real one. We can't say there isn't any education policy because there isn't any concrete document on it. Portugal waited twelve years for the Education System Act to be written; this did not mean it lacked an education policy in the

meantime.

2. Education policy is always in evolution. Eleven years after the Law was published, it was rewritten (Law 115/97) with the introduction of an important measure on teacher education: The degree of *licenciado* is now absolutely necessary for the teaching of all levels (nursery and primary teaching included).
3. Education policy does not only depend on the pronouncements of politicians. It depends on the efforts of each of us—administrators, professors, teachers—in our day- to-day work. We can corrupt wonderful principles or we can give real meaning even to insipid political pronouncements.

Notes

1. This article was presented under the name "Portuguese Experience," at the ATEE Spring Conference "Changing Education in a Changing Society", at Klaipeda University, Lithuania, May, 1999.
2. The Editor thanks Alfinio Flores for translations of selected portions of the Education System Act.

References

Constituição da RepÚblica Portuguesa de 1976 (Actualised by the Constitutional Law nº 1/97, 20 September.

Commission Européenne, DG XXII: Education, Formatin et Jeunesse. (1996). *Réformes dans l'enseignement obligatoire: 1984-1994*. Bruxelles: Eurydice.

D'Hainaut, L. (1980). *Educação. Dos fins aos objectivos*. Coimbra: Livraria Almedina.

European Commission (DG XXII). (1999). *Development in the field of education at the national level*. Brussels: Eurydice National Units.

Eurydice. CEDEFOP. European Commission. (1995). *Structures of the Education and Initial Training Systems in the European Union*. Brussels. Luxembourg: ECSC-EC-EAEC.

Lei de Bases do Sistema Educativo (Law nº 115/97, 19 September)

Lei de Bases do Sistema Educativo (Law nº 46/86, 14 October)

Ministério da Educação. (1998). *Curriculo, Programas e Aprendizagens*. Lisboa: DES.

Monteiro, A. R. (1975). *Educação, acto politico*. Porto: Edições O Professor.

White Paper on Education and Training (Memo/96/162). *European Commission white paper teaching and learning: Towards the learning society*.

About the Author

Jesus Maria Sousa

University of Madeira. Department of Education Sciences
Campus Universitário da Penteada, 9000-390 Funchal, Portugal

Office phone number: 00-351-91-705203
Fax number: 00-351-91-705229
Email: angie@dragoeiro.uma.pt

Jesus Maria Sousa holds a PhD in Teacher Education from the Centre d'Études et de

BEST COPY AVAILABLE

Recherches en Sciences de l'Éducation in the University of Caen (France) and an MA in Curriculum Analysis and Organisation from the University of Minho (Portugal). Since 1989, she has served as a professor in the Department of Education Sciences in the University of Madeira (Portugal), being responsible for the area of Curriculum Theory and Development. From 1996 to 1998, Dr. Sousa was Vice-Rector of the University of Madeira for Pedagogical and Students Affairs, member of the Regional Education Council, and member of National University Evaluation Council--Fundação das Universidades Portuguesas. She is part of the Research and Development Centre 19 of the Association of Teacher Education in Europe (ATEE), working on Perspectives on Curriculum in Teacher Education, and is member in charge of the international office of the Provisional Bureau for the Société Européenne d'Ethnographie en Éducation, located at Lecce University (Italy). Currently, her research focuses on personal dimensions in teacher education, education ethnography and flexible curricula.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy AnalysisArchives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (480-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Andrew Coulson
a_coulson@msn.com

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Richard M. Jaeger
University of North Carolina - Greensboro

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
New York University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
Ann Leavenworth Center

Jay D. Scribner
University of Texas at Austin

for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatistica
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

EPAA • enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 290 times since January 11, 2000

Education Policy Analysis Archives

Volume 8 Number 6 January 11, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Higher Education Finance Reform in the Czech Republic: Transitions in Thought and Practice

Matthew S. McMullen
University of Pittsburgh

Abstract
Throughout Europe and especially the former communist countries of Central and Eastern Europe, universities and governments are evaluating ways to finance higher education other than the current dominant model of almost total government support. With government pressure to use limited funds in other areas (e.g., health care, environment, and the like) higher education institutions are being encouraged to become more economically self-sufficient. Some of these reforms have included establishing closer ties with regional businesses and introducing tuition and user fees to offset some of the costs of university operations. The particular focus of this report is on the new methods of financing higher education in the Czech Republic.

Introduction

In addition to the economic and political changes that began in 1989, Czechoslovakia peacefully separated in 1993 into the Czech and Slovak Republics. Their higher education system and society in general had to adapt to the initial political and economic transition in 1989 and then yet another transition in 1993 when it split into two countries. For this report, the term Czechoslovakia will sometimes be used in describing and analyzing the country for historical events prior to 1993, after it will be referred to as the Czech Republic.

Background: Higher education during the communist regime

When World War II ended, the Soviet higher education model was imposed as the dominant model for Central and Eastern Europe. The Soviet model

(communism) of political and economic development had some distinctly negative effects: The central economic planning model was inefficient and inflexible and was unable to adapt to changes in the world economy; the bureaucratic control of human rights and freedoms; the presence of internal security forces and the use of informants; the constant attempts to suppress dissident thinking and activity; the use of groups and organizations in service to the state (e.g., universities and mass media); and the use of Marxism/Leninism as a justification for all actions (Mauch & Fogel, 1992). The negative features of communism that affected both government and the economy also affected higher education. For example, government - run science academies did most of the research. The academies and universities were under strict government (Communist Party) control. Communist officials were afraid of politically unreliable faculty members who might influence students and often these faculty would work at the academies where they would not have contact with students (Kallen, 1991, Koucky, 1990).

The government rigidly centralized and politicized higher education in terms of access, curriculum, staffing, resource allocation and planning. Each successive five - year plan was designed to provide the planned state economy with personnel to meet the needs of the state. State planning limited the university's role in intellectual development and left little room for the inclusion of new scientific developments. It reduced universities to manpower training institutions and even this was not successful as realistic data were lacking on the national needs for skilled manpower. In time, the fulfillment of the five - year plans for higher education and individual institutions became goals in and of themselves and such plans were fulfilled whether or not they were appropriate. Thus, each successive five - year plan discouraged any assumption of responsibility on the part of university personnel. Also, higher education deteriorated as a result of political interference which often led to massive dismissal of many of the most competent staff (Kallen, 1991).

Despite the many negative features, the legacy of communism has had some positive effects: The state offered free public education from early childhood through the university level; eradicated widespread illiteracy; the educational level of the adult population in much of the region was raised to a level comparable to that of Western Europe; educators had designed innovative approaches to adult training; and there was a substantial increase in female participation in education. In addition to these positive features, the educational infrastructure (buildings, some equipment, etc.) was adequately developed and, as a result, future reforms can proceed with more of a focus on the content of the system (Kallen, 1991, Von Kopp, 1992).

The following is a basic description of the functions of higher education under communism. The functions listed were the ideology and not necessarily what was put into practice. *Five functions of higher education under communism*

1. Socio-political, economic, and cultural needs are filled.
2. Knowledge is created in association with individual and social consciousness--Attitudes, views, ideas, values, and aspirations.
3. Individual needs and experiences of academic staff members are developed and valued. 4) The training is used for modern and humanistic educational concerns (Holmberg & Wojtowicz, 1990, p. 10).

The communist system's goals and objectives for higher education were dictated by government officials concerned with creating the "communist man," someone for whom the good of the collective was more important than individual achievements. The Socialist Countries Conference for Ministers of Higher Education held in Prague in 1986, provided examples of how socialist education was directed by the ideology of the Party. The conference concluded by demanding that new strategic guidelines should be aimed at the full utilization of a new social system requiring good professional training and political and ideological maturity; code words for conformity to Party goals (Fischer-Galati, 1990). The principles, ideals, and functions of the higher education system were organized and controlled by the federal government and or the Communist Party officials.

Communism in Czechoslovakia

The Communist Party spent 40 years trying to remold Czechoslovak higher education into the image of the Soviet Union's system and the principles of international communism. The Party not only controlled all levels of higher education it also used institutions as instruments for controlling and educating students' minds to create the "communist man." National committees, which reported to the Ministry of the Interior, administered the system. All senior appointments in the Ministry of Education and in the National Committees were to Party members. The authority of the Ministry was minimal and confined to the administration of grants to universities and to the production of curricula and related textbooks. Membership in the Party was an important criterion for the highest academic posts. How closely an institution conformed to the planned system was the paramount means for evaluating the effectiveness of each institution no matter its output (Koucky, 1990, Kotasek, 1991, Yazdgerdi, 1990).

Summary of higher education under the communist system in Czechoslovakia

- The aims, tasks and resources in teaching and research were defined by the Communist Party and implemented by the state.
- Planning was comprehensive and an instrument of political control. Higher education institutions were accountable to the Communist Party and there was very limited institutional autonomy.
- There was almost no strategic planning at the institutional or sub-unit levels.
- The incentive system was based on the achievement of goals set by the Party.
- Higher education institutions were totally dependent on the state for financing and followed a rigid line-item budgeting process.
- The state set manpower planning with projections in the labor market.

(Holmberg and Wojtowicz, 1990; Bok, 1991; Cerych, 1993; Daniel, 1991; Mitter, 1990; Rupnik, 1992).

Changes in University Financing in Czechoslovakia

After an initial surge in student enrollments after WWII, growth in higher education slowed in the 1960s and the system of state funding reflected this trend. The financial decision- making process in higher education institutions started to change in the following ways:

1. The influence of technocrats on labor distribution planning in the national economy was growing, which meant their influence on the number of students admitted to each higher education institution was growing as well.
2. The participation of academics in the management of the higher education system was increasing.
3. The influence of political leadership was being replaced by the influence of technocrats.

During this transition period following the 1960s, the funding for higher education institutions took the form of incremental budgeting. For example, in a given year higher education institutions received the same funding as in the previous year plus a certain bonus based on their demands and the means available. The amount was based on constant negotiations between the state administration and the individual institutions of higher education. The increments depended to a large extent on each higher education officials' ability to negotiate an increase in financing (Holda, Cermakova, & Urbanek, 1994).

Problems in the methods used for funding higher education focused on the following areas:

- Ineffectiveness: The traditional scheme of budgetary base plus increment meant that institutions were expected to spend all of the entire current year's

budget, thus preparing the highest possible budget for the following year. This often meant a waste of resources since they would have been more efficiently used if they were allowed to be transferred to the next year. The negotiations on increments often took the form of political and personal arguments, rather than educational needs and concerns. In sum, the system did not reward superior performance.

- Lack of Transparency: Although the final budget of an institution was very strict and closely monitored, there were essentially no general rules for the funding of higher education institutions. Financial allocation was the result of a great number of private and opaque negotiations. Because of unclear rules, there were many subjective decisions.
- Lack of flexibility: As the budget was based on the previous years allotment, it could not respond to developments both inside and outside the institution (e.g., labor market, changing needs of the economy, etc.). Most important, the budget was not based on the number of students enrolled and thus did not reflect changes in these totals. (Heyneman, 1994, Holda, et al., 1994).

The transition to democracy and a market economy in Czechoslovakia (beginning in 1989) has had a pronounced influence on higher education. These changes have shown a movement away from political control of institutions and a change of thought as to the methods used to fund higher education operations (at least in the Czech Republic).

Higher Education in Transition

The sluggish economy and the growing frustration with the inefficient system eventually led to pressure for radical changes in university operations. Pressure to reform higher education came from academics, students and social groups. This pressure built up throughout the 1980s and came to a breaking point in 1989. Shortly after the student demonstrations of November 1989, which helped to focus and mobilize opposition to the old regime, individual groups of educators, students and members of the intelligentsia began to meet and discuss how the education and research system could be democratized and modernized. These meetings eventually culminated in the passage of the University Act of May 1990 which replaced the Higher Education Act of 1980 (Daniel, 1991).

The Czechoslovak Higher Education Act of 1990

The Higher Education Act of 1990 set out a democratic structure for the guidance of higher education and allowed academic freedom in many areas. State control and administration had been minimized and the authority of academic bodies increased. Unlike the previous system of decision making, academic institutions have the power to discuss and create policy. The Act revived the academic senate, which was abolished under the communist system, as an important governing body within universities. The revived senates (representing faculty, students and staff) were provided a large measure of control over their curriculum choices, hiring practices and research goals.

Under the 1990 Act, universities had the freedom to make their own economic decisions. For example, in 1991 higher education institutions received financial allocations from the state, as in previous years, by the system of 'basis and increment'. The difference was that the money was not earmarked for a specific function. In assigning funds, the Ministry of Education, advised by the university councils, assign funds to universities according to estimated annual capital and other expenditures. It became the responsibility of the individual universities (e.g., rectors, academic and faculty senates) to decide the specific distribution of these funds (Daniel, 1991). The only limits were the total amount of wages and general operating funds (e.g., buildings, etc.). In addition to these fiscal freedoms, the state allocated money to institutions without specifying how many students they should educate (Holda, et al., 1994).

The importance of the law on colleges and universities passed on May 4, 1990 can not be overstated. It put substantial decision making power back into the

BEST COPY AVAILABLE

hands of the university and its faculty and students. The law emphasized academic rights and freedoms as important principles of democracy and envisioned democracy in terms of self government and autonomous decision making within the higher education community. Through the 1990 Act and subsequent legislation, the post communist model of higher education is being developed.

Summary of the developing post-communist model in the Czech Republic

- Increasing importance of academic freedom, competition for students and funding and representation of academics in decision making bodies.
- Less direct central state control.
- Institutions accountable to constituencies such as students, government, business etc. and autonomy and academic freedom are determined by this accountability.
- Need to find multiple sources of financing and budgeting.
- Limited line-item budgeting process with a move to a formula method based upon the number of students enrolled.
- Higher education's relation to the labor market is significant, but often indirect, primarily the result of meeting market demands not dictated directly by the government, but by the market.
- Strategic planning by governing bodies within institutions seen as essential for the development of the institution.

University financing after 1989

Budget Allocations. In 1990, higher education consumed 17% of the total education budget. This is 1.7% of total education expenditures and .8% of the country's GDP. Of this amount, 40% were costs attributable to personnel, 30% to goods and facilities, 11% to research and 19% for students welfare and fellowships (Harbison, 1991a, 1991b). In 1991, budget resources were allocated as in the past (incremental) but government officials in the Czech Republic insisted that 10% of the overall higher education budget was to be distributed according to a new method of financial allocation. This new method was based on the number of students and a cost per student comparison across disciplines (a formula method). In 1992, universities implemented the new method. The budget was divided into three parts: Normative (the general costs of operating the institution such salaries, building costs, etc.); above normative (additional costs such as research, new projects etc.); and reserves. Thus, for the first time, the major part of the budget (normative) was to be allocated on a formula based on the number of students times the average costs of educating each student depending on their discipline (Mauch & Fogel, 1993). This was implemented, in part, to address the significant differences in the per student annual costs which range from a low of 16,000 Kcs per student of Economics to 79,000 Kcs for students in the Fine Arts. This difference in cost is because there is a higher teacher/student ratio in Economics (30/1) and a very low and not cost efficient ratio in fields such as the Arts (8/1) (Mokosin, 1995).

In 1992, the formula as applied yielded a great variation in the budgets of individual institutions. Some were cut in the extreme and others increased in comparison to 1991. The government decided to add a supplement to the funding provided by the state so that no institution would suffer too great a difference in one year. For example, in 1992 the University of South Bohemia had a total of 2,196 students in various disciplines. The normative amount determined by the Ministry in 1992 was 16,921 Kcs. This was roughly the average instructional cost per student in higher education. Multiplying that times 3,352 (an adjusted figure, only in part including the number of students) gave the university 56,722,000 Kcs as a 1992 budget, a 22.2% cut in the normative budget from the year before (see Table 1).

Table 1
Application Of The 1992 Budget For The University of South Bohemia
(In 1992, There Were Nine Study Fields)

	SocSci	Educ	Tech	Agr	Med	NatSci	Chem	Vet	Arts	Total
Univ. of South Bohemia	83	1,257		747	67	42				2,196
Ratios by Faculty	1.00	1.25	1.65	1.90	2.55	2.55	2.55	3.00	3.50	

University of South Bohemia Operational Expenditures for 1992 (thousands of Kcs.)

	Normative				Above Normative				
'91 Budget	Application of Ratios	Applic.%	Adjust.%	'92 Budget	Room/Board	Foreig. Lect.	Foreign Stud.	Sport	Total
72,901	56,722	-22.2%	-8.5%	66,693	8,328	420	0	10	75,451

Source: Budget documents from the *Czech Ministry of Education and Sport*, 1992. In Mauch and Fogel (1993).
Normative in Kcs = 16,921.

As a result of the application of the ratios, some of the 23 institutions in the Czech Republic received severe cuts and others great increases. The Ministry was forced to apply a correction factor in order that no institution would receive a cut or increase of more than 10%. For the University of South Bohemia the decrease turned out to be 8.5% which gave a normative budget of 66,693,000 Kcs. Adding in the above normative amount, the total budget for 1992 was 75,451,000 a severe cut from 1991 (Mauch & Fogel, 1993).

As stated earlier, the above normative budget was designated for activities above basic instructional costs, (e.g. student room and board, stipends for foreign students, sports, and special programs). The proportion of the budget derived from normative and above normative varies greatly by institution. It was suspected that one reason the budget is separated into these two categories is to enable the state increasingly to restrict the above normal budget by asking the users to pay ever increasing amounts until these activities are self-sufficient. Given restrictive budgets, it could be argued that universities may find it necessary to admit more students, release unnecessary or incompetent faculty, and attend to social demand (Daniel, 1991). This scenario has only partially developed.

Government's new role in the financial development of academic institutions

The government, through the Higher Education Act of 1990, has provided higher education institutions with additional opportunities to obtain non-governmental funding. Universities have been freed by the state to earn money through conferences, tourism, consulting, publishing, research, university enterprises, bookstores, lecture notes, exams, student fees, franchises and licensing arrangements. Universities may keep additional income in their own institutional accounts and the 1990 law exempts university enterprises from taxation (OECD, 1992). New laws have also allowed universities to seek donations and bequests and they can set up foundations to continue the work of the university in perpetuity.

Contributions from the private sector

A plan developed by the Ministry of Finance and implemented as part of the new tax system established on January 1, 1993, called for tax relief for private sector enterprises who donate funds to organizations or institutions with activities deemed to be in the public interest. Higher education institutions fit into this category (OECD, 1992). In this way the government is encouraging private sector enterprises to donate a portion of their earnings to higher education. While the potential is great, there are limitations. First, in the near future, funds from this source will be small because in the

current stage of the country's economic transition, firms are still struggling and profits are small. Donations from multi-nationals are not yet significant. Also, higher education institutions will have to compete with other institutions (e.g., museums, theaters, social service organizations, etc.). To secure this income, universities will have to find ways to make their programs attractive to donor groups unaccustomed to philanthropy.

When Czechoslovakia split into the Czech and Slovak Republics in 1993, initially there was little change in the higher education system. However, weaknesses in the 1990 Act especially within the area of financial decision making and academic management needed to be addressed if reform was to continue. There required specific plans and needs for each system and as such the Czech Republic developed its own higher education act in 1998.

Higher Education Act of 1998

A new Higher Education Act was approved by the Czech parliament in April 1998 which was designed to address many of the issues in management and financing that had developed since the implementation of the 1990 Act. The 1998 Act differed from the law passed in 1990 in that it allowed for the further creation of new programs, institutional diversification and a basic change of property rights.

The 1998 Act is a continuation of legislation on economic management of state property. The ownership of the property will be transferred from the state to the institutions of higher education, thus fundamentally altering their financial management concerning property and budgeting. The change in property rights transforms state higher education institutions into public legal entities. As a result there is a change in internal management, making institutions more self-determined by having self government rights in the use of their property (e.g., the right to collect fees for use of the property). Through this new method of management and ownership came the establishment of a new body in public higher education institutions, the Board of Trustees, consisting of academic and business leaders (Ministry of Education, Youth and Sport, 1998). Through this and other measures, the government further promotes the concept of multi-source financing by making institutions more self-reliant and decentralized.

The method for government funds to be distributed to higher education institutions will also change. Continuing with the method started in the early 1990s, funding will be focused on a formula funding method based on the number of students enrolled although it will affect significantly more than the 10% of the overall higher education budget that was indicated in 1992 (the exact amount was still not finalized during the writing of this report). It is believed that this will make the process more effective and transparent as it will depend on the institutions to develop programs to attract students and thus increase their funding from the government and fees imposed on the students. This method of funding will also be a means of competition among institutions for students. Creating programs in demand and improving existing programs will be important to attracting more students. This flexibility of operations will prove important to drawing in more funding from government and business.

The 1998 Act also introduces the concept of study fees for students of public higher education institutions. Before this Act, there were no tuition fees and students' families received an allowance, tax relief and stipends. Educational materials, housing and meals were also subsidized. In most cases these subsidies or stipends have been drastically reduced or eliminated slowly throughout the 1990s. Because of the 1998 Act, public higher education institutions can set the entrance fees (e.g., exams), but a maximum level is determined by the Act. As far as further fees for study (e.g., tuition, etc.), the minimum lower limit is prescribed by the Act and the maximum amount is left to the discretion of higher education institutions. Students who stay a year longer than is determined by the study program will be required to pay additional study fees. These funds will be used as a scholarship endowment to be expended within the institution. For private institutions, whose development is made possible through the 1998 Act, the study fees are not adjusted by the Act. The determination of their amount is completely at their discretion (Ministry of Education, Youth and Sport, 1998).

The diversification of higher education financing

With the implementation of the Higher Education Acts of 1990 and 1998, the democratization of society and further collaboration with the West, some necessary reforms are gradually being implemented to make higher education institutions more financially self-sufficient. These reforms have come in the form of a diversification in higher education institutions and programs. This diversification is an attempt to make the funding of institutions more flexible and adaptive to the needs of the economy by tying them more closely with business and government in their region. This in return is designed to provide them with additional revenue for their development. These reforms are occurring through a focus on regional higher education, bachelors studies and private higher education institutions among other areas.

Regional higher education institutions

After 1989, new universities and faculties were established that had a considerable influence on the regional structure of higher education. Since 1989, the share of the total number of students in the traditional university centers of Prague and Brno dropped by about 4%, as regional educational centers increased enrollments. Under 40% of students studied in Prague (the capital) in 1998, compared with 43% in 1989, and in Brno (the second largest city), 19% compared to 23% (Ministry of Education, Youth and Sport, 1998).

Some universities have become actively engaged with their regions and municipalities and have attempted to merge academic activities with local concerns. For example, in Liberec and Olomouc the universities have developed training and re-training programs in teaching, local administration and architecture, in close collaboration with their municipalities (Mokosin, 1995). Some regional universities have attempted to adapt to their reduced funding (in relation to inflation) from the government by developing ties with industry. Currently, the principal involvement of the universities in industrial re- organization is in the area of re-training managers and workers. In the future, the active engagement of university research and teaching on issues of regional concern is likely to flow from structured and regular consultations between scientists and teachers on one hand, and representatives of economic and social organizations and local government on the other.

As new laws have been passed in the area of tax exemption for non-profit organizations, it is expected that collaboration between higher education and industry will increase throughout the country which will further regionalize higher education and its ties with local business. This is designed to aid in the development of the regional economy. If innovative enterprises grow in numbers and the financial capability of these companies expands, this sort of collaboration could increase and be mutually beneficial to these businesses and the higher education institutions.

Bachelors Studies

Higher education institutions in the Czech Republic are attempting to meet changing skill level needs in the economy by offering more intensive courses that can be completed in a shorter period of time. One of the programs designed to do this is the bachelors studies program created in 1992. The bachelors study program usually lasts three years, but occasionally four. The degree of magister or engineer, the first and only level of undergraduate study prior to 1992, usually lasts five years (Mokosin, 1995, Winkler, 1993). The bachelors program does not replace the established method of study, but rather provides students with a more condensed, specialized option. Many bachelors study programs are designed to anticipate the future demand for high quality professionals in fields whose relevance to the economy has changed dramatically. These fields include; economics, engineering, business, mathematics, physics, law, public administration, and the like. (Ministry of Education, Youth and Sport, 1998).

According to the 1998 Act, the bachelors study program can lead to the awarding of the degree as a basic unit of higher education studies (Bachelors of Art, BcA) and there is now a bachelors degree offered at most institutions. bachelors courses are now offered at over 50 faculties in 18 higher education institutions. There are over 160 specializations within the faculties, many of which are offered with a part-time option (Prucha and Halberstat, 1993). Not surprisingly, most of the programs are located in the small provincial higher education institutions whereas the large well-established

BEST COPY AVAILABLE

universities in Prague or Brno are somewhat resistant to this non-traditional method of study. Of the over 160 specializations, only about 30 are in the two largest universities; Charles University in Prague and Masaryk University in Brno. A common thread among the different bachelors programs is the concept of a self-contained cycle leading to specific qualifications not previously offered in any of the existing institutions. These programs are often established to meet local needs at the request of regional authorities.

Regional sites have established separate fields of study such as the Textile and Engineering school in Liberec (technical school) which is developing a bachelors program in technical engineering in co-operation with Skoda works and its parent company, Volkswagon, in the neighboring town of Mlada Boleslav. The Liberec/Skoda bachelors program also has the support of the Ministry of Industry and is one of the few cases of close inter-ministerial collaboration in the sphere of higher education. The Faculty of Law in the University of Olomouc has a bachelors study program in the field of Public Administration, and several schools of Education have a bachelors cycle in studies qualifying engineers or other specialists to teach in professional secondary schools (Prucha and Halberstat, 1993).

The number of fields of study offered as well as the number of students taking bachelors degree programs is growing steadily. In the 1997/1998 academic year, the proportion of students taking bachelors degrees of the total number of undergraduates was 24.3% compared with only 11.1% in 1992/1993. The number of applicants for the bachelors programs continues to grow and enrollments have tripled in six years. (See Table 2.)

Table 2
Development of the number of students taking bachelors programs and their share in the total number of undergraduates in the Czech Republic (1992-1998)

Academic year	Students of bachelors programs	Undergrad. as a whole	Students taking bachelors programs as a % of total undergraduates
1992/93	12,628	114,185	11.1%
1993/94	15,624	122,456	12.8%
1994/95	28,147	129,453	21.7%
1995/96	34,821	139,774	24.9%
1996/97	36,668	156,868	23.5%
1997/98	39,410	162,373	24.3%

Source. Ministry of Education, Youth and Sport, 1998

With the addition of tuition and other user fees, these programs represent a growing source of additional income.

Private education

Private higher education did not exist in Czechoslovakia under communism. The 1990 law, while not forbidding the introduction of an alternative or binary system of higher education (both private and public institutions), did not authorize the establishment of private institutions. Legislation stated that: "It shall be the exclusive right of institutions of higher education to provide academic-scientific degrees to graduates and organize post-graduate studies" (Mokosin, 1995). As a result of the very restricted levels of privatization within Czech society prior to 1989, along with limiting legislation within the 1990 Higher Education Act, private higher education institutions had not been established to any significant extent since 1989.

As a means of diversification, coinciding with the increasing privatization of

government owned industry, government and academic policy makers through the 1998 Higher Education Act attempted to address the need for private higher education by making it significantly easier for the creation of these institutions. Institutions dealing with educational, scientific, research, development, or other creative activity can be founded after acquiring state permission. They are responsible for establishing their own fees for study (Ministry of Education, Youth and Sport, 1998).

In sum, higher education policy makers, in collaboration with government officials are seeking to diversify their financial sources and operations through the development of bachelors programs, private institutions and closer ties with the regions in which they are located. Through these methods, institutions are attempting to become more economically self- sufficient, either through the addition of fees for study or collaboration with business. Each of these programs increases academic decision making and creates opportunities for the development of financial resources outside of government funds, thus increasing their autonomy.

Conclusion

Higher education in the Czech Republic is going through an important transition, both politically and economically. New methods of financing university operations are necessary during the transition to a market economy as government funds are increasingly being drawn to other areas. Government and academic officials have worked together in the development of the Higher Education Acts of 1990 and 1998, both of which provide more academic freedom and opportunities for higher education institutions to develop programs that will meet their economic needs. Some of the key elements of change and diversification in higher education were:

- The regionalization of higher education through the tying of regional institutions to some financing from the region's industry and increasing the role of local government support
- The creation of bachelors programs and their expansion of enrollments in which is expected to account for at least 20% of the flow of higher education graduates by the year 2000.
- A shift in student financial support from the government to students and families (e.g., tuition fees and private education).

Because of the similar political and economic structures in all former post-communist countries, policy makers and educational researchers in transitional countries around the world may find the Czech transition useful in finding alternative methods of financing higher education. As the process is still developing, further research in this area after a longer period of implementation should lead to an evaluation of the alternative methods currently being undertaken in the Czech Republic and other countries in the region. As the countries of Central and Eastern Europe continue to move toward democracy and capitalism, higher education must move with it and create opportunities for itself now and in the future.

References

Bok, D. (1991). *Universities in transition: Observations and recommendations for Hungary and Czechoslovakia (CDC Report)*. Washington, DC: Citizens Democracy Corp.

Cerych, L. (1993). Particular context of present day East and Central Europe (editorial). *European Journal of Education, 28*(4), 377-379.

Daniel, D. (1991). National higher education and research systems of central Europe. Bratislava, Slovakia: Slovak Academic Information Agency.

Fischer-Galati, S. (1990). The impact of modernization in the education system: A comparative survey. *East European Quarterly, 25*(2), 275-282.

Harbison, R. W. (1991a). Education and training in Czechoslovakia. Unpublished report to the Czechoslovak government. Durham, England: Bicks, Sinclair and Associates, Inc.

Harbison, R. W. (1991b). Education and training in Czechoslovakia. Unpublished report to the Czechoslovak government. Durham, England: Bicks, Sinclair and Associates, Inc.

Heyneman, S. (1994). Education in the Europe and Central Asian Region: Policies of Adjustment and Excellence. Unpublished Report to the World Bank.

Holda, D., Cermakova, Z., & Urbanek, V. (1994). Changes in the funding of higher education in the Czech Republic. *European Journal of Education, 29* (1), 75-82.

Holmberg, C., & Wojtowicz, W. (Eds.). (1990). The Polish school system: Some social and historical aspects. Linkoping University, Sweden, Department of Education and Psychology.

Kallen, P. (1991). Academic exchange in Europe: Toward a new era of co-operation. The open door: Pan European cooperation. Bucharest, Romania: UNESCO European Center for Higher Education.

Kotasek, J. (1991). Czechoslovakia. Pp. 643-650 in P.G. Altbach & B. Johnstone (Eds.), *International higher education: An encyclopedia*. New York and London: Garland Publishing Inc.

Koucky, J. (1990). Czechoslovak higher education at the crossroads. *European Journal of Education, 25*(4), 361-377.

Mauch, J., & Fogel, D. (1992). Academic administrators in Hungary and Czechoslovakia: New roles and responsibilities. Paper presented at the Comparative Education Conference, Prague, Czechoslovakia.

Mauch, J., & Fogel, D. (1993). Issues in funding higher education in Eastern Europe: The case of Czechoslovakia. Pp. 207-237 in P.G. Altbach & B. Johnstone (Eds.), *The funding of higher education: International perspectives*. New York and London: Garland Publishing Inc.

Ministry of Education, Youth and Sport an the Center for Higher Education Studies. (1998). Higher Education in the Czech Republic. Prague, Czech. Republic.

Mitter, W. (1990). Education in Eastern Europe and the Soviet Union in a period of revolutionary change: An approach to comparative analysis. Unpublished manuscript.

Mokosin, V. (Ed.). (1995). Higher education in the Czech Republic, Prague, Czech Republic: The Center for Higher Education Studies.

OECD Report. (1992). Higher education policy review in the Czech and Slovak Federal Republic. Unpublished Manuscript.

Prucha, J. & Halberstat, L. (1993). The development and diversification of the higher education system: bachelors study. In OECD Report: Higher Education in the Czech Republic 1992-1993. Prague, Czech Republic: The Center for Higher Education Studies.

Rupnik, J. (1992). Higher education reform process in Central and Eastern Europe. *European Journal of Education, 27*(½), 145-153.

Von Kopp, B. (1992). The Eastern European revolution and education in Czechoslovakia. *Comparative Education Review, 36*, 101-113.

Winkler, J. (1993). Can one stop the pendulum? Managing change at Czech University. Unpublished manuscript.

Yazdegerdi, T. (1990). Changes in the educational system. *Report on Eastern Europe 1*, 14-18.

About the Author

Matthew S. McMullen
802 William Pitt Union
University of Pittsburgh
Pittsburgh, PA 15260

Voice: (412) 648-7421
Fax: (412) 383-7166

Email: Mcmullen+@pitt.edu

Matthew McMullen is a Research Associate of the Center for Russian and East European Studies and Visiting Faculty member at the Institute for International Studies in Education, University of Pittsburgh. His PhD (1996) is from the University of Pittsburgh, in Administrative and Policy Studies (International Development Education Program). He holds a Graduate Studies Certificate from the Center for Russian and East European Studies at Pitt and and from Charles University, Prague, Czech Republic, (1994) in the area of Economics and Political Science. His publications include McMullen, M., Donnorummo, R. and Mauch, J. (Eds.) (2000). *Higher Education and Emerging Markets: Development and Sustainability*. (Garland Publishing: New York) and McMullen, M. and Prucha, J. (2000). The Czech Republic: A Country in Transition....Again" in *Higher Education and Emerging Markets: Development and Sustainability*.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy AnalysisArchives* is http: epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (480-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina -- Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

BEST COPY AVAILABLE

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodriguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autonoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doc.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

BEST COPY AVAILABLE

Este artículo ha sido consultado 149 veces desde el 12 de janero de 2000

Education Policy Analysis Archives

Volume 8 Number 7 janero 12, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Indicadores de la implementación en procesos de reforma educativa en Uruguay: Una aproximación cualitativa

Marcos R. Sarasola
Universidad Catolica del Uruguay

Resumen

En este trabajo el autor estudia, desde una perspectiva cualitativa, la problemática de la implementación de innovaciones educativas. Toma por caso la reciente reforma del Ciclo Básico en Uruguay. Con base en los conceptos de van der Vegt y Vandenberghe (1992), analiza las "funciones guía" ejercidas por el director para poder regular el flujo interno de la implementación. La primera de ellas es la "claridad conceptual", que tiene que ver con las posibilidades de proveer a los profesores de una clara visión de lo que ha de lograrse con la implementación y con concretar esa visión en términos de saber profesional y habilidades de los docentes. La segunda es la "presión direccional" que refiere a un nivel operacional de la implementación; es decir, cómo se vinculan las actividades diarias con los objetivos de la innovación. La "función de apoyo", refiere al apoyo que brinda el director para la gestión de los recursos (materiales, emocionales, técnicos y administrativos) para que ellos efectivamente respalden el trabajo en el centro. Por último la "definición de laxitud", o sea, la definición que hace el director sobre el grado de autonomía que tienen los docentes frente a los objetivos externos de la innovación.

Implementation indicators in educative reform process in Uruguay:

BEST COPY AVAILABLE

A qualitative perspective

Abstract
In this paper the author investigates, from a qualitative perspective, the problems associated with implementing educational innovations. He studies the recent case of the Basic Cycle reform in Uruguay. Based on the concepts of van der Vegt and Vandenberghe (1992), he analyzes the "functions guides" exercised by the director in order to be able to regulate the internal flow of the implementation. The first of these is "conceptual clarity", which has to do with providing the professors with a clear vision of what will take place within the implementation, and with specifying that vision in terms of the professional knowledge and abilities of the faculty as well. The second of these is "directional pressure," which refers to an operational level of the implementation; that is to say, how daily activities mesh with the objectives of the reform. Next, "function of support" refers to the support offered by the director for the management of resources (material, emotional, technical, and administrative resources), so that the resources may then effectively support the work in the center. Finally is "definition of latitude," that is, the degree of educational autonomy that the faculty have with respect to the external objectives of the reform.

Introducción

En forma similar al concepto de "tensión esencial" que Kuhn (1987) utiliza para describir la dinámica de tradición e innovación como motor de la investigación científica, podemos afirmar que en el campo de la educación existe también una tensión básica entre permanencia y cambio que moviliza la doble función de las instituciones de enseñanza: mantener y trasmitir lo que ya es (Durkheim, 1974) y renovar la formas de enseñanza y aprendizaje a través de la innovación educativa.

En el caso de Uruguay, que nos ocupa en el presente artículo, se pueden detectar esta tensión en el pensamiento pedagógico que recorre la historia educativa del país. Un ejemplo son las charlas que Carlos Vaz Ferreira daba a los docentes en la década del veinte: Vaz Ferreira sostenía que el problema fundamental para la implementación de innovaciones en educación era suponer que todo lo anterior era malo. En contraste, cuando algo era concebido como "bueno" se lo llevaba hasta tal extremo, se lo dogmatizaba de tal manera, que terminaba por convertirse en una gran equivocación. Llamó a este fenómeno "la exageración pedagógica" (Vaz Ferreira, 1921).

Desde luego, para analizar la problemática de la innovación educativa es preciso acudir a una conceptualización más amplia. Como punto de partida resulta útil el esquema de House (1988) que identifica tres grandes clases de innovaciones: tecnológica, política y cultural. La perspectiva tecnológica, según el citado autor, se rige claramente por el concepto de la producción: es posible mejorar la educación si se introducen nuevas tecnologías. En la perspectiva política, una innovación educativa es un fenómeno en el que se ponen en juego intereses de grupos con poder político dentro de la sociedad. A diferencia de la perspectiva tecnológica, que pone el acento en la innovación en sí misma, ahora el énfasis se coloca en la innovación ubicada en determinado contexto político y social. Por último, la perspectiva cultural se centra en el contexto en que se ha de desarrollar la innovación, apela a conceptos como los significados generados por la comunidad, sus valores puestos en juego y el camino del consenso como vía privilegiada de elaboración colectiva. Estas tres perspectivas han estado presentes en la implementación de innovaciones en nuestro sistema educativo, aunque en diversos grados cada una de ellas. No han sido comunes, sin embargo, las experiencias guiadas desde una perspectiva cultural.

Un conjunto de cambios se ha introducido en el sistema educativo uruguayo, en los últimos años. En particular en la Enseñanza Secundaria, se manifiestan a través de la Experiencia Piloto en el Ciclo Básico donde el centro educativo como unidad de gestión adecuada para la consecución de buenos y suficientes aprendizajes, distingue la estrategia de las autoridades de la educación. Éstas han propuesto un nuevo Modelo de Centro sostenido por un nuevo estilo de gestión de

los directores, de relación entre los docentes y de ellos con los alumnos y sus padres. Una forma distinta de aproximarse y construir el conocimiento, a través de modificaciones en el curriculum.

En dicho contexto, el presente trabajo intenta, desde una perspectiva cultural, ensayar pistas lo suficientemente confiables como para ser recorridas nuevamente en posteriores trabajos a la hora de estudiar la implementación de una innovación en un centro educativo. Para ello hemos trabajado en dos liceos urbanos, uno en la capital del país y otro de una capital departamental del interior del país. Estos centros fueron elegidos a condición de que se hubieran integrado a la experiencia desde el comienzo y que tuvieran una alta implementación. En el aspecto metodológico, se trabajó con base en tres fuentes de evidencia: Entrevistas (el director, 16% de los profesores de docencia directa y un adscripto); Observación (el director en actividades cotidianas, reuniones de profesores, reuniones de coordinación y reuniones con padres). Análisis documental (Proyecto de Centro e informes sobre las acciones desarrolladas en el Centro).

1. Antecedentes y conceptos

1.1. La Experiencia Piloto en el Ciclo Básico

En 1996 se inicia una reforma experimental en ocho liceos y tres escuelas técnicas, para el primer año del Ciclo Básico. En 1997, la experiencia se expande a tres nuevas escuelas técnicas y a doce liceos más, totalizando veintiséis centros. Durante el año 1998 han participado de esta experiencia, treinta y tres de los trescientos cincuenta centros del Ciclo Básico.

Dos son los objetivos de la Experiencia Piloto del Ciclo Básico, declarados en la Exposición de Motivos de la Rendición de Cuentas y Balance de Ejecución Presupuestal para el Ejercicio 1996 (ANEP, 1995). El primer objetivo, es lograr la creación de una comunidad académica con un reducido número de profesores. Excepto los docentes de Expresión y del Curriculum Abierto, los profesores concentran treinta horas de trabajo en un turno, dedicando veinticinco de ellas a la docencia directa y las cinco restantes a la Coordinación y atención de los estudiantes fuera del aula. El segundo objetivo es entregar al estudiante elementos para el desarrollo de su capacidad para aprender, elementos para integrarse al mundo presente y bases inteligibles del saber científico.

Estos dos objetivos se articulan en torno a un nuevo Modelo de Centro sustentado por el Proyecto de Centro y a una modificación en la estructura curricular, que facilite la exitosa consecución de ellos. Este nuevo Modelo de Centro, supone la presencia de un director capaz de trabajar en equipo con sus docentes y capaz de hacer trabajar al equipo en la toma de decisiones y la resolución de problemas. Para ello se ha reducido el número de integrantes del cuerpo de docentes, así como también se ha logrado su permanencia en el centro durante todo el turno. Durante la permanencia del docente en el centro, se incluye un espacio específico de coordinación.

El espacio de la Coordinación, se espera que sea un ámbito exquisito para la construcción del Proyecto de Centro donde exista un compromiso de todos los actores institucionales con los procesos pedagógicos que allí ocurren y con sus resultados. Un lugar privilegiado para que el director pueda comunicar sus expectativas a los docentes, como elemento indispensable para la construcción de una visión común.

Se ha instrumentado, al servicio de este modelo de gestión y para mayor facilidad del desarrollo de los procesos pedagógicos, un nuevo curriculum. Los liceos pilotos, así concebidos, se han conformado en las unidades de implementación de los cambios en este nivel del sistema educativo. Éstos han establecido implícitamente una suerte de compromiso en llevar adelante los cambios propuestos. El carácter voluntario en la participación de los centros a través de sus directores y la manera diferente en que los docentes acceden al desempeño de sus cargos (Nota 1)en esas instituciones, les otorga un status diferente al del resto de los liceos.

Cada comunidad de la Experiencia Piloto, entonces, por su misma definición, se encuentra especialmente motivada a priori para trabajar de forma tal que se logren los objetivos. Es así que establecen acciones a partir de un status contractual (van der Vegt y Vandenberghe, 1992) con la innovación que se pretende implementar en ese

nivel.

En el marco de este status contractual, se han producido procesos de configuración del patrón de implementación propios, tal como se afirma en las conclusiones del Seguimiento de la Experiencia Piloto: "del estudio realizado surge la evidencia de que la propuesta innovadora fue puesta en práctica de manera diferente en cada centro" (ANEP, 1997b, p. 76). Así, las instituciones se han involucrado en un muy complejo proceso de cambio, que admite diferencias en la forma y profundidad en que los actores de esas organizaciones se han comprometido con él.

Los elementos claves que hacen a las diferencias en los grados de implementación, y que son sustentados por la literatura, han sido confirmados por los recientes estudios sobre la implementación de la Experiencia Piloto. Éstos establecen que, "los Centros Piloto que lograron los mayores niveles de cambio en su capacidad de gestión lo hicieron en base a la presencia de un director que construye una visión del centro con su equipo, a la existencia de un Proyecto de Centro que articula la gestión, al trabajo técnico desarrollado en equipo y a un nuevo patrón de inserción de los docentes en el centro" (ANEP, 1997b, p. 31).

1.2. El Proyecto de Centro

La implementación de una innovación a través de un Proyecto, no es un proceso automático y menos aún, seguro. Según Berman y McLaughlin (1978), la implementación puede seguir tres procesos distintos: no-implementación, cooptación o adaptación mutua. Como no-implementación, entienden el proceso por el que no se efectúan ajustes o alteraciones al proyecto inicial. La cooptación ocurre cuando los actores adaptan el proyecto a sus propias necesidades, sin que haya cambios en los comportamientos tradicionales de la institución. Por último, expresan los autores, la adaptación mutua ocurre cuando el proyecto y su puesta en marcha sufren modificaciones y cuando se efectúan ajustes en las funciones y las estructuras del centro, en relación con los objetivos externos, fruto de las particularidades de la comunidad educativa. Si bien este proceso no garantiza en sí mismo el éxito de la implementación, "es el único proceso que promueve el cambio en los docentes", es decir, los " docentes cambian si (y sólo si) trabajan en ajustar el diseño original del proyecto a su centro educativo" (Berman y McLaughlin, 1978, p. 17).

El desafío, entonces, parece ser conseguir un balance adecuado entre la formulación de las acciones y su implementación y la adecuada fundamentación desde un marco axiológico. "El reto es encontrar el equilibrio entre el uno y el otro, para que el documento exprese y refuerce la acción, y ésta pueda explicitarse, comunicarse y afianzarse a través del documento", como bien expresan Alsinet y Muñoz (1995, p. 70). Equilibrio difícil pero posible, si se está atento a la coherencia entre **la misión** del centro y **las acciones** que en él se desarrollan.

La opción por un proyecto innovador que impulse acciones dentro del centro, debe ir acompañada de la necesaria lucidez en el análisis de la cultura, para que en definitiva, en su concepción, no se esté generando su propio fracaso. La confección del documento es un momento importante en el proceso del proyecto, pero no es el momento central. Es referente para el colectivo y puede ayudar a dar claridad, pero es dentro del complejo entramado institucional donde se desarrolla lo medular. Se debe tener claro que no todo es posible en una organización dada, sino que "ésta parece actuar como filtro, que no deja pasar más que algunas iniciativas o ciertas acciones y rechaza otras" (Friedberg, 1988, p. 8).

No todo es posible en la organización, por lo que la viabilidad de la puesta en práctica de una innovación no puede pasar como un momento más en la elaboración de ella. La viabilidad también tiene un carácter eminentemente dinámico, esto es, además de un minucioso detalle en el momento de elaborar el proyecto, ha de tenerse claro que "la viabilidad de los cambios institucionales se construye" (Aguerrondo, 1992, p. 162). Es parte del mismo devenir del proyecto, la construcción de las condiciones que aumenten la probabilidad del éxito de las acciones.

1.3. El impacto sobre la cultura organizacional

A la hora de pensar en la implementación de un nuevo modelo de gestión de

un centro educativo, no es posible soslayar el hecho de que es una organización. Esto claramente da un marco para la acción, el cual presupone que un cambio implique modificar aspectos de la organización. En tal sentido es consistente la literatura que sostiene la necesidad de condiciones dentro de la organización para que se faciliten y desarrollen las innovaciones. Fullan (1985, 1986a, 1986b) habla justamente de algunas condiciones para el éxito de las innovaciones, tales como la necesidad de una cultura donde se haga el ejercicio del trabajo colectivo, donde existan creencias y visiones compartidas y esté claramente delimitada la función de los directores.

El Equipo de Docentes, es considerado como estructura idónea para el mejoramiento de la tarea y la posibilidad de generar transformaciones. Pero esta estructura dentro de la organización, supone generar una cultura no muy extendida en nuestros centros educativos (Achard, 1995; Ravela, 1989).

Los docentes son habitualmente formados en el individualismo. Éste es legitimado en la posterior práctica profesional, en una cultura del quejido que hace de los encuentros entre los docentes (Sala de Profesores, Reuniones de Evaluación, recreos, Salas Docentes, etc.) verdaderas instancias de lamento institucional, en lugar de momentos que ayuden a la mejora de la práctica como acertadamente afirma McLaughlin (1990). Los escasos encuentros previstos oficialmente como parte de la tarea docente, no dejan de tener una fuerte tendencia a ser de trámite puramente administrativo. De esta manera se fomenta, una vez más, la privacidad del docente en su clase. En este sentido es interesante lo que algunos autores como Lortie (1975) y Sarason (1982), han llamado celularismo: cada profesor en su clase con sus alumnos y su tarea. Con ésto se garantiza el actual funcionamiento, reforzando la cultura que hoy prevalece y preservándola, como es evidente, de cualquier cambio sustancial.

La perspectiva que plantea Escudero (1988, p. 91) de "la escuela como unidad de cambio y como lugar privilegiado para la formación de los profesores", ayuda a reforzar las acciones sobre los centros educativos, porque es en ellos donde han de consolidarse los cambios. Pero no es posible cambiar profundamente, sin modificaciones en la cultura interna de la escuela.

Sin embargo, la tarea de incidir sobre la cultura no es nada sencilla, ya que no significa incidir sobre una suerte de entelequia, sino sobre algo que ha sido construido diariamente. Algo que pertenece a la construcción del colectivo y que, como expresa Sathe (1983), es un conjunto de ideas (a menudo no expresadas) que comparten los miembros de una comunidad, o sea, una forma de ser y una forma de hacer.

En el momento de establecer algunas estrategias de trabajo orientadas a la cultura institucional, o mejor aún, a la inclusión o modificación de algunos elementos que son parte de ella, la cautela ha de signar los procedimientos. Es interesante en tal sentido, la afirmación de Rossman, Corbett y Firestone (1988, p. 126): "La aversión al cambio varía con el carácter de las normas a ser desafiadas y con lo novedoso del desafío". Por más claros que a veces puedan aparecer los cambios necesarios para el mejoramiento de la práctica, las resistencias lo son aún más. Staessens (1991) propone algunas pistas de trabajo sobre tres áreas de la cultura: (a) el director como constructor y transmisor de la cultura, (b) el consenso en los objetivos y (c) las relaciones profesionales entre los docentes.

Parece claro que nadie mejor que el director, a la hora de trasmitir qué es lo que importa a la institución. El director ocupa un lugar privilegiado en el concierto institucional, y tanto, que los factores que podrían suponerse como determinantes de su acción (el tamaño del centro, o la procedencia de su alumnado o su misma estructura), no son decisivos. Es más, tienen carácter secundario, como afirma Ball (1989), con respecto a su estilo, los tipos de influencias, las coaliciones de grupo u otros factores micropolíticos. Para centros educativos de enseñanza Secundaria pública de nuestro país, Aristimuño (1996) confirma tales afirmaciones y encuentra fuerte evidencia de que el rol del director, como constructor de la cultura, es decisivo a la hora de implementar innovaciones.

Así es que estamos convencidos de que los docentes tienen las potencialidades para transformar su quehacer, en una educación centrada en los aprendizajes. Paradoja de nuestra educación que expresa la posibilidad de que los docentes se centren en los aprendizajes. Paradoja de nuestros centros educativos en los que los

temas relativos al aprendizaje no son relevantes aunque sea posible, felizmente, detectar cierta relación entre la cultura organizacional y la calidad del trabajo docente en el aula (Aristimuño, 1996).

1.4. La implementación

Comencemos definiendo implementación. Para ello parece adecuada la definición dada por Berman (1981, p. 273): "La implementación consiste en la adaptación de una idea innovadora en su puesta en acción dentro de una institución. La efectiva implementación por parte de los actores, parece estar caracterizada por una adaptación mutua de ellos a través de la claridad que tengan acerca de los objetivos de la innovación y los comportamientos funcionales requeridos". Esta afirmación contiene los elementos esenciales a los que atender, en el momento de focalizar esta etapa de la innovación. En primer lugar porque toma en cuenta la importancia que tienen los diferentes actores, protagonistas de la puesta en acción. Las instancias en las que intervienen directamente las personas que ponen en funcionamiento la innovación desde su propia práctica, son instancias en las que la probabilidad de dar muerte a esa innovación es máxima. Desconocer esto, significa un error. Existen sobradas experiencias en nuestro país de modificaciones con un marcado acento en lo administrativo. El resultado es que poco han logrado en sus intentos y poco impacto han tenido en los sectores donde pretendían incidir.

Es esclarecedora la síntesis de razones que exponen Fullan y Pomfret (1977) para estudiar la implementación. La primera razón, es simplemente que no es posible saber qué es lo que ha cambiado hasta tanto no intentamos su conceptualización y su medida directamente. La segunda razón, es porque permite entender por qué tantos cambios en la educación fracasan cuando se trata de ponerlos en acción. La tercera razón, es que muchos fracasos provienen de haber ignorado la implementación o haberla confundido con otros aspectos propios del proceso de cambio. La cuarta razón, por último, es que hasta que no se estudie la implementación independientemente, se hace difícil interpretar los resultados del aprendizaje y su relación con los factores que lo determinan.

1.5. Las Funciones Guía

"La figura clave en la construcción de la respuesta institucional al cambio, es la del director" (ANEP, 1997b, p. 32). Pero, ¿de qué manera? ; ¿cuáles son las funciones del director que se vinculan directamente con esa construcción? ; ¿a través de qué funciones facilita la implementación?.

Se desprende del citado documento, que se trata de un director que construye una visión del centro con su equipo. Conclusión consistente con recientes investigaciones (Vandenberghe y Staessens, 1991), sobre la construcción de la visión a través del director. Cabe, entonces, transformar las preguntas en una nueva: ¿a través de qué funciones el director construye, junto con su equipo, la visión del centro?. Concretamente van der Vegt y Vandenberghe (1992), definieron las Funciones Guía ejercidas por el director para poder regular el flujo interno de la implementación. La primera de ellas es la **Claridad Conceptual**. Tiene que ver con proveer a los profesores de una clara visión de lo que ha de lograrse con la implementación y con concretar esa visión en términos de saber profesional y habilidades de los docentes. La segunda es la **Presión Direccional** y refiere a un nivel operacional de la implementación. Es decir, cómo se vinculan las actividades diarias con los objetivos de la innovación. La función de **Apoyo**, refiere al apoyo que brinda el director para la gestión de los recursos (materiales, emocionales, técnicos y administrativos) para que ellos efectivamente respalden el trabajo en el centro. Por último la **Definición de Laxitud**, o sea, la definición que hace el director sobre el grado de autonomía que tienen los docentes frente a los objetivos externos de la innovación.

2. El estudio

2.1. Los Liceos

La implementación, reiteramos el concepto, consiste en la adaptación de una idea innovadora en su puesta en acción dentro de una institución. Por lo que, para ser

hecha efectiva por parte de los actores, estará caracterizada por una adaptación mutua de ellos a través de la claridad que tengan acerca de los objetivos de la innovación y por la ejecución de las acciones que ellos desarrollen. El proceso de implementación de una innovación en educación es esencialmente un proceso de "ida y vuelta" (Farrar et al. , 1980), en el que las estrategias son modificadas para que se adecuen a cada institución. Por tanto, la implementación es un proceso que puede pensarse en términos de proceso organizacional en el que sus resultados surgen de los arreglos particulares que ocurren dentro de la propia organización para llevar adelante las estrategias. En nuestro estudio es posible ver diáfanamente cómo el Proyecto de Centro en cada liceo pone al nuevo Modelo de Centro a atender las particularidades, las necesidades, que los propios actores definen para la consecución de los objetivos. Los dos proyectos tienen su origen en diagnósticos que los enraízan en las singularidades de cada realidad pedagógica. Se eligen momentos y estructuras, formas diferentes de participación, pero, sin embargo, en los dos son herramientas hábiles para la consecución del objetivo primordial: el aprendizaje de los alumnos.

Otro elemento característico de este proceso de adaptación mutua, ha sido el espacio de Coordinación. La misma estructura pensada externamente a la institución, es modificada adaptándose al estilo y necesidades de las diferentes organizaciones. Si bien en los dos centros es un lugar de trabajo valorado positivamente por los docentes y usado estratégicamente por las direcciones, es posible encontrar diferencias que les dan un perfil propio, como decíamos, a la misma estructura.

2.2. Los directores

Las peculiaridades de cada centro han condicionado, tal como hemos visto, la elaboración del Proyecto de Centro y el funcionamiento del espacio de la Coordinación. Una de esas peculiaridades es claramente decisiva en el proceso de adaptación: la presencia del director. No caben dudas que, en cualquiera de los dos centros estudiados, su presencia es muy fuerte como constructores de la cultura organizacional. A pesar de tener grandes diferencias, hemos podido detectar muchos puntos de contacto entre ellos. Los docentes, fundamentalmente, pero también el análisis de documentos y las observaciones realizadas, permiten afirmar que en ambos casos los directores han logrado proveer a sus profesores de una clara visión de lo que ha de lograrse con la implementación: que los alumnos aprendan.

Es justamente para que los alumnos aprendan, que se ha dispuesto un nuevo Modelo de Centro. En tal sentido, hemos podido establecer que los profesores tenían claro que la Coordinación (como lugar de encuentro profesional y antídoto contra el aislamiento) y la elaboración por consenso del Proyecto de Centro (como marco referencial de las acciones en el aula), son dos herramientas fundamentales del Modelo, para la consecución de aquel objetivo. Los directores de los centros estudiados tienen en el ejercicio de las Funciones Guía puntos de contacto. Pero el ejercicio de dichas funciones se da en el contexto de una gestión que también tiene puntos de contacto.

En síntesis, es posible caracterizar a los directores de los centros educativos estudiados como profesionales prontos para el cambio antes de que éste llegara como oportunidad a través de la Experiencia Piloto. Cambio en la gestión de la organización que ahora es puesta al servicio de los docentes, para que faciliten los aprendizajes de los alumnos. En otras palabras, profesionales que ponen en juego su capacidad de hacer uso del sistema al servicio de los docentes y la consecución de los aprendizajes. Pero también una gestión signada por claridad en los objetivos y un gran respeto a los tiempos personales e institucionales.

Las funciones de Presión Direccional y de Claridad Conceptual son ejercidas sabiamente por ambos directores, con manifiestas particularidades. Esta singularidad se vio traducida en el compromiso que cada uno de los profesionales tiene para con sus alumnos (para que aprendan y lo disfruten), sus docentes (para que sean "profesionales-en-relación", construyendo juntos), y su centro (como lugar dispuesto para que en él ocurran buenos y suficientes procesos de enseñanza y de aprendizaje).

3. Conclusión

En primer lugar, reafirmamos que la metodología cualitativa es la herramienta más adecuada para el estudio de estos complejos procesos de implementación, en los que la compresión de la experiencia humana no es simplemente una cuestión de causas y efectos. No es posible soslayar el impacto que la implementación tiene en la cultura de cualquier organización; no es posible soslayar a las mujeres y a los hombres, distintos, únicos y exquisitos constructores de su propia realidad. Son ellas y ellos quienes actúan desde su propia situación y protagonizan el cotidiano vivir de las instituciones.

En segundo lugar, se hace necesario profundizar en las funciones de Apoyo y Definición de Laxitud. Es probable que haya que redefinirlas para que sea menos complejo su análisis, porque sin lugar a dudas no es menor su ejercicio por parte de los directores.

En tercer lugar, del copioso material que resultó del trabajo de campo hemos observado una veta de análisis a profundizar en cuanto a dos aspectos propios de la organización. Uno de ellos es la forma en que se resuelven los problemas, ya sean aquellos que pertenecen al cotidiano, como los que surgen en las reuniones de trabajo, muy especialmente en la Coordinación. El otro aspecto que consideramos rico para su profundización es la toma de decisiones, tanto en términos de cómo se realizan, como del impacto que ellas tienen en el funcionamiento del centro educativo.

Por último, queremos concluir diciendo que las Funciones Guía efectivamente y tal como queríamos demostrar, parecen servir, por lo menos en los casos estudiados, a los efectos de ayudar a la determinación del proceso de implementación de una innovación educativa. Sin dudas, habrá que seguir trabajando en la mejora de los instrumentos utilizados y en la atención a otros elementos de análisis que parecen también incidir en esos complejos procesos.

Nota

1. Es importante destacar que la forma de acceder a los cargos ha ido sufriendo modificaciones durante los tres años de implementación. Asimismo necesidades de funcionamiento han implicado diferencias en la forma de acceder al centro, por parte de los distintos actores.
2. The Editor thanks Michele S. Moses for translation of the Abstract.

Referencias

ACHARD, I. (1995) El rol del director de Enseñanza Secundaria en Montevideo. (Tesis de Maestría, no publicada). Montevideo: Universidad Católica del Uruguay.

AGUERRONDO, I. (1992) Planificación de las instituciones educativas. En: Frigerio, G. y otros. *Las instituciones educativas. Cara y ceca.* Buenos Aires: Troquel-Flacso.

ALSINET, J. y MUÑOZ, E. (1995) Proyecto de Centro, un universo de posibilidades. *Cuadernos de Pedagogía.* España. 238 (julio- agosto): 70-73.

ANEP (1995) *Exposición de Motivos de la Rendición de Cuentas y Balance de Ejecución Presupuestal para el Ejercicio 1996.* Montevideo: CODICEN

ANEP (1997a) *Aportes Para La Reflexión. Jornadas de sensibilización de la Experiencia Piloto.* Montevideo: CODICEN

ANEP (1997b) *Estudio de Seguimiento de la Experiencia Piloto del Ciclo Básico. Resultados de la implementación en 1996.* Montevideo: Programa MESyFOD.

ANEP (1998) *La Reforma Educativa en Uruguay.* Montevideo: CODICEN

ARISTIMUÑO, A. (1996) Schools do matter: a study about the implementation of remedial courses in four public high schools in Uruguay. (Tesis de Doctorado, no publicada). Lovaina: Katholieke Universiteit Leuven.

BALL, S. (1989) *La micropolitica de la escuela. Hacia una teoria de la organización escolar*. Madrid: Paidós-MEC.

BERMAN, P. (1981) Educational change. An implementation paradigm. En R. Lehming & M. Kane (Eds), *Improving schools. Using what we know*. (253-286) Beverly Hills: SAGE Publications.

BERMAN, P. y McLAUGHLIN, M. (1978) Implementing and sustaining innovations. En *Federal Programs Supporting Educational Change*, Vol. VIII. Santa Mónica, California: RAND.

DELGADO, J. y GUTIÉRREZ, J. (1995) *Métodos y técnicas cualitativas de investigación en Ciencias Sociales*. Madrid: APG.

DURKHEIM, E. (1974) *El carácter y las funciones sociales de la educación, Educación y Sociología*. Argentina: Shapire.

ESCUDERO, J. (1988) La innovación y la organización escolar. En: Pascual R. (coord.). *La gestión educativa ante la innovación y el cambio*. Madrid: Narcea.

FARRAR, E.; DESANCTIS, J. y COHEN, D. (1980) Views fron below: Implementation Research in Education. *Teachers College Record*, 82, 1, 77-100.

FRIEDBERG, E. (1988) Análisis sociológico de las organizaciones. *Fichas de capacitación*. Serie a. Nº 66. Montevideo: CLAEH.

FULLAN, M. (1982) *The meaning of educational change*. New York/Toronto: Teachers College Press/OISE Press.

FULLAN, M. (1985) Change Processes and Strategies at the local level. *The Elementary School Journal, 85*, 3, 391-421.

FULLAN, M. (1986a) La gestión del cambio educativo. *Simposium de innovación educativa*. Murcia.

FULLAN, M. (1986b) *Support Systems for Implementing Curriculum in School Boards*. Ontario: Ministry of Education.

FULLAN, M. y POMFRET, A. (1977) Research on Curriculum and Instruction Implementation. En *Review of Educational Research*, 47, 1, 335-397.

HERSEN, M y BARLOW, D. (1976) *Single-case experimental design. Strategies for studying behavior*. New York: Pergamon.

HOUSE, E. (1988). Tres perspectivas de la innovación educativa: tecnológica, política y cultural. *Revista de Educación* 286, 5-34.

LORTIE, D. (1975). *School Teacher: a sociological study*. Chicago: The University of Chicago Press.

MAJCHRZAK, A. (1984). *Methods for policy research*. Beverly Hills: SAGE Publications.

McLAUGHLIN, M. (1988). Ambientes institucionales que favorecen la motivación y productividad de los profesores. Ponencia del II Congreso Mundial Vasco. En: Villa, A. (Coord.), *Perspectivas y problemas de la función docente*. España: Narcea.

RAVELA, P. (1989). El liceo como organización. Estudio de la gestión educativa en la Enseñanza Media. (Tesis de Maestría, no publicada). Montevideo: FLACSO-CIEP.

ROSSMAN, G.; CORBETT, H. & FIRESTONE, W. (1985). Professional Cultures, Improvement Efforts and Effectiveness. Findings from a Study of Three High Schools. Philadelphia, PA: Research for Better Schools.

SARASON, S. (1982). *The Culture of Schools and the Problems of Change (2nd Edition)*. Boston: Allyn and Bacon.

SATHE, V. (1983). Implicatios of corporate culture: a manager's guide to action. *Organizational Dynamics, 12* ,2.

STAESSENS, K. (1991). The professional culture of innovating Primary Schools. Nine Case Studies. Annual meeting of the American Educational Research Association. Chicago.

STAKE, R. (1998). *Investigación con estudio de casos*. Madrid: Morata.

VANDENBERGHE, R. & STAESSENS, K. (1991). Vision as a core component in school culture. Annual meeting of the American Educational Research Association. Chicago.

VAZ FERREIRA, C. (1921). La exageración de la Pedagogía. En *Estudios Pedagógicos por Carlos Vaz Ferreira*. Montevideo: Barreiro y Cía. y Librería Nacional.

VEGT, R. VAN DER & VANDENBERGHE, R. (1992). Schools implementing a central reform policy. Findings from two national educational contexts. Annual meeting of the American Educational Research Association. San Francisco.

WITTROCK, M. (1989). *La investigación en la enseñanza II. Métodos cualitativos y de observación*. Madrid: Paidós.

YIN, R. (1989). *Case study research. Design and Methods*. Londres: SAGE Publications.

Acerca del Autor

Marcos R. Sarasola

Marcos R. Sarasola es Profesor de Innovacion Educativa en la Universidad Catolica del Uruguay; el estudiante del Doctorado en Filosofía de la Universidad de Deusto, Bilbao, España.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
aacosta@cucea.udg.mx

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatistica
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

EPAA Editorial Board

BEST COPY AVAILABLE

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

BEST COPY AVAILABLE

This article has been retrieved 426 times since January 13, 2000

Education Policy Analysis Archives

Volume 8 Number 8 January 13, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Asymmetry in Dual Language Practice: Assessing Imbalance in a Program Promoting Equality

Audrey Amrein
Arizona State University

Robert A. Peña
Arizona State University

Abstract

The capacity for dual-language programs to deliver specific benefits to students with different primary and secondary language skills continues to be debated. Individuals favoring dual language assert that as it relies upon a reciprocal approach, dual language students acquire dual language proficiency without the need for teachers to translate from one language to another. By utilizing and conserving the language skills that students bring, dual language students also gain cross-cultural understandings and an expanded opportunity to realize academic success in the future. Research that explores whether these programs meet the needs of monolingual and bilingual students is limited. The intent of this study is not to criticize dual language practice. Instead, it is to describe a newly implemented dual language immersion program that exists and operates in Phoenix, Arizona. In particular, this study examines the practices of dual language teachers at Leigh Elementary School and the challenges encountered as school personnel worked to provide students with different primary and secondary language skills increased opportunities to learn.

Introduction

While the efficacy of language programs remains a widely debated topic in educational discourse, researchers and planners agree that language programs do not exist within a vacuum, and that the benefits accrued by participating in these programs are likely to differ for individual students. This conclusion suggests that

language programs need to be analyzed on a case-by-case basis as their success is largely affected by the context in which the language program is developed. Further, researchers indicate that micro-level and macro-level issues related to planning and implementation must be examined to understand how the sociopolitical context of schools may favor or impede planning, language program development, and the access students are provided to become proficient in using a second language for example (Freeman, 1996).

Studying dual language practice in its context is important for addressing specific language education issues. For example, investigating a recently developed language program together with its context provides opportunities to identify school factors contributing to language acquisition and loss during the early stages of that program's implementation. In addition, studying dual language practices and the context in which those practices take place provides opportunities to explain why language programs experience varying levels of success in preparing students to be bilingual and biliterate.

This paper investigates a recently developed language program in its school context. In particular, the practices of teachers in a dual language program at Leigh Elementary School are examined. Further, the challenges encountered as school personnel struggled to provide students from majority and language minority backgrounds with increased opportunities to learn through dual language are investigated.

Dual Language Theory and Practice: A Review of the Literature

A review of the literature suggests that dual language programs strive to develop enhanced second language skills in all students (Valdés, 1997). Freeman (1996) suggests that effective dual language instruction occurs when teachers define bilingualism and cultural pluralism as "resources to be developed" (p. 558). Teachers in effective dual language programs generally adopt a language as resource rather than a language as problem orientation while providing instruction. She adds that language majority and language minority students are typically combined across dual language classroom settings in an effort to promote change by socializing students in ways that differ from how they are socialized in mainstream society.

In some models, language minority and majority students conduct their academic work using a language with which they are most familiar while being immersed in the language to be learned. Students receive language arts instruction, for example, in their native languages and receive all other content area instruction in the two languages of focus. Cummins (1979) suggests that allowing students to access curriculum using their native language results in their experiencing greater academic success and in students acquiring improved cognitive abilities. Cummins (1979) and others add that acquiring improved higher order thinking skills in their native language allows language minority students to acquire higher order thinking skills in a second language as well (see for example Christian, 1996; Hakuta, 1986; Krashen, 1991; Pucci, 1994; Riojas-Clark, 1995, and Valdés, 1995).

Christian (1995) explains that dual language programs integrate language minority and majority students and "provide instruction in, and through, two languages" (p. 66). The (L1) language describes the primary or the first language of the student, and the (L2) language describes the second language or the language to be acquired. To achieve a maximum benefit from dual language, Christian (1995) indicates that students from the two language backgrounds are together in each class for most or all of their content instruction. She suggests that dual language classrooms are formed to promote positive attitudes for students towards both languages and cultures, and that dual language programs emphasize full bilingual proficiency for native and nonnative speakers.

While researchers of dual language suggest that variability exists between different programs, they nonetheless indicate that most dual language programs have three goals in common (Christian, 1995). First, dual language programs are created to help students develop high levels of proficiency in their native and a second language. Second, these programs stress that students perform at or above grade level in academic areas in both languages. Third, developers of dual language programs emphasize that students acquire positive cross-cultural attitudes and

enhanced levels of self esteem.

Researchers indicate that developers and teachers of dual language programs stress students learning language primarily through content (Snow, Met, and Genesee, 1989). These individuals suggest that language is best developed within a content-based curriculum, rather than as the focus of classroom instruction. In addition, researchers, developers and dual language teachers emphasize carefully structuring the social interactional characteristics of programs as combining L1 and L2 students in the same instructional setting is believed to promote increased and better opportunities for language acquisition and development (Christian 1996). These individuals reason that by integrating students from two language groups in a mixed classroom setting, dual language offers the language learner access to practitioners and students who serve as L1 models. Additionally, these individuals suggest that this additive approach supports the ongoing development of the students' native language skills while a second language is being learned.

Christian (1996), Gonzales and Lezama (1974) indicate that dual language programs generally use one of two models. The first, or "90/10" model, finds Spanish, for example, being used for approximately 90% of the instructional time. The use of English as the medium of instruction is gradually increased until the proportion of instruction is "50/50". Under the "90/10" model, students whose primary language is English are immersed in Spanish, while students with a primary language other than English receive L1 instruction with a gradual introduction to English as the primary mode of instruction. In this case study, a "50/50" dual language model was used. In the "50/50" model, the percentage of L1 and L2 instruction is equal from the beginning (Christian, 1996; Gonzales and Lezama, 1974).

Methodology

The description of the methods used for collecting the data and completing this study are separated into five parts. Part one describes the documents that were collected and studied to learn about the operation of the dual language program. Part two describes the techniques used to complete the observations. Part three describes procedures that were followed during interviews with participants. Part four describes methods of data analysis, and part five introduces the theoretical framework used to complete this study.

Documentation

At the onset of data collection, a three-ring binder containing statistical and demographic information about Leigh and Leigh's community was provided to the researchers. Included in this folder were test score results, the school calendar, publications written in two languages used to recruit parents and students into the program, and other school publications describing the dual language program. In addition, advertisements and other announcements that were made available to the general public and throughout Leigh's campus were gathered and studied.

Observations

The sample included in this study was deliberately chosen and observed in each participating classroom. This resulted in six different classrooms being observed. Specifically, observations were completed in two classrooms per kindergarten, two classrooms per the 1st grade, and two classrooms per the 2nd grade. Although the program operated through the 3rd grade, observations in these classroom settings were not conducted.

Over a period of two years, approximately 50 hours of observation time, of which most was spent in the Spanish- speaking classrooms, were completed. The lengths of each observation ranged widely. Two or three of the observation periods lasted as long as 4 hours in a particular classroom setting while other observation periods lasted no more than 15 minutes in another classroom. Observation periods were determined in relation to daily classroom activities, and by using teachers' suggestions regarding key opportunities that should be observed. Observations were conducted as a complete observer, and neither the primary investigator or the

co-author of this study participated in the activities of the classroom whatsoever.

The first year of this study was no more than an introduction to the site and the program. Although some preliminary assertions emerged within this phase, these assertions were only hunches and were not in anyway found to be supported by data. Continuing on with the second year of this study in order to test those preliminary assertions, additional observation data was compiled to investigate other themes and to conduct an in-depth analysis of the dual-language program as it existed in its school context.

Interviews

Two formal interviews were conducted with the program director. The first was introductory. Findings from this interview almost entirely dealt with programmatic issues, guidelines, operations, and objectives. A second interview with the program director was held with a different intent. This interview came at a strategic time in the research. During this interview the main goal was to compare data generated during the observations with the director's perceptions of the program. Although some programmatic issues were discussed, this second interview delved more into theoretical issues that were related to working hypotheses. As such, this interview served as one of two total member checks. The second member check was conducted after a final draft of this paper was composed. The program director read the manuscript and provided feedback and other ideas to consider, many of which were re-worked into the manuscript.

Later, one informal interview with a board member and many other informal interviews with the teachers were conducted. These informal interviews occurred between class periods, on walks to the cafeteria, and sometimes, although efforts were made to avoid this practice, during instructional time.

Data Analysis

According to Erickson (1986), "one basic task of data analysis is to generate [these] assertions, largely through induction" (p. 146). In this study, the entire data corpus was analyzed for underlying themes. Following Erickson's (1986) procedures of data analysis, the data resources were converted by the primary author into items of data by rereading and revisiting the data corpus. Next, the data were coded by circling, in colored ink, analogous instances that related to the working assertions. From this, various instances and fragmented pieces that supported each assertion were sorted in order to "make clear to the reader what is meant by the various assertions, and to display the evidentiary warrant for [each of] the assertions" (Erickson, 1986, p. 149).

Through data analysis, it was especially important to be sensitive to "discrepancies between the ideal plan and its implementation" (Freeman, 1996, p. 563). One of the fundamental principles of dual language/bilingual immersion programs relates to insuring equal access to educational opportunity. In reference to bilingualism and bi-literacy, Freeman (1996) advises that "the explicit goal is for all of the students to master skills in both Spanish and English through equal representation and evaluation of Spanish and English" (p. 579). Moreover, equal attention and respect are to be given to the two languages spoken by the community's population, Spanish and English, in order to promote equal appreciation and involvement with the two languages, and to develop practices that are effective for schooling all Leigh students.

Theoretical Framework

It may be argued that symmetry is one of nature's wonders. In almost every shred of nature there exists some kind of underlying order. In fractals, repeated iterations of basic yet random shapes create symmetrical beauty. The simplest thread of a leaf can be reiterated millions of times to create a poised tree or the simplest geometric shape can be reiterated thousands of times to create a flower whose whorls are equalized. Each small portion of the shape, when magnified, can reproduce exactly a larger portion. Wheatley (1992) states that "Fractals, in stressing

qualitative measurement, remind us of the lessons of wholeness," lessons of order, and lessons of balance (p. 129). It may also be argued that asymmetry, defined as a lack of proportion, also occurs and is atypical. As such, imbalances or inequalities may be antagonistic and may impede what is essential to complete development and balance.

Asymmetry in this paper describes the tool used to study the dual language program at Leigh. This program proposes to promote balance, fairness, and equality. To that end, instances of asymmetry must be noticed and made apparent in order to rebalance the scale and provide individuals experiencing dual language equal opportunities to learn.

Instances of symmetry were noticed when the program promoted fairness and equality. For example, this program ensured that all school publications were printed in both Spanish and English. Ideally, this pattern was to be carried across this program to ensure an equal representation of both languages. The logistics developed in the planning period also promoted this principle of equality completely. Instances of asymmetry occurred, however, when the planners attempted to move theory to practice.

Finally, in addition to fixing a study in its contextual place, assessing the effectiveness with which program offerings provide symmetry in the form of equal opportunities for students to learn probably also requires that researchers account for the duration of the program's operation. In this research, the dual language program was in its second year of implementation. This is essential in that any assertions derived are limited by the newness of the program. On the other hand, because this program is in its infancy, an excellent opportunity to investigate how it operated within its sociopolitical context, and how it was challenged to address the call to provide equal access during its earliest stages of development was provided.

Findings

Findings taken from the data are divided into two parts. Part one provides demographic and background information as understanding dual language program development and practice requires examining the sociopolitical context in which these activities took place (Freeman, 1996). Part two introduces assertions on asymmetry and is comprised of three areas. Labeled instructional asymmetry, the first area describes instances when and where pedagogical imbalances occurred. The second area, labeled resource asymmetry, describes occasions when discrepancies in the availability of materials emerged. Area three is labeled student asymmetry describing characteristics of the student population and the students themselves that made providing equal opportunities to learn problematic.

Demographic and Background Information

Leigh Elementary School District experienced enormous and rapid changes in its student demographic makeup over the past several years. In 1997, 7,746 students were enrolled in the district. From 1990 to 1997, there was an 83% growth in total enrollment, a 77% growth in students classified as having a low socioeconomic status, a 132% growth in the population of ethnic minorities, and a 203% growth in students classified as Limited English Proficient (LEP). These demographic changes were accompanied by low student tests scores and by calls for school officials to develop an improved program for educating students.

According to district reports, Leigh Elementary is the most diverse of the district's elementary schools. At the time of this study, Leigh served 1250 students, a population composed of 11% ethnic majority and 89% ethnic minority students. Of the 89% ethnic minorities, 81% were Mexican- American, 4.9% were African-American, 2.5% were Native- American, and .3% were Asian-American. In contrast, Leigh's student population was socio-economically homogeneous. Almost 97% of the population participated in the free and reduced lunch program at the time this study was conducted. Further, Leigh's population was linguistically dichotomous. The proportion of Leigh's LEP students increased from 21.6% in 1993 to 70% in 1998. Spanish and English were the dominant languages at home and few students were bilingual upon admittance to Leigh.

In 1996 Leigh Elementary was awarded a Title VII Grant that funded a language program entitled the "Two-Way Bilingual Immersion Literacy in Two Languages" program. This program was developed to promote bilingualism for Leigh elementary students, regardless of their language proficiency status. By this, the program was developed to enhance access to educational opportunities for all Leigh students by providing increased opportunities for students from diverse language backgrounds to learn. This program focused on dual language immersion with the languages of focus being Spanish and English, the representative languages of the school's population.

The 1996-1997 school year was the year of planning. In the first year of implementation (the 1997-1998 school year and the second year of the grant), the program served approximately 160 students. As noted earlier, this program was still in its puerile stage just ending its second year of operation, and while Leigh's dual language program was viewed as a success by many, little external research had actually been conducted to assess this program's nature and effectiveness. On the other hand, research completed by Peña (In Press) does provide additional information about the elementary school district, the Title VII grant that funded the dual language program, and the individuals involved in developing and implementing the program.

Assertions on Asymmetry

Instructional Asymmetry

One finding that became apparent early during the conduct of this study was that the Spanish teachers were bilingual and the English teachers were monolingual. As such, the teachers were classified as either Spanish speakers or English speakers, and the classrooms were classified as being places where either Spanish or English was used as the sole language of instruction. Freeman (1996) suggests that the ideal dual language program calls for "the English-dominant teacher to speak and be spoken to only in English and for the Spanish-dominant teacher to speak and be spoken to only in Spanish" (p. 576). This also requires that the classroom teacher should not translate during instruction or when questions emerge. In other words, teachers in dual language programs must "be true" to their respective languages and their languages of instruction. In this sense, and consonant with the research, students should be able to identify teachers with one particular language and a specific classroom setting. Through this instructional formula, the students could also be ensured equal exposure to both languages and opportunities for language and cognitive development.

Instructional asymmetry resulted in this study when the teachers switched language codes. Again, all of the Spanish- speaking teachers were bilingual and the English-speaking teachers were monolingual. As such, the Spanish-speaking teachers were able to switch language codes. They had a greater capacity and tendency for not being "true" to the instructional language because they were fluent in two languages. For example, if a student did not comprehend what the Spanish-speaking teacher was saying, it was not unusual for the bilingual teacher to translate her message into English in order to reduce the student's confusion. None of the English-dominant teachers were able to speak Spanish. "making teacher code-switching impossible" (Freeman, 1996, p. 576). Because the English-speaking teachers were monolingual, the Spanish-speaking children were forced to comprehend English. In contrast, because the Spanish-speaking teachers were bilingual, the English- speaking children learned to rely on the on the Spanish-speaking teachers' tendency to translate.

Instructional asymmetry also resulted when teachers treated students unequally in communications. Invariably, when an English-speaking student posed a question to the Spanish-speaking teacher, the student would ask the question in English. Since the teacher was bilingual, the teacher could understand the question in English and could then respond to the question in Spanish. However, when the Spanish-speaking student posed a question, the English- speaking teacher could not understand and, therefore, would force the student to repeat the question in English. In this, the Spanish-speaking students were required to both speak and comprehend

English while the English-speaking students were only required to listen to the Spanish. The Spanish-speaking teachers did not force the spoken language while the monolingual English-speaking teachers forced the spoken language because they were monolingual. In this regard, the shortage of bilingual teachers not only resulted in the students experiencing different expectations, but the monolingual English speakers were provided with fewer opportunities to speak and master a second language.

In this study, one of the three bilingual teachers would not code-switch or translate from English to Spanish. This teacher would deflect questions back onto the English- speaking students requiring them to either tap into a language broker or try to understand Spanish on their own. This teacher performed in accordance with program guidelines, and was able to satisfy dual immersion principles related to furthering equal access.

These examples of instructional asymmetry are largely due to the newness of the program and to the shortage of bilingual teachers. Although the program guidelines state that only one language is to be used to ensure full immersion, analyses of data compiled for this study suggest it is especially difficult for the Spanish-speaking teachers to withhold instruction and other types of support when they are fluent in two languages. The teacher-participants were compelled to help students experiencing frustration to learn. The program's director noted that the teachers were increasingly becoming more accustomed to staying in, or being true to the target language and not translating, but as with any new program, following these requirements appeared to take a concerted effort and time.

Finally, the primary language of the teacher and the teacher's perceptions about dual language learning appeared to have affected this program's capacity to provide students with equal access. For example, while observing an English-speaking teacher teach her mixed language science class, the teacher approached the principle investigator of this study at the back of the room to talk. This teacher said that she had been an ESL teacher up until the present year. When asked how she liked the program, she replied that she had never seen kids at this grade level learn "English" faster. From a discourse analysis perspective, her response spoke directly to her perceptions regarding dual-language instruction. Her statement implied that having the students acquire English was her priority. Her objective as a teacher in this program, in other words, may have been to emphasize English acquisition over Spanish acquisition, while not promoting both languages equally. According to Cummins (1986), reforms are dependent on the extent to which educators redefine their roles with respect to the minority. In this study, the teacher's preference for having her mixed language students improve their English proficiency may have conjured distorted perceptions relative to how the students judged themselves, their peers, their native tongue, and the need to acquire a second language.

This last observation suggests that the future success of both the students and program are probably related to the importance that educators attribute to language acquisition and to how students learn. Success may also be connected to each teacher's skill, training, and personal ideology. Cummins (1996) states that "educators who see their role as adding a second language and cultural affiliation to their students' repertoire are likely to empower students more than those who see their role as replacing or subtracting students' primary language and culture" (p. 25).

Resource Asymmetry

Classroom resources describe children's literature books, resource manuals manipulatives at learning stations and games. According to dual language research (Freeman, 1996) and the program's guidelines, a Spanish-speaking teacher should only have Spanish resources within the classroom, and the English-speaking teacher should only have English resources within the classroom. In this study, the teacher's classroom environment was arranged at each teacher's discretion; likewise, the teachers were encouraged to stock their classrooms using materials written in the appropriate language of the room.

In this instance, an asymmetry occurred as the Spanish teachers utilized resources written in Spanish and English, and as the English teachers utilized resources that were written only in English. This resulted in students in the

BEST COPY AVAILABLE

Spanish-speaking classrooms accessing resources in both English and Spanish while students in the English-speaking classrooms could only access resources written in English. It also resulted in opportunities to learn or read in Spanish in the Spanish speaking classes being fewer than those opportunities for students to learn or read English in their English classrooms.

Analyses of the data collected indicated that the classroom environment as designed by the teacher was also out of balance. The posters and other classroom decorations in the Spanish-speaking classrooms were, for the most part, available in Spanish and English, while posters and decorations in the English-speaking classrooms were written in English only. The dual language posters available in the Spanish-speaking rooms translated from English to Spanish and back again, and may have been instructionally useful as such. In the English-speaking classrooms however, English was the only language used on the posters and throughout the classroom environment.

The school library and the resource room demonstrated a similar pattern. Resources available in Spanish were scarce overall, while the appropriateness of these same materials for students at different levels of development was also severely limited. For example, materials written in Spanish constituted less than 20% of the total shelving area; thus, the potential for a student to find a book written in English was five times as likely as it was for a student to select a book written in Spanish. This concurs with Pucci's (1994) findings that "the school library holdings of Spanish reading materials [were] far below what even the bare minimum would warrant" (p. 78).

Furthermore, findings taken from this study suggest that the materials available in Spanish were separate from other resources and located in an isolated section of the library's shelves. This suggests that access to these resources may have been even more difficult to gain as some monolingual Spanish-speaking students could feel uneasy and struggle with selecting materials that would separate them from their peers, and, as in Pucci's (1994) study, involve them in using books and learning aids in a "section of the library [that] was easily observable" (p. 74).

As with the case of bilingual teachers, this imbalance in classroom resources may also have had disparate implications for providing students with equal opportunities to learn. Access to resources was not balanced. This suggests that the pool of available resources was deeper for the English speaking students, and that these resources may have been geared toward English-speaking students, and toward making those students with more limited skills become more proficient in English. This lack of proportion may have reflected the newness of this program. More likely, however, this disproportion illustrated a hegemonic condition that is prevalent in U.S. society.

Student Asymmetry

According to Freeman (1996), "language majority students' participation in dual-language facilitates the development of academic competence in Spanish" (p. 571). In other words, equal numbers of English-speaking and Spanish- speaking students need to participate for a "50/50" model of dual language immersion to operate effectively. Further, equal numbers of students are needed during student interactions to provide balance, and so students can be readily available as peer resources.

Characteristics related to the student population at Leigh introduced additional challenges to developing the dual language program and providing students with equal opportunities to learn. For example, Leigh's population to begin with was lopsided. Leigh's high attrition rate and high rates of student mobility also kept the program numbers in constant flux. During the second interview with the program director, she noted that "population percentages range from 54%:46% to 70%:30% (Spanish:English)." Similarly, observations revealed that the makeup of students in their classes was usually weighted heavily on the Spanish-speaking side because the program lacked English speakers to complete the "50/50" balance.

Observations of classroom experiences also revealed that separation according to language occurred widely among the students. Although the program director stated "our kids play together, our kids recess together, our kids do learning together,

and that's got to impact how they think about the others... everyone is mixing with everybody in the program," separation among the students participating in the program was observed. According to the data, students separated themselves voluntarily into language cliques during formal instruction, free class time, and outside of the classroom setting. Although some of the classrooms were deliberately arranged by dual language teachers to integrate language speakers and prevent in-class separation, separation nonetheless occurred when students were allowed to make choices regarding peer interactions. For example, analyses of the data revealed that if students were allowed to seat themselves within the classroom at random or were allowed to form their own groups for group work, the students would break off into homogeneous language groups. This separation usually resulted in students associating with students who spoke the same language in other words. Furthermore, the grouping of students with similar languages and backgrounds reflected imbalances existing in the larger society. Consistent with Freeman's (1996) study of dual language programming, in other words, groupings between and among students "correspond[ed] to racial, ethnic, or class lines in society" (p. 579).

Finally, and in keeping with previous dual language research (Freeman, 1996), students acting as language brokers were expected to facilitate in the language learning process as well. Language brokers were encouraged to translate for and contribute to peers becoming bilingual and biliterate. However, due to their penchant for separating themselves from other students, the language brokers were observed as neither accessible to all students nor easy to "tap into." In short, observations revealed that the language brokers were more likely to associate with other language brokers and more likely to join the English monolingual groups rather than to interact with the Spanish monolingual students.

In this sense, these students hastened their assimilation into the dominant culture by choosing to speak the language of the dominant language group. This finding suggests that along with language brokers being viewed as members of an education elite, students with stronger bilingual and biliterate skills preferred to associate with other students who were prized because they shared enhanced bilingual proficiency. Consistent with findings taken from his study of cultural differences, "success in school came more readily for those willing to understate, separate from or deny their Mexican culture" (Peña, 1997, p.13).

Theoretical Discussion on Asymmetry

Although "English only" laws have not been voted into the U.S. Constitution, "English only" is practiced in many areas throughout the U.S. regardless of written policy. Freeman (1996) and Shannon (1995) suggest that as English is the language of the majority, equality and opportunity in the U.S. come first to those who master the English language. Relatedly, languages other than English always have had, and always may have, a secondary status according to these thinkers. As a result, it may be argued that English is the language of choice. The Bilingual Education Act of 1988 in itself mandates that students be given the opportunity to master English while not emphasizing that students improve or maintain their native tongue.

This emphasis on English only is likely to affect programs striving to promote equality through dual language instruction. As dual language programs attempt to value two languages equally, in other words, it may be predictable for programs like Leigh's to encounter resistance in moving from dual language theory to practice given the nature of their sociopolitical context. Furthermore, Freeman (1996) suggests that given internal and external societal pressures, "leakage between the ideal plan and its implementation is not only understandable but to be expected" (p. 565).

According to Fairclough (1989), the sociopolitical context describes the "dynamic interrelationships among situational, institutional, and societal levels that influence each other in important ways" (Freeman, 1996, p. 559). A crucial issue that needs examining then is how the socio-political context affects dual language program practice and reform. Further, researchers need to account for factors related to time and the relative newness of programs and school reforms. In this study, characteristics of the larger sociopolitical context and the newness of the program

combined to create asymmetry and influence the lack of equal opportunities that were provided to students.

In reference to *instructional asymmetry*, it seems that a citizenry that does not favor bilingualism may not encourage educators to cultivate bilingual students in public schools. Similarly, results taken from this study suggest that while being fluent in English enhanced communication between bilingual teachers and English-speaking students, this pattern of communication may have combined with social and political preferences to encourage dual language students to become proficient in English, native English speaking students to be apathetic about mastering a second language, and dual language students to believe that English is superior to Spanish.

Furthermore, instructional asymmetries occurred due to a shortage of bilingual teachers. The aforementioned instances of instructional asymmetry occurred as a result of the Spanish-speaking teachers' capacity and tendency to communicate using English. Hence, it seems that an equal dispersion of bilingual teachers across classroom settings would prevent these inequalities, but this is not plausible. If teachers with bilingual skills were equally available in the English-only and Spanish-only classrooms, only illusions of instructional symmetry would appear. It is true the teachers' language skills would be balanced across classrooms, but the potential for code-switching and language favoritism would now occur in both classrooms, doubling instructional errors. The instructional errors would infringe upon the program's quality by promoting inadequate, instead of unequal, opportunities to learn. Ironically then, given the findings in this study, promoting equality by equalizing the numbers of bilingual teachers would result in reduced program quality. It is possible that if teachers with bilingual skills were readily available in equal proportions, this program, and other dual-language programs for that matter, would become even more mediocre.

It may be that monolingual Spanish and monolingual English teachers would facilitate an ideal match between instructional theory and program practice. In this scenario, the instructional asymmetries that emerged in this research would more likely vanish and the program's quality could be maintained. Developing a dual-language program with monolingual teachers, however, might introduce an array of other challenges related to developing dual language programs, and to providing students with different language skills equal opportunities to learn.

In reference to assertions regarding *resource asymmetry*, findings in this study suggest that materials and resources in Spanish were most difficult to obtain. Further, being that Spanish resources are fewer in comparison to English resources in the community, materials available in Spanish are likely not only to be more scarce, but more costly to purchase. Pucci (1990), who conducted a survey of booksellers in the Los Angeles area in 1990, noted, for example, that prices for resources in Spanish are typically 20-200% higher than resources written in English (Pucci, 1994, p. 78). This scarcity of resources, when combined with higher costs, is likely to result in poorer districts like Leigh not being able to reinforce the Spanish language in the manner by which the programmatic guidelines and objectives articulated.

According to Pucci (1994), a "commitment must evidence itself in terms of tangible resources, as well as thoughtful policies" (Pucci, 1994, p. 78). Results taken from this study indicate that not only must dual language programs have such a commitment and make a deliberate effort to equalize resources, but in order for equal educational opportunities to be provided to Leigh's native Spanish speakers, extraordinary steps may be needed to purchase resources in Spanish that are not only likely to be significantly more expensive, but more burdensome for poor schools like those in the Leigh Elementary School District to afford.

In reference to assertions about *student separation*, the findings presented earlier stand as an example at the school level of what happens in the larger social context. The Spanish language may not have clout or political sway in U.S. society. Although it was developed to be a great "equalizer," this program catered to the English speakers and the bilingual students more often than those of students who spoke Spanish only.

Research cited in Cummins (1986) supports the efficacy of dual language immersion programs if the native language has a high status and is strongly

reinforced in the larger society (p. 20). In this study, asymmetry resulted in the English language being viewed with a higher status. English was perceived as more prevalent and necessary making the acquisition of a second and less esteemed language that much less desirable.

Conclusion

This study was important as it provided the opportunity to examine the relationship between dual language theory and practice in six dual language classroom settings. What transpired at Leigh holds meaning for how other schools develop and conduct their dual language programs. Without a systematic review of their practices, dual language programs may be subjecting students to inequality, to fewer educational opportunities, and to policies and practices that separate students according to race, ethnicity, and language orientation. Furthermore, lacking systematic study, schools working to implement dual language programs may continue to reproduce the inequalities and injustices that characterize the wider society thus making more failures inevitable (Cummins, 1986, p. 33).

Although Leigh's program demonstrated discontinuities between theory and practice, Leigh's successes should also be recognized. The program, especially with respect to its sociopolitical context and infancy, is providing educational opportunities by offering dual language to its students. This in itself represents a departure from how language minority students typically experience schooling. However, lacking greater symmetry, the benefits of dual language may never be fully realized.

References

Bartlett, J. (1968). *Familiar quotations: A collection of passages, phrases and proverbs traced to their sources in ancient and modern literature (14th ed.).* Toronto, Canada: Little, Brown and Company.

Christian, D. (1996). Two-way immersion education: Students learning through two languages. *The Modern Language Journal, 80* (1), 66-76.

Covaleskie, J.F. (1994). The educational system and resistance to reform: The limits of policy. *Educational Policy Analysis Archives, 2.* (Entire issue). [Available on line: http://olam.edu.asu.edu/epa/v2n4.html]

Cummins, J. (1986). Empowering minority students: A framework for intervention. *Harvard Educational Review, 56* (1), 18-36.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*, 222-251.

Erickson, F. (1986). Qualitative methods in research on teaching. In M. Wittrock, (Ed.), *Handbook of research on teaching* (3rd ed.).

Freeman, R.D. (1996). Dual-language planning at Oyster Bilingual School: It's much more than language, *TESOL Quarterly, 30* (3), 557-582.

Graue, M.E. & Walsh, D.J. (1998). *Studying children in context: Theories, methods, and ethics.* Thousand Oaks, CA: SAGE Publications.

Hakuta, K. (1986). *Mirror of language: The debate on bilingualism.* New York: Basic Books.

Krashen, S.D. (1991). Bilingual education: A focus on current research. (Focus Occasional Paper No. 3). Washington, DC: National Clearinghouse for Bilingual Education.

Merriam-Webster's ninth new collegiate dictionary (9th ed.). (1986). Springfield, MA: Merriam-Webster.

BEST COPY AVAILABLE

Peña, R.A. (In Press). A case study of parental involvement in a conversion from transitional to dual language instruction, *The Bilingual Research Journal.*

Peña, R.A. (1997). Cultural differences and the construction of meaning: Implications for the leadership and organizational context of schools. *Educational Policy Analysis Archives, 5*,10. (Entire issue). [Available on line: http://olam.ed.asu.edu/epaa/v5n10.html]

Pucci, S.L. (1994). Supporting Spanish language literacy: Latino children and free reading resources in schools, *Bilingual Research Journal, 18* (1&2), 67-82.

Riojas-Clark, E. (1995). How did you learn to write in English when you haven't been taught in English? The language experience approach in a dual language program, *The Bilingual Research Journal, 19* (3&4), 611-627.

Snow, M.A., Met. M., & Genesee, F. (1989). A conceptual framework for the integration of language and content in second-foreign language instruction, *TESOL Quarterly, 23*, 201-217.

Valdés, G. (1997). Dual-language immersion programs: A Cautionary note concerning the education of language- minority students, *Harvard Educational Review, 67*(3), 391-429.

Valdés, G. (1995). The teaching of minority languages as academic subjects: Pedagogical and theoretical challenges, *The Modern Language Journal, 79*(3), 299-328.

Wheatley, M.J. (1992). *Leadership and the new science: Learning about organizations from an orderly universe.* San Francisco, CA: Berrett-Koehler Publishers.

About the Authors

Audrey L. Amrein
College of Education
Arizona State University
Division of Educational Leadership and Policy Studies
PO Box 872411
Tempe, AZ 85287-2411

Voice: 480-965-3267
Fax: 480-965-1880

Email: amrein@asu.edu

Audrey L. Amrein is currently a graduate student enrolled in the Policy Studies PhD program in the College of Education at Arizona State University in Tempe, Arizona. Her research interests include the study of school policy, urban education, and students at-risk.

Robert A. Peña
College of Education
Arizona State University
Division of Educational Leadership and Policy Studies
PO Box 872411
Tempe, AZ 85287-2411

Voice: 480-965-5371
Fax: 480-965-1880

Email: rpena@asu.edu

Robert A. Peña, Ph.D., is currently an Assistant Professsor in Educational Leadership and Policy Studies at Arizona State University in Tempe, Arizona. His research interests include urban education, students at-risk, leadership, ethics, and the study of poverty.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy AnalysisArchives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (480-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apemberton@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina - Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois - UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Maria Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
.torres@gseisucla.edu

prior vols | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 253 times since January 24, 2000

archives | abstracts | editors | board | submit | comment | subscribe | search | epaa.asu.edu

Education Policy Analysis Archives

Volume 8 Number 9 January 24, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Gender Related Differences in Career Patterns of Principals in Alabama: A Statewide Study

William A. Spencer
Auburn University

Frances K. Kochan
Auburn University

Abstract

The purpose of this research was to determine the status of women administrators in the Alabama in terms of demographic and career patterns. A survey was sent to all principals in Alabama. Five hundred-fifty, or 42% of the principals responded. In Alabama, women principals are generally more recent in their position, are somewhat more likely to have come directly from the classroom, and have less mobility in acquiring the position.

Introduction

In many fields research has shown that women fare differently from men in terms of their career patterns. In cases such as engineering, there are far fewer women than men recruited into the educational programs which prepare them for the career field and those women experience higher levels of attrition than do their male counterparts (Riehl and Byrd, 1997). This unequal situation is compounded by the fact that women also tend to receive less compensation than their male counterparts, advance within the organization at a slower rate, and generally interrupt their professional careers in order to devote time to raising a family (Gupton & Slick, 1996). In K-12 education, females comprise 83 % of the elementary and 54% of the secondary teaching populations. Yet they constitute only 52 % of the principalships in elementary schools and 26 % of the high school positions (Henke, Choy, Geis, & Broughman, 1996). Only 7 % of the school superintendents in the United States are

women (Shakeshaft, 1998).

There is a general consensus that the administrative leadership of a school is the key element to the effectiveness of the school (Wallace, 1992; Short & Greer, 1997). While not disregarding the obviously critical role of teachers and parents, a poor principal or superintendent can nullify even the best of teachers' and parental efforts. Therefore it is essential that schools have effective, quality leaders. When examining women's capacity to serve as school leaders, some researchers believe that males and females have different leadership styles. (Nogay and Beebe, 1997; Irby and Brown, 1995). As Fisher (1999) put it,

> ". . . Sociologists, anthropologists, psychologists, even business analysts have extensively described this multifaceted gender difference: women's interest in personal contacts, their drive to achieve interpersonal harmony, and their tendency to work and play in egalitarian teams versus men's sensitivity to social dominance and their need to achieve rank in real or perceived hierarchies. "(p. 29)

Both Grogan (1996) and Aburden & Naisbett (1992) report that women's leadership style tends to be more transformative and inclusive than that of their male counterparts making females more capable of adopting a collaborative management, approach than men. These researchers add that this style is the preferred one for today's schools.

Others disagree with these assertions and argue that males and females do not differ significantly in the ways in which they lead (Astin & Leland, 1991; Dobbins & Platz, 1986; Eagly & Johnson, 1990). Mertz and McNeely (1996) suggest that the either/or, male/female dichotomy is too simplistic and that a multidimensional approach, which examines context, ethnicity, and other factors is required when conducting research on the issue of leadership style.

Whether differences exist in female and male leadership styles and whether one style is preferable to another is unresolved and merits further research. However, the research supports the fact that females are at least as effective in their leadership roles as men (Shakeshaft, 1990). Thus there is no apparent reason why women should not fill these positions in proportion to their presence in the educational field.

Alabama, like most of the nation, is entering a decade in which there will be a significant turnover in the principalship. Within 5 years, 40% of present principals expect to retire. Another 30% expect to leave these positions within 10 years (Kochan & Spencer, 1999). It is imperative that an ample supply of high quality professionals will be available to fill the vacancies these retirements will create. If there are factors which hinder the recruitment of able women into leadership positions, then public education and the state will pay a price in lost credibility and potential in securing quality leaders for its schools.

Purpose of the Study

The purpose of this study was to determine the status of women administrators in the Alabama in terms of demographic and career patterns. We sought to discover the degree to which females were represented in the administrative ranks and whether there were any discernible barriers hindering their entrance into these positions.

Methodology

Data Collection

A survey was developed around demographic questions and the state principals' competencies. The survey was sent to all principals in Alabama. The mailing included an explanatory letter, guaranteeing anonymity, and a postage paid self-addressed envelope. Questions addressed demographic issues of gender, ethnicity, age, and number of years in position. Principals were also asked about retirement plans and how they acquired their leadership styles. The last part of the survey asked principals to rank order the Alabama principal competencies and then to rank their own capabilities on these skills.

Data Analysis

Descriptive statistics were used to analyze most of the demographic data. Differences between men and women, reasons for retirement and experiences which influenced leadership styles were counted and placed in rank order. Mean scores were computed for responses to the importance and competence principals assigned to each of the Alabama principal competencies.

Findings

Demographic Characteristics

Five hundred-fifty, or 42% of the principals responded. Of these, 514 included a designation of gender and only those responses are included in these findings. Sixty-three percent of those responding to the gender question were males and thirty-seven percent were females. Eighty-four percent of the principals were white, non- Hispanics, 15 % were African American, and the remaining 1% were other minorities. Almost 90% of the principals are 40 years of age or older while forty-three percent are 50 years of age or older. The average age is 48.3. This is slightly higher that the last reported national average of 47.7 (Henke et al., 1996).

Educational Preparation

Data related to educational preparation indicates a difference between males and females. Male principals as a group have somewhat lower levels of professional education than do their female counterparts. Table 1 displays the educational degree and post-degree levels of female and male principals. Almost half of the males have a Master's degree. Slightly less than one-third have post Master's work or a Specialist Degree and less than a quarter have a post-Specialist work or a Doctorate. Females, on the other hand, are virtually evenly distributed across the three levels with more than one third having post Masters work or Specialist Degrees and more than one-third having post Specialist work or Doctoral Degrees. Using a Chi square analysis, these differences were found to be significant at greater than the .001 level (chi-square (df=2) = 15.332, $p < .001$).

Table 1
Educational Levels of Principals by Gender

	Masters or less	Post Masters or AA	Post AA or doctorate	Total
Male	151 (46.6%)	101 (31.2%)	72 (22.2%)	324
Females	59 (31.1%)	63 (33.2%)	68 (35.8%)	190
Total	210 (40.9%)	164 (31.9%)	140 (27.2%)	514

chi-square (df=2) = 15.332, p < .001

Consistent with this finding, the data also show that males have lower levels of professional certification than do female principals (Table 2) with about twelve percent more females having "AA" certification. These differences in formal preparation were also statistically significant (chi-square (df=1) = 5.67 (Corrected), p < .05).

Table 2
Certification Levels of Principals by Gender

	"A" Certification Principal	"AA" Certification Superintendent	Total
Males	130 (42.2%)	178 (57.8%)	308
Females	56 (30.9%)	125 (69.1%)	181
Total	186	303	399

chi-square (df=1) = 5.67 (Corrected), p < .05

Another difference between the groups is in the undergraduate preparation of principals. As shown in Table 3, female principals are much more likely to have majored in education as undergraduates than males. Men were more likely to have undergraduate majors in social science, natural science, mathematics or engineering than females. In part this may simply reflect the fact that at the elementary level principals are more generally female while at the middle school and high school levels, males predominate as principals. Again these differences are statistically significant (chi-square (df=4) = 55.44, p < .001.

Table 3
Background Preparation of Principals

	Education	Social Sciences	Humanities	Nat. Sci, Math or Engineering	Business or Other	Total
Male	176 (58.5%)	48 (15.9%)	10 (3.3%)	50 (16.6%)	17 (5.6%)	301
Female	160 (86.5%)	3 (1.6%)	8 (4.3%)	5 (2.7%)	9 (4.9%)	185
Total	336 (69.1%)	51 (10.5%)	18 (3.7%)	55 (11.3%)	26 (5.3%)	486

chi-square (df=4) = 55.44, p < .001

Length of Tenure in Position

As can be seen in Table 4, females have fewer years in their current positions than do their male counterparts. From those in their first year as principal up through about 8 years in the position, females are more prominent than males. Beginning with the ninth year and going forward, males are overrepresented. The maximum time in the job for a female principal was 21 years whereas the maximum for the males was 32 years. It is largely this highly skewed distribution that accounts for a significant difference in the average years in position for females vs. males (5.53 years vs 7.41 years). Thus women's entrance into the principalship roles appears to have increased in recent years.

Table 4
Years in Current Position

	0 - 4	5 - 9	10-14	15-19	20 or more	Total
Male	151 (46.5%)	82 (25.2%)	45 (13.8%)	25 (7.7%)	22 (6.8%)	325
Female	98 (51.6%)	64 (33.7%)	15 (7.9%)	12 (6.3%)	1 (.5%)	190
Total	249 (48.3%)	146 (28.3%)	60 (11.7%)	37 (7.2%)	23 (4.5%)	515

chi-square (df=4) = 18.10, p < .01

Entry into the Principalship

An important dimension of recruitment is whether leadership of an organization is provided by individuals who are already employed by that organization or by individuals who come from outside the organization. Another important issue is whether these leadership positions are open to all or whether some individuals have limited access to them. As shown in Table 5, principals in Alabama exhibit a marked tendency to come from within their own system. More than 80 percent became principals in the system in which they were already employed. However, of those who did come from outside the system, more than 75 percent were males. Thus females are somewhat more likely to become principals in their own systems than are males. This difference is also statistically significant (chi-square (df=1) = 7.48 (Corrected), p < .01).

Table 5
Origin of Principals

	Within Current System	From Outside System	Total
Male	253 (79.1%)	67 (20.9%)	320
Female	169 (88.9%)	21 (11.1%)	190
Total	422	88	510

chi-square (df=1) = 7.48 (Corrected), p < .01

A related issue of interest, is the position principals previously occupied prior to assuming their current principal role. Again, we observe a somewhat different pattern between males and females. As displayed in Table 6, females are proportionally more likely than males to have come from the central office or other supervisory position or from the classroom while males are proportionately more likely to accede to the principalship from either an assistant principal position or from being a principal in another school or system. Moreover these differences are significant (chi-square (df=2) = 19.9, p < .001). In spite of these differences, the trend for both groups is to become principals after being either an assistant principal or a principal in another school.

Table 6
Position Held Prior to This Principalship

	Supt, Asst or Assoc Supt,Supervisor	Principal or Asst Principal	Teacher, Coach or Other	Total
Male	12 (3.8%)	242 (77.6%)	58 (18.6%)	312
Female	15 (8%)	110 (58.8%)	62 (33.2%)	187
Total	27	352	120	499

chi-square (df=2) = 19.9, p < .001

Retirement Prospects

While mobility from one principalship to another may leave vacancies in a school system, overall the number of principals would appear to be relatively stable. However this appears to be changing in Alabama. A large proportion of current Alabama principals plan to retire in the near future. In Alabama, all public school employees belong to the Alabama Teachers Retirement System. After 25 years of service, they are eligible to retire but are not required to do so. According to the data shown in Table 7, over the next five years almost 75 percent of male principals will be eligible for retirement but only about 62 percent of female principals will be eligible. Thus female principals can anticipate a longer service career ahead before they would be eligible to retire.

Table 7
Eligibility for Retirement

	Now or This Year	Next Year	Next Five Years	Next Ten Years	More than 10 Years	Total
Males	29 (9.2%)	101 (32%)	104 (32.9%)	42 (13.3%)	40 (12.7%)	316
Females	15 (8.1%)	45 (24.2%)	56 (30.1%)	45 (24.2%)	25 (13.4%)	186
Total	44 (8.8%)	146 (29.1%)	160 (31.9%)	87 (17.3%)	65 (12.9%)	502

chi-square (df=4) = 10.97, p < .05

Being eligible to retire and actually retiring are, of course, different things. Therefore we examined current principals plan to retire in the near future. We also looked at whether there was a difference between males and females in this regard. The results, contained in Table 8, show that while there are differences between the genders in this regard, these differences were not statistically significant. Thus we would conclude that the two groups likely do not differ in the time frame within which they actually plan to retire.

Table 8
Planned Retirements

	This Year	Next Year	Next Five Years	Next Ten Years	After Ten Years	Total
Males	6 (2.2%)	13 (4.9%)	109 (40.7%)	81 (30.2%)	59 (22%)	268
Females	0	10 (6.9%)	51 (35.2%)	54 (37.2%)	30 (20.7%)	145
Total	6 (1.5%)	23 (5.6%)	160 (38.7%)	135 (32.7%)	89 (21.5%)	413

chi-square (df=4) = 6.18, n.s.

Reasons for Retiring

Turnover among principals is the result of many factors. Using information from the literature, we listed 14 reasons principals retire in the survey and asked the principals to indicate those which applied to them. Respondents were also given the option of adding any other reasons. Table 9 displays the list of reasons these principals would retire and their relative ranks based upon how frequently the respondents chose them. The number one reason given for retiring was to assume a better position. Thus technically, they are not leaving the professioin, but they are leaving the State of Alabama. But when one looks at the reasons these respondents selected for leaving this role through retirement, the correlation between the relative ranking of reason for retiring is fairly high between males and females (Spearman $r = .82$, $p < .001$), with a few notable discrepancies. Females rank frustration of goals as second highest in importance while males rank it sixth. Similarly females place more importance on a lack of fulfillment than do males. They also ranked the need for having more time with family at a much higher level than males. Females also more often than their male counterparts ranked the time needed to do the job as a reason to retire. At the same time, they have less problem apparently in dealing with the external mandates than do male principals and are somewhat less inclined to seek a new position out of state.

Table 9
Importance of Reasons Given for Retiring

Stated Reason	Male N (Rank)	Female N (Rank)
Better Opportunity Elsewhere	222 (1)	118 (1)
Too Much Community Politics	100 (2)	56 (2-tie)
Burn Out	91 (3)	46 (4)
Take Another Position in Another State	85 (4)	40 (7)
Too Many External Mandates	83 (5)	25 (11)
Too Much Frustration of My Goals	65 (6)	56 (2-tie)
Job Requires Too Much Time	60 (7)	43 (5-tie)
Too Many Financial Problems in My School	58 (8)	27 (10)
Lack of Fulfillment with Job	53 (9)	33 (8)
Need More Time with My Family	44 (10)	43 (5-tie)
Deteriorating Relations within School and Community	33 (11)	24 (12)
Other Reasons	28 (12)	28 (9)
Too Much Influence of Teachers' Organization	9 (13)	2 (13-tie)
Inadequately Prepared for the Job	2 (14)	0 (15)
Maternity Leave	1 (15)	2 (13-tie)
$r_s = .82, p < .001$	N = 325	N = 191

Importance of Specific Skills and Self Evaluation

To understand more fully why there might be differences in the desire to retire between males and females, a portion of the survey was dedicated to assessing (1) what principals now on the job believe to be the most important skills that a new principal would need, and (2) how those principals would assess their own level of proficiency in those same skills. As a basis for this, the researchers utilized a set of skills which the Alabama State Department of Education uses to evaluate principals in the field. Table 10 contains a list of these skills and their level of importance as seen by principals. While the relative importance level of each skill is the same for both males and females ($r = .985$), females tend to place more importance on the skills overall than do males. On balance there is about one fourth of a point difference which is statistically significant, $t(16) = 18.04$, $p < .001$.

Table 10
Importance of Principal Skills

BEST COPY AVAILABLE

Skill	Males	Females
Evaluates staff according to state and local policies and procedures	4.35	4.52
Demonstrates problem solving skills	4.35	4.49
Demonstrates organizational skills	4.29	4.48
Takes a leadership role in improving education	4.3	4.45
Communicates standards of expected performance	4.28	4.49
Improves professional knowledge and skills	4.18	4.53
Demonstrates skills in the recruitment, selection and assignment of school personnel	4.24	4.34
Manages Instruction	4.10	4.38
Implements clear instructional goals and specific achievement objectives for school	4.06	4.34
Establishes clear instructional goals and specific achievement objectives for school	4.04	4.29
Implements evaluation strategies for improvement of instruction	3.86	4.05
Understands special education laws and requirements	3.77	4.03
Understands the state's education accountability law and requirements	3.77	3.91
Understands legislative (political) processes that impact schools	3.67	3.68
Understands impact of the New Foundation Program for funding public schools	3.45	3.62
Understands the state's education trust fund and reports to board and community on finance issues (proration, etc.)	3.29	3.32
Understands the state's new accounting system for education	3.07	3.34

$r = .985$, $p < .001$; Mean diff = .23 (Females higher), $t(16) = 18.04$, $p < .001$

Self Rating of Principals

Using the same list of skills principals were asked to rate their own level of competence on each and the results are shown in Table 11. Again the results are similar to the previous case. Both males and females again are in basic agreement on their relative strengths and weaknesses. And again females tend to rate themselves slightly higher (Mean = .19) than do males, but the difference is statistically significant $t(16) = 8.57$, $p < .001$.

Table 11
Self Rating of Principal Skills

Skill	Males	Females
Evaluates staff according to state and local policies and procedures	4.43	4.68
Demonstrates problem solving skills	4.56	4.80
Demonstrates organizational skills	4.60	4.79
Takes a leadership role in improving education	4.53	4.73
Communicates standards of expected performance	4.57	4.79
Improves professional knowledge and skills	4.44	4.78
Demonstrates skills in the recruitment, selection and assignment of school personnel	4.60	4.77
Manages Instruction	4.57	4.75
Implements clear instructional goals and specific achievement objectives for school	4.57	4.84
Establishes clear instructional goals and specific achievement objectives for school	4.61	4.82
Implements evaluation strategies for improvement of instruction	4.34	4.64
Understands special education laws and requirements	4.42	4.70
Understands the state's education accountability law and requirements	4.17	4.42
Understands legislative (political) processes that impact schools	3.85	4.15
Understands impact of the New Foundation Program for funding public schools	3.97	4.19
Understands the state's education trust fund and reports to board and community on finance issues (proration, etc.)	3.58	3.81
Understands the state's new accounting system for education	4.04	4.17

$r = .977$, $p < .001$; Mean diff = .19 (Females higher), $t(16) = 8.57$, $p < .001$

Discussion

The Status of Females in the Principalship

Female respondents in this survey comprise 37% of the principals, which is slightly lower than the state figure of 38% and the national average of 42%. From the perspective of women seeking these positions, there is "good news" and "bad news." The findings suggest that although there has been an increase in the number of females entering the principalship in recent years, those who are in these positions have higher levels of education and more teaching experience than their male counterparts. This may be a factor in why females ranked their competence on the Alabama Principal Competencies more highly than males. Their higher levels of education and experience may have raised their competency levels and/or levels of confidence in their knowledge and skills. While it appears that opportunities are opening up, one-third of the females moved directly to the principalship from their teaching role.

That may mean it requires more time for them to become familiar and comfortable in the job. This may partially explain why the workload and the time the job takes was ranked more highly by females than males in retirement decisions. However, since this explanation seems to contradict females ranking their competence more highly than males, it is also possible that the time pressures females feel are related to family needs, a retirement decision factor ranked more highly by females than males. The impact of moving from a teaching position to a principalship requires further examination. The reasons a higher percent of females move from district office positions to the principals also bears further study.

An issue that may also be troubling for females is that while most principals are appointed to positions within the county in which they work, those selected for

these positions from outside their county are predominately male. Whether this is the result of females having less mobility than males or is an indication of some type of discriminatory attitude in educational systems is something that bears further investigation.

Potential Actions

The role of the principal in today's schools is a complex and difficult one for males and females alike. However, our data suggest that females may have to deal with more stresses and difficulties in acquiring and functioning in this role. The actions recommended below may help overcome some of these difficulties. Although these recommendations focus on the role of women, we would like to stress the need for all principals to receive support and guidance. Thus strategies should be developed that support the needs of all principals regardless of gender.

The disparity of females in the principalship relative to their numbers in the teaching force, may be the result of many factors: tradition, hiring practices, female unwillingness or reluctance to seek the role (Griffin, 1997), or issues related to family needs. This finding bears further study and examination within the state and school system structures. However, it is apparent that universities and school systems should take some actions to help deal with the disparate status of women in these positions. Programs of educational administration and school systems should consider establishing programs to identify, educate, and encourage females to enter the administrative ranks.

School districts should also examine their hiring practices and/or establish programs to groom and prepare female leaders in a systemic manner to assure that opportunities for advancement are made more apparent and equal between the genders. The lack of adequate role models is another issue systems should address. While the lack of a role model may have the advantage of allowing a new principal to be more open to new ideas it can also be the source of many difficulties including making political or technical errors and displaying a lack of confidence (Greenfield,1983). Having a role model provides validation for those entering a new role which is particularly important for traditional outsiders, such as women. This suggests that the advantages of having a role model outweigh the disadvantages (Hart, 1995; Pence, 1995). Since mentoring is seldom available for these women, school systems and educational leadership programs should consider creating mentoring opportunities for them to provide support and guidance (Funk & Kochan, in press; Crow, Mecklowitz & Weekes, 1992). In addition, "women-friendly" promotion structures that recognize the special career patterns of females related to childbearing and childbearing, proposed by Griffin (1997) and the alternate career model proposed by Grant (1989) should be reviewed and considered as avenues for assuring fair and equitable opportunities are available for females to enter the administrative ranks.

Implications

While this study has by no means been an exhaustive exploration of all gender differences in the principalship in Alabama, it has been sufficient to indicate that women principals are generally more recent in their position, are somewhat more likely to have come directly from the classroom, and have less mobility in acquiring the position. A cursory look at the figures indicates that females have assumed the principalship in larger numbers and percentages than in the past suggesting that barriers to females assuming school administrative roles are being overcome. However, there are some cautions that flow from the results. First, there is no reason to believe that the increases in female principals will continue exponentially over time. In fact, some of the data indicate that barriers and pressures may deter females from seeking or being selected for these positions. The data demonstrate that females are hired more often in places they are known and have worked and are seldom hired outside of their school systems. Thus their opportunities for employment as principals appear more limited than those of males.

Second, there is the issue of whether females will seek these positions at all and if they get them, one wonders if they will remain in them. Data related to

reasons for retirement indicate that family pressures fall more powerfully on females than on their male counterparts. When this is combined with the fact that women must have higher levels of education and more years of experience than males to get the position, some of them may decide not to seek these positions.

Third, the fact that many women come to the principalship without having been assistant principals may be an indication that they are getting principalships in schools where there are no assistant principals. This may be one of the reasons they selected the time spent on their job as a retirement factor more often than men. Further data should be gathered on this issue.

Most states, like Alabama, will be facing massive administrative retirements over the next decade (Muse & Thomas, 1991; National Association of Secondary School Principals, 1998). Likewise, the percent of female principals in Alabama is similar to the field in general. Therefore it is probable that our findings have uncovered meaningful issues that are present not just in Alabama, but in other states and school districts thoughout the country. It might be helpful for them to conduct similar studies to determine the status of females in the principalship in their settings. We believe that this statewide study poses questions not only for our state but for other states and for the field in general to consider. Among them are:

1. Despite recent increases in females entering the principalship, are they being held to a higher educational standard than males before being placed in these positions?
2. Are hiring practices free from gender-bias, particularly when "outsiders" are being considered to fill positions?
3. Are females being consistently placed in principalships where they are the only administrator?
4. How can female administrators be given support and mentored when there are so few role models to guide them?

Although we have focused on females, the future of our schools will be largely determined by the quality of our leadership. Alabama and the nation cannot afford to limit the potential or quantity of the pool of individuals who can provide this leadership. This study indicates that there are limits and barriers being faced by women who are qualified to fill the principalship in our state. Although progress has been made, particularly during the last five years, not all is "right with the world." Fairness and the needs of our state dictate that the issues raised and the questions posed be addressed not only by those who educate and hire school administrators in Alabama, but by those who do so throughout the nation.

References

Aburdene, P., & Naisbitt, J. (1992). *Megatrends for women*. New York: Villard. Myers

Astin, H.S., & Leland, C. (1991). *Women of influence: Women of vision*. San Francisco: Jossey-Bass.

Bishop, P.W., & Mulford. W.L. (1996). Empowerment in four Australian primary schools: They don't really care. *International Journal of School Reform, 5*(2),193-204.

Chase, S. E. (1995).*Ambiguous empowerment: The work narratives of women school superintendents.* Amherst, MA: University of Massachusetts Press.

Crow, G. M., & Glascock (1995). Socialization to a new conception of the principalship. *Journal of Educational Administration, 33*(1), 22-43

Crow, G. M., Mecklowitz, B., & Weekes, Y. N. (1992). *From teaching to administration: A preparation institute*. Lancaster, PA: Technomic.

Dobbins, G. H., & Platz, S. J. (1986). Sex differences in leadership: How real are they? *Academy of Management Review, 11*, 118-127.

Eagly, A.H., & Johnson, B.T. (1990). Gender and leadership style: A meta-analysis. *Psychological Bulletin*, 233-256.

Elmore, R.F. (1995). Structural reform and educational practice. *Educational Researcher, 24*(9), 23-26.

Funk, F., & Kochan, F. (In Press) Journeys in mentoring from a female perspective. In C. Mullen & D. Lick (Eds.). *New directions in mentoring: Creating a culture of synergy*. London: Falmer Press.

Grant, R.(1989). Heading for the top-the career experiences of a group of women deputies in one LEA. *Gender and Education, 1*(2), 113-25.

Greenfield, W. D. (1983). Career dynamics of educators: Research and policy issues. *Educational Administration Quarterly, 19*(2), 5-26.

Griffin, G.(1997). Teaching as a gendered experience. *Journal of Teacher Education, 48*(1), 7-18.

Gossetti, P.P. & Rusch, E. (1995). Reexamining educational leadership: Challenging assumptions. In D. M. Dunlap & Patricia A. Schmuck (Eds.) *Women leading in education* (pp. 11-35). Albany State University of New York Press.

Grogan, M. (1996).*Voices of women aspiring to the superintendency*. Albany, NY: State University Press.

Gupton, S. L., & Slick, G. A. (1996) *Highly successful women administrators: The inside stories of how they got there.* Thousand Oaks, CA: Corwin.

Hart, A. W. (1995). Women ascending to leadership: The organizational socialization of principals. In D. M. Dunlap & Patricia A. Schmuck (Eds.) *Women leading in education.*(pp. 105-124). Albany; State University of New York Press.

Helgeson, S. (1991). *The female advantage: Women's ways of leadership.* Garden City, NY: Doubleday

Henke, R. R., Choy, S. P., Geis, S., & Broughman , S. P. (1996). *Schools and staffing in the United States: A statistical profile, 1993-94* (NCES No. 96-124). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Hill, M. S., & Ragland, J. C. (1995) *Women as educational leaders: Opening windows, pushing ceilings*. Thousand Oaks, CA: Corwin.

Irby, B. J., & Brown, G. (1995, April). Constructing a feminist-inclusive theory of leadership. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Kochan, F. K. (1999). A collage of voice and form: A summary of the findings. In F. K. Kochan, Barbara L. Jackson, & Daniel Duke (Eds) *A thousand voices from the firing line: A study of educational leaders, their jobs, their preparation, and the problems they face* (pp.104- 110). Columbia, MO: UCEA Press

Kochan, F. K., Jackson, B., & Duke, D. (In press).*Voices from the firing line: How principals and superintendents view their jobs. Phase 1: Assembling the Pieces.* Columbia, MI: UCEA Press.

Kochan, F., & Spencer, W. (1999). The principalship: The practitioners'

perspectives. *Mid-South Educational Research Journal, 6*, 10.

Leithwood, K., & Menzeos, T. (1998). Forms and effects of school-based management: A review. *Educational Policy, 12*(3), 325-346.

Loden, M. (1985). *Feminine leadership or how to succeed in business without being one of the boys*. New York: Times Books.

Mertz, N., & McNeely, S. R. (1998) Women on the job: A study of female high school principals. *Educational Administration Quarterly, 34*(2), 196-222.

Muse, I., & Thomas, G. J. (1991). The rural principal: Select the best. *Journal of Rural and Small Schools, 4*(3), 32-37.

Myers, D. B., McKeegan, H. F., & Bieger, G. R. (1986, April). Male and female elementary principals: Development press, control press, and job satisfaction. Paper presented at the annual meeting of the American Educational Research Association. San Francisco, CA.

National Association of Secondary School Principals (1998). Is there a shortage of qualified candidates for openings in the principalship: An exploratory study. Arlington, VA: Educational Research Service.

Nogay, K., & Beebe, R. J. (1997). Gender and perceptions: Females as secondary principals. *Journal of School Leadership, 7*. 246-65.

Pence, J. L.(1995). Learning leadership through mentorships. In D.M. Dunlap & Schmuck, P.A. (Eds.), *Women leading in education* (pp. 125-144). Albany: State University Press.

Riehl, C., & Byrd, M. A. (1997). Gender differences among new recruits to school administration: Cautionary footnotes to an optimistic tale. *Educational Evaluation and Policy Analysis, 19*. 45-64.

Robinson, E., & Kochan, F. (1999). Theory and practice: Can they meet? Unpublished manuscript.

Rosener, J. B. (1990). Ways women lead. *Harvard Business Review*. 119-125.

Rosener, J. B., McAlester, D. J., & Stephens, G. K (1990). Leadership study: The women's forum. Unpublished manuscript, University of California at Irvine.

Schmuck, P. A., & Schubert, J. (1994). Women principals' views toward sex equity. In D. Dunlap & P. L. Schmuck (Eds.) *Women leading in education* (pp. 234- 237). Albany: State University of New York Press.

Shakeshaft, C. (1986). A female organizational culture. *Educational Horizons, 65*. 117-122.

Shakeshaft, C. (1989). The gender gap in research in educational administration. *Educational Administration Quarterly, 25*(4), 324-337.

Shakeshaft, C. (1990). *Women in educational administration*. Newbury Park, CA: SAGE Publication

Shakeshaft, C. (1998). Wild patience and bad fit: Assessing the impact of affirmative action on women in school administration. *Educational Researcher, 27*(9), 10-12.

Spencer, W., & Kochan, F. (1998, November). Gender related differences in principal career patterns in Alabama. Paper presented at the Annual Meeting of the Mid- South Educational Research Association New Orleans, LA.

Note

Adapted from a paper presented at the Annual Conference of the Midsouth Educational Research Association, New Orleans, LA, November, 1998.

About the Authors

William A. Spencer

EFLT Department
4036 Haley Center
Auburn University, AL 36849

Phone: (334) 844-3073
Fax: (334) 844-3072
Email: spencwa@auburn.edu

William A. Spencer is professor of education and has conducted research both in the U.S. and abroad on factors influencing the academic performance of middle school youth, as well as the causes and consequences of dropping out of high school. He has also studied the relationship between principal leadership and school climate as well as the impact of teacher involvement on teacher perceptions of principal leadership.

Frances K. Kochan

Truman Pierce Institute
2195 Haley Center
Auburn University, AL 36849

Phone: (334) 844-4488
Fax: (334) 844-0558
Email: kochafr@auburn.edu

Frances K. Kochan is Director of the Truman Pierce Institute and Associate Professor in Educational Leadership. Her research interests focus on creating collaborative communities, the relationship of beliefs and practice, and the role of educational leaders. She is co-editor of a book on leadership and has published in numerous journals including *Theory into Practice*, *Journal of Teacher Education*, and *Planning and Changing*.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy AnalysisArchives* is http: epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (480-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

BEST COPY AVAILABLE

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
aacosta@cucea.udg.mx

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Maria Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Hugo Casanova Cardiel
Universidad Nacional Autónoma de México

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

BEST COPY AVAILABLE

This article has been retrieved 343 times since January 27, 2000

archives | abstracts | editors | board | submit | comment | subscribe | search | epaa.asu.edu

Education Policy Analysis Archives

Volume 8 Number 10 January 27, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Collegiate Grading Practices and the Gender Pay Gap

Alicia C. Dowd
Cornell University

Abstract
Extending research findings by R. Sabot and J. Wakeman-Linn (1991), this article presents a theoretical analysis showing that relatively low grading quantitative fields and high grading verbal fields create a disincentive for college women to invest in quantitative study. Pressures on grading practices are modeled using higher education production functions.

The gender pay gap has narrowed in the United States since the 1970s, but is still of sufficient magnitude to warrant concern about the equal employment and status of women. The decrease in the size of the gap can be explained in part by the increasing numbers of college women who responded to expanded opportunities in the labor market and chose to enter technical and applied fields, particularly business (Eide, 1994; Loury, 1997). Women entering fields requiring quantitative skills can expect a greater return on their educational investments, because such skills are a relatively scarce human capital input (Paglin & Rufolo, 1990). Numerous studies have demonstrated that, all else equal, college graduates with quantitative skills will earn more than their counterparts without such skills (Berger, 1992; Eide, 1994; James & Alsalam, 1993; Rumberger & Thomas, 1993; Sharp & Weidman, 1989). However, women continue to be disproportionately represented in the humanities and social sciences and underrepresented in mathematics and the applied and physical sciences (National Center for Education Statistics, 1997). The theoretical analysis presented in this article shows that one way to increase the participation of college women in quantitative fields, and potentially reduce the pay gap even further, is to institute uniform collegiate grading practices in quantitative and nonquantitative fields.

Previous research (Kuh & Hu, 1999; Sabot & Wakeman-Linn, 1991) has provided evidence that grade inflation and compression has occurred in collegiate

disciplines at different rates, creating non-uniform (or divergent) grading practices. One factor contributing to the underenrollment of women in quantitative fields may be the use of relatively high grading practices in nonquantitative, or "verbal," fields and low grading practices in quantitative fields. This article has two purposes. The first is to show that grading disparities between academic disciplines have a significant impact on the curricular and career choices of female students. The second is to apply the analytical tool of the higher education production function to explain pressures on assessment practices from within and outside the academy that lead to divergent grading practices. This analysis also considers from which quarter pressure might come to change such practices. The discussion takes account of the public and private nature of institutions of higher education, noting that human capital formation is not their only, or necessarily even primary, function.

Theoretical Framework

Students earn college credits and degrees by investing time, money, and effort. At the majority of institutions of higher education, student performance in classes is evaluated with grades, and students must receive passing grades to receive credit for coursework. Students must also earn a sufficient number of credits in prescribed areas to be granted a degree in any given field of study. Variation in the effort students must expend to successfully complete coursework in different fields creates variation also in the costs of earning credits in those fields. The full costs of that effort will be tempered by a student's motivation and interest.

A student might pay the same tuition to major in mathematics or in English, but if she has strong mathematical skills and weak writing skills, she will have to invest more time to earn passing grades in English than in mathematics. Thus, the cost of earning a degree in a given field depends on the effort a student must expend to complete courses with a passing grade, or, for students with higher standards, to be satisfied with his or her own performance. In addition, some fields have more numerous or rigorous requirements, which raises the cost of study in that field relative to other fields for any student. (Note 1) The grades students receive inform them of their area of comparative advantage in completing coursework in a subject, the probability of successful completion of a course of study, and the costs (in time and effort) of obtaining a degree (Altonji, 1993).

The analysis presented in this article is based on an economic approach (Becker, 1976) to understanding the curricular and career choices of college students. Educational choices are treated as investment decisions, influenced by pecuniary and non-pecuniary costs and benefits. By her curricular choices, a student determines the specific type of human capital she will acquire. She thereby influences potential future returns to the educational investment and her ability to maximize her "utility," or satisfaction. The economic approach to understanding human behavior makes a number of assumptions about the way in which individuals conceive of their well being. Self-interest is conceived of broadly, beyond the pursuit of material concerns, to include a wide range of values and preferences. Individuals are considered to be forward-looking, to have consistent preferences over time, and to seek to maximize their welfare. There are a number of constraints on a person's capacity to pursue his or her self-interest and these include time, income, incomplete information, and lapses in judgment (Becker, 1996).

Altonji (1993) has highlighted the fact that individuals make educational choices under considerable uncertainty regarding their ability to complete a course of study in their selected field. His analysis (p. 51) models how "new information about preferences and academic performance, and new information about payoffs influence choice of major and the decision to stay in school." Within this human capital framework, as individuals gain new information, they make their curricular choices, transferring from one field to another or dropping out of college, based on an estimation of their ability to complete degree requirements. The probability of completion is influenced by their stock of knowledge, academic ability, and by degree requirements. The utility function indicated by Altonji's analysis also includes educational and occupational preferences and the present value of lifetime earnings.

Prior Research

BEST COPY AVAILABLE

In a 1991 article published in the *Journal of Economic Perspectives*, Sabot and Wakeman-Linn examined the influence of collegiate grading practices on student course choice. They documented the existence of grade inflation and compression (low variation) and also observed grading patterns that characterized high and low grading departments. They concluded that students face a disincentive to study in low grading fields, which, in their study of a small but varied sample of U.S. colleges, were predominantly quantitative fields. They found that economics, chemistry, and mathematics are consistently low grading fields, while art, English, music, philosophy, psychology, and political science are consistently high grading fields. In a survey administered to a small sample of English majors at a research university (Dowd, 1998), I also found that responding students believed that the average grades in biological sciences, physics, computer science, and chemistry at their institution was a B-; in political science, philosophy, economics, and mathematics a B; and in foreign languages, English, sociology, and history a B+. Consistent with Sabot and Wakemann-Linn's study, the low grading fields included quantitative subjects and the high grading fields included verbal subjects.

Davis (1966) argued that college students assess their areas of comparative advantage (where their skills and aptitudes put them ahead of their peers) based on the local competition for grades at their institution. Students then shape their career plans based on the feedback grades provide. However, Sabot and Wakeman-Linn (1991) observed that due to varying rates of grade inflation and compression among academic departments, "grades as a signal of relative strengths and weaknesses {are} more difficult for students to interpret." They noted (p. 167) that students do not adequately adjust their perception of differentially-scaled grades in order to gain a sense of their relative strengths and weaknesses, because "the incentive effects of absolute grades on course choice are far more powerful" than the indicators of comparative advantage that are weakened by non-uniform grading. Sabot and Wakeman-Linn argued that arbitrary differences in grading policies should be eliminated, because they provide incentives for some students to move away from academic areas where they are comparatively strong. Conversely, the effect of more-uniform grading policies would be to encourage greater numbers of students to take courses in the currently low grading departments, which are those that place emphasis on quantitative skills. While the labor market, through high earnings, provides an incentive to invest in quantitative study, under divergent grading—where quantitative fields are low grading relative to others—colleges create a disincentive to investment in quantitative study.

Divergent Grading and Labor Market Supply

The following simple utility-maximizing model extends Sabot and Wakeman-Linn's (1991) analysis to highlight the influence of non-uniform grading practices, where they exist, on the supply of college graduates with quantitative skills. The model is intended to facilitate a policy analysis of the implications of divergent grading for gender equity in earnings.

Under divergent grading practices, when a student decides in which fields of study to invest her time, she faces greater costs to obtain the valuables associated with college study in a quantitative rather than a verbal field. To obtain a certain number of credits in a quantitative rather than a humanities class with a grade of B would on average require more effort, because quantitative classes have lower mean grades. The relative costs of the effort to earn a degree through study in quantitatively or verbally oriented fields may be represented by the ratio E_Q/E_V, where E_Q represents the costs, psychic and otherwise, associated with quantitative study, and E_V represents the costs associated with verbal study. I assume that this ratio is fixed for each individual (disregarding the fact that costs would vary as students make marginal investments in either field).

We can also represent the ratio of the different compensation packages offered by employers to individuals with strong quantitative and strong verbal skills as W_Q/W_V. Again, I assume that this ratio is fixed. A forward-looking student with complete information about her future wage potential could determine whether to

invest in quantitative or verbal study by comparing W_Q/W_V and E_Q/E_V. If $W_Q/W_V > E_Q/E_V$, she would choose to invest in quantitative study. If $W_Q/W_V < E_Q/E_V$, she would choose to invest in verbal study, and if the two ratios are equal, she would be indifferent to these two options. For example, if the wage ratio is 2:1 (Q:V), then the student should invest her time pursuing quantitative study as long as earning credits in quantitative fields is less than twice as difficult (accounting for all costs, both psychic and material) as earning credits in verbal fields. The forward-looking student in this scenario would need to take into account lifelong earnings and career satisfaction, as well as the continuing education required to succeed at the occupations pursued.

The college's assessment systems and grading policies affect a student's decision to choose to study in a quantitative or verbal field by the fact that the differential between average grades in these two types of fields is one component (along with ability, motivation, and interest) establishing the ratio E_Q/E_V. As the differential increases, the value of E_Q/E_V also increases, and a greater number of students will determine it is not a wise investment to study in a quantitative field. In this way, the divergent grading system is a contributing factor determining the proportion of the population of college graduates who enter the labor market with quantitative skills. Student perceptions of the relative wages offered for quantitative and verbal skills also influence the proportion of students who enter different fields of study (as Freeman (1978) has illustrated with his cobweb model of curricular and career choice).

College graduates with different types of interests and abilities encounter different opportunities in the labor market. As strong quantitative skills are scarce relative to strong verbal skills, quantitative skills are compensated at a higher rate in the labor market than are verbal skills. Recent studies indicate earnings advantages over comparison groups of humanities and education majors of 23% to 61% for engineers, up to 25% for business majors, 13% to 35% for students of mathematics and the physical sciences, and 8% to 24% for social scientists (Angle & Wissmann, 1981; Berger, 1992; Bishop, 1994; Daymont & Andrisani, 1984; Eide, 1994; Griffin & Alexander, 1978; James & Alsalam, 1993; Rumberger & Thomas, 1993; Sharp & Weidman, 1989). When students are influenced by divergent grading practices to invest in verbal skills rather than in quantitative skills, the supply of verbal skills provided by college graduates to the labor market increases over the supply of graduates who would have made this choice, given their aptitudes and interests, under uniform grading practices. Labor economic theory indicates that the impact of this supply shift would lead to a decrease in wages paid to graduates offering verbal skills to employers (Ehrenberg & Smith, 1993).

Influences on the Curricular Choices of Women

Divergent grading leads to a greater quantitative- skills deficit among women than among men for several reasons. The first relates to the distribution of quantitative skills among men and women. In the population of college-bound high school graduates, women are less likely to be among those with the strongest quantitative skills. In addition, the measured quantitative and verbal skills of men show greater variance than that of women (Cole, 1997), and those students at the tails of the quantitative and verbal skills distribution are least affected by divergent grading. Students who have average skills in both quantitative and verbal fields are those who are most likely to receive misinformation about their comparative skills advantage as a result of low grading in quantitative fields and high grading in verbal fields. On the basis of their abilities, these students should be indifferent regarding choice of field. However, the degree of misinformation they receive is the full difference between average quantitative and verbal grades, and they are then motivated to choose verbal fields. Students with close to average quantitative and verbal skills are also likely to receive erroneous feedback. Students with a quantitative/verbal skills differential so large that the grading differential does not change the direction of the signal regarding their area of comparative advantage are not affected.

Second, women may be more affected by the quantitative/verbal grading differential because they may already face higher costs of study in quantitative than

in verbal fields as a consequence of participating in a learning environment that is oriented toward men. Sandler, Silverberg, and Hall (1996) have described a "chilly classroom climate" for women, which is exacerbated in traditionally male fields. In such a climate, women would experience psychic costs as they find their intellects and class contributions devalued. In particular, the competitiveness of study in quantitative fields relative to verbal fields may create high costs for women who pursue quantitative study (Dowd, 1998; Strenta, Elliott, Adair, Matier, & Scott, 1994). Even when women have equal measured abilities and aptitudes as men in quantitative fields, they have been found to enjoy science courses less than their male counterparts and to choose at greater rates to exit the field (Ware, Steckler, & Leserman, 1985). Prior research has shown that women persist in quantitative fields at greater rates if they attend women's colleges (Jacobs, 1996; Solnick, 1995), which suggests that women find a more welcoming environment in all-female classes, experiencing lower costs than those imposed by a male- centered environment. However, the findings on the effect of women's colleges on female educational attainments are not conclusive (Riordan, 1994; Smith, Wolf, & Morrison, 1995).

Finally, women may also give greater weight in making their curricular choices to their present or "local" status, to use Frank's term (1985), in the collegiate environment than to their future economic status. Loury (1997) found that women are less motivated than men by the college wage premium in making the decision to attend college. Frank (1996) and Daymont and Andrisani (1984) found that women place greater value than men on moral and personal dimensions of career satisfaction. These findings suggest that women are less concerned than men with future monetary returns to education. This disinterest may cause women to spend less time acquiring information about salaries and to underestimate the relative economic returns to quantitative and verbal fields of study. Disinterest may also be fostered by greater uncertainty concerning labor market participation, due to the fact that child-rearing responsibilities often interrupt women's careers. As Polachek (1981) observed, the prospect of discontinuous employment may provide an incentive for women to acquire human capital that does not depreciate quickly during their time outside the labor force and lead them to avoid rapidly changing technological fields. However, England (1982) countered that available data do not support this hypothesis.

The Higher Education Production Function

The discussion above has shown that divergent grading creates a disincentive to study in quantitative fields. Further, it demonstrates that these disincentives are likely to have a greater influence on the curricular choices of women than of men. At this point, beginning with an overview of relevant aspects of several theories of the higher education production function, I evaluate the factors creating patterns of low and high grading in quantitative and verbal fields of study. The need for and purposes of grading can be understood as part of a higher education production function, and the existence of divergent grading practices suggests that quantitative and verbal fields experience a different kind or degree of pressure to produce grades.

Production functions consider the outcomes of schooling as educational "outputs" resulting from various inputs including faculty, quality of students, and physical and financial capital. The demand for these outputs, which include teaching, research, and public service, comes from students, private and public funding agencies, and donors (Garvin, 1980; Hopkins, 1990; Hopkins & Massy, 1981; James, 1990). Production functions typically are based on the assumption that the goal of a private firm is to maximize profits. It is further assumed that market forces create an imperative that firms produce at the most efficient technological boundary of production. These assumptions do not apply to higher education, however, and in modifying the production function model for the higher education context, researchers have proposed several other objectives, including the maximization of administrative scope, income, and prestige. The role played by grading in the production function varies depending on the outcome to be maximized.

Niskanen (1971) described universities as "mixed bureaus," non-profit

BEST COPY AVAILABLE

organizations with public and private characteristics, due to the fact that they are funded through grants as well as through revenues generated by selling their output at a per-unit rate. He viewed universities as income-maximizers, whose administrators and faculty gain utility by increasing the size and scope of their bureaucracy. Breneman (1976) observed that faculty members seek to optimize departmental prestige, and Garvin (1980) elaborated on this and other research to develop a model of institutions as a whole as prestige maximizers. Faculty members gain utility from increasing levels of prestige associated with their departments in the form of higher salaries, better quality graduate students who can be attracted at a lower price, higher caliber colleagues, and greater success rates in seeking internal or external funding.

Zemsky and his colleagues drew on elements of the prestige- and bureaucracy-maximizing utility models to argue that faculty members increasingly expend their energies toward individual goals, away from the goals of the institution (Pew Higher Education Research Program, 1990; Zemsky, Massy, & Oedel, 1993). They attribute this phenomena to misplaced incentive structures that motivate faculty to focus on their research at the expense of teaching and advising. Faculty members maximize prestige in their disciplinary labor market by publishing academic papers. Teaching, the quality and value of which is difficult to present to external observers, carries little reward, they argued.

The Demand for Grades

The prestige- and bureaucracy-maximizing production models of higher education provide a theoretical basis for examining the characteristics of high and low grading departments. In this section, I extend these models to explain the pressures on departments at four-year research institutions to adopt high or low grading practices. I also use a utility maximization analysis to describe the interests students have in the prestige of their institutions and the demand they create for grades.

As Breneman (1976) and Garvin (1980) have illustrated theoretically and empirically, departments at research universities maximize prestige through research and scholarly output. They can increase their output by hiring very productive faculty members or by increasing the total number among the faculty. As increasing student enrollments provide a rationale for additional faculty hiring, there is a derived demand for a larger quantity of students. As faculty members prefer to work with talented students, there is also a demand for higher quality students. When departments attract external research funds from the government, foundations, or corporations, they can afford to lose a share of university resources allocated on the basis of student enrollment. The availability of external funding creates pressure to "weed out" less talented students and reduce enrollments. Departments that attract a lesser share of external research dollars will attempt to maximize enrollment, a goal that would relax pressures for competitive grading practices intended to dissuade the least capable students to leave the field.

Under certain conditions, students themselves create a demand for competitive grading, in a way that the other agents in the higher education output-demand system do not. Funding agencies, such as the government and foundations, are primarily interested in the outputs of research and teaching, as they make investments in higher education to develop public goods and promote social welfare. For students, higher education is both a consumption and an investment good (Schultz, 1961). The immediate value of their consumption is affected by the quality of teaching and learning, including modes of assessment. The value of their investment benefit is influenced by the status of their college (Heath, 1993).

Heath (1993) has illustrated theoretically that students value both local and global status, where local status is defined as a student's academic standing at her institution. As was discussed above, local status informs a student's understanding of the investment costs of completing a degree in any given field of study (Altonji, 1993). Local status also has psychic costs and benefits (Frank, 1985) and contributes to determining the consumption value of a student's education. In Heath's analysis, global status is determined largely by a college's ability to place graduates in high paying occupations and in graduate and professional programs. Global status is

influenced by an institution's academic rigor and the quality of enrolled students, with greater rigor attracting an academically talented student body. Students value the positive effects of higher standards on their global status, but fear the potentially negative effects on their local status and the increased costs of completing their work.

Student interest and influence on collegiate grading practices stem from their investment and consumption decisions. Students can be expected to endorse competitive grading practices, in which performance is graded on a curve and where average grades are low relative to other fields, if they perceive that such practices enhance their global status and ability to compete for high paying jobs. Students who are competing for scarce places in lucrative professions will have the greatest concern for global status. Under heavy interest, access to an occupation becomes limited and institutions have a prestige-maximizing incentive to certify only a portion of their students for entry into that field. The response to this incentive is the adoption of assessment practices that are designed to motivate or require those who are least capable to leave the field of study (Breneman, 1976).

Students who are not career oriented and who place a greater value on higher education as a consumption good can be expected to resist competitive grading and to avoid such practices when making their course choices, because it imposes immediate psychic costs and reduces the consumption value of their classes. If a field of study does not provide a closely articulated link to lucrative and competitive career paths, students will demonstrate a lack of interest in the credentialing function of grades. These students may value grades intrinsically as a reflection of their talents, but they do not create a demand for comparative rankings. In the absence of preprofessional student pressures, the field has an income- and resource-maximizing incentive to become high grading in order to attract enrollment.

In summary, the prestige-maximizing and bureaucracy-maximizing model of higher education production provides a theoretical basis for understanding the pressures on collegiate grading practices. External research dollars enable departments to maximize prestige and income while "weeding out" the least successful students from their programs. Student careerism also creates pressures for competitive grading, as students wish to enhance their global status. The model clearly predicts the behavior of departments experiencing a combination of low student careerism and low external funding (high grading practices) and high careerism and high external funding (low grading practices). As quantitative and applied fields are influenced much more greatly by research interests and strong links to employers than are arts and letters fields (Becher, 1989; Breneman, 1976), they are more likely to adopt low grading practices to maximize prestige. Verbal fields, with weak ties to employers and low levels of research funding, are more likely to adopt high grading practices to maximize administrative scope and enrollment.

Traditions of Scholarship

The educational production function identifies the utility-maximizing goals of scholars in different disciplines and provides a model that predicts grading practices in response to different output-demand systems. Internal features of departments stemming from disciplinary traditions and epistemologies may also account for different assessment practices. In Academic Tribes and Territories, Becher (1989) characterizes modes of scholarship in academic disciplines. His four-part taxonomy of "hard pure," "hard applied," "soft applied," and "soft pure" fields bears resemblance to the simpler quantitative/verbal dichotomy I have used. Hard fields are quantitative and soft fields, which include the humanities, social sciences, and "social professions" (education, social work, law), may or may not employ quantitative analyses. The applied fields, whether hard or soft, are those whose research practices are influenced strongly by practitioners and a search for practical knowledge. Becher's applied fields are those I have described as having ties with employers. Whether these employment relationships influence grading practices depends on the level of competition among students for entry into related occupations and professions. These relationships can be influential in a hard applied field, such as engineering, as well as in a soft applied field, such as business.

As Becher (1989) indicates, the modes of scholarship in the applied fields

follow from those of their pure counterparts, but are altered by the focus of applied fields on generating solutions to practical issues outside academe. For this reason, the epistemological distinctions that Becher observes between hard pure fields (natural sciences and mathematics) and soft pure fields (humanities and social sciences) describe the predominant disciplinary traditions and cultures that may influence grading practices. He offers a set of contrasts that, in sum, indicate that hard pure fields have a more clearly defined body of knowledge than the soft pure fields. First, Becher (1989, p. 13) observes, knowledge in hard pure fields is cumulative through the work of generations of researchers building on each others' findings relevant to clearly defined and bounded questions. In contrast, soft pure fields address issues that retain their currency over time. Researchers in soft pure fields make contributions, not by generating new knowledge, but by providing insights into familiar topics. Soft pure fields lack the clear boundaries that specify areas of investigation in hard pure fields. Second, while hard pure fields "break down complex ideas into smaller components," in soft pure fields "complexity is regarded as a legitimate aspect of knowledge, to be recognized and appreciated" (p. 14). Third, in hard pure fields, scientists make "strong" arguments based on mathematical models, measurement, and observed regularities. In soft pure fields, where explanation revolves around numerous concepts and the absence of clearly defined variables, scholars make apparently weak arguments and rely more heavily on "judgment and persuasion" (p. 14). Finally, soft pure knowledge recognizes and admits the "intentionality" of the scholar, while hard pure fields convey knowledge as "impersonal" and "value-free" (pp. 14-15).

Becher, himself, does not comment on differences in assessment practices between fields. This likely stems from the fact that participants in his case study at "elite departments" defined their membership in their academic professions "in terms of excellence in scholarship and originality in research, and not to any significant degree in terms of teaching capability" (p. 3). For this same reason, grading practices may be given peripheral attention, be little affected by disciplinary norms, and be easily modified by external influences. Or, they may follow closely from the research traditions. In the latter case, the openness of soft pure fields to divergent viewpoints combined with the acceptance of unresolved complexities in subject content would be consistent with assessment practices that allow numerous "correct" answers. In contrast, hard pure fields would be expected to rely on assessment practices that test students' abilities to convey their understanding of established subject content and to make greater distinctions between right and wrong answers. Low grading practices in hard pure (quantitative) fields and high grading practices in soft pure (verbal) fields may, therefore, have epistemological roots. This explanation is not completely persuasive, however, because the soft pure fields awarded lower grades on average in earlier times (Kuh & Hu, 1999; Sabot & Wakeman-Linn, 1991). Understanding of the relative influence of external demands and internal traditions of scholarship on assessment practices would require a study of changes in external and internal departmental environments in relation to changes in grading over time. To my knowledge, such a study has not yet been conducted.

Empirical Tests

Though little research has been conducted that tests the predictions of the production function model of grading practices, two recent studies present relevant findings. Freeman (1999) investigated the predicted relationship that departments with graduates entering lucrative professions have low average grades. He hypothesized (p. 344) that "given equal money prices per credit hour across disciplines, departments manage their enrollments by 'pricing' their courses with grading standards commensurate with the market benefits of their courses, as measured by expected incomes." Using data from the National Center for Education Statistics on 648 U.S. institutions of higher education, he confirmed that fields associated with higher starting salaries had lower GPAs than those associated with greater "income risk" (p. 350). His research provides evidence that departments manage student enrollment through their grading practices. Those experiencing higher student demand due to positive salary prospects for graduates are more likely to grade more rigorously. Freeman's work did not also estimate the influence of

available research dollars on grading practices.

Kuh and Hu (1999) investigated the causes of grade inflation from the mid-1980s to the mid-1990s, providing evidence that average grades have increased during that time period. However, their models do not include variables representing changes in labor market returns to field of study or changes in availability of externally funded research dollars, so the work does not provide a test of the production function model of grading practices. Their results do provide some relevant empirical evidence to evaluate the model, however. Using a large national data base including students from approximately 600 four-year colleges and universities, they find (p. 306) that grades in the humanities increased at a faster rate than grades in science and mathematics, with the grade increase in the science and mathematics cluster observed to be minimal. This finding supports the idea that quantitative fields, which have greater opportunities to attract research support, are resistant to inflationary pressures on grading. Grades in the social sciences and preprofessional fields were on average lower than those in science and mathematics, which, if the included social sciences were applied fields, supports the aspect of the model that indicates that preprofessional students will create a demand for rigorous grading.

In addition, Kuh and Hu found (p. 304) that while "grades increased across the board the increases were greatest at [research universities]," which suggests that some fields at research universities felt the greatest pressure to increase grades. Under the production function model, these fields are expected to be those attracting few external research dollars, though they could only have had the observed impact on the average grades if they were, indeed, departments with high enrollments. However, disaggregating the broader results, Kuh and Hu find (p. 314) that grades in general liberal arts colleges and in the humanities and social sciences were actually deflated in private institutions during the period under study. These findings may provide evidence contradictory to the production function model. Alternatively, they may indicate that humanities and social science fields without a significant preprofessional student body do not assume inflationary practices unless they are in a competitive situation with low grading preprofessional and research- oriented fields, which are more likely to be found at public and research universities. The latter interpretation of their results is appropriate if the sample included a significant number of private liberal arts colleges among the private institutions, but it is not possible to draw this conclusion from the article.

Discussion

The existence of divergent grading indicates that high grading and low grading departments are subject to different output-demand systems for grades. Institutions themselves are not likely to insist on uniform grading practices across their departments without a change in that demand system. If we assume that departments are maximizing their utility under existing practices, from what quarter might change toward uniform grading come? As discussed above, students, with their sometimes conflicting interests in global and local status, and agencies such as corporations, foundations, and the state, with their interests in the outputs of research and teaching, are the primary consumers of higher education. In this section, I discuss the potential motivations of the state and of students to create a demand for change. Foundations with an interest in social justice and economic development may play a role analogous to that of the state discussed below. Corporate sponsors of research will be most interested in private returns to their investments, but corporations too have an interest in an adequate supply of college graduates who have quantitative training.

As a matter of social justice, the state has an interest in promoting equal employment opportunities for women. As a matter of economic development, it has an interest in encouraging women to develop human capital in quantitative fields if market mechanisms are not providing an adequate incentive. Through research grants and internship programs, in its role as an employer, and through direct funding of colleges and universities, the state creates a demand for research and teaching. Through specialized programs, it structures some of that demand to create opportunities for women. These opportunities do not attract as many women in the

presence of divergent grading as they would under uniform grading (as some women continue to choose verbal fields despite the offer of an incentive, due to the higher cost of earning a degree in a quantitative field). The state could potentially increase the enrollment of women in quantitative fields by putting regulative pressure on colleges to adopt uniform grading practices.

However, as Strike (1997) has argued, when state regulatory processes require educational institutions to promote human capital formation as the goal of schooling, the resulting regulations promote a particular conception of what constitutes a good life. Such an intrusion as defining human capital formation as the goal of education, at the exclusion or expense of other legitimate schooling goals, is beyond the purview of the state. Colleges and universities do not have an obligation to motivate female students to plan their educational investments with an eye toward future economic success. The traditional liberal arts curriculum has been intended to produce people who are "virtuous, of good taste and liberated interests" (Strike), not people whose educational and life goal is to attain high earnings. Liberal arts colleges may very legitimately wish to structure the curriculum, including grading practices, to require or encourage students to take liberal arts courses. If liberal arts colleges choose to promote enrollment in liberal arts courses by intentionally lowering the psychic costs of study in those courses, that approach may well be consistent with institutional goals. Pressures for uniform grading might therefore come from the state, not in a regulatory mode but in its capacity as a consumer. The state addresses its human capital concerns by supporting educational programs that provide training in areas it deems valuable, thereby increasing the attractiveness of those areas to prospective students (by reducing associated tuition costs or by providing enhanced instructional facilities, for example). To further increase enrollment of women in quantitative fields, the state could attempt to alter aspects of the learning environment in those fields that create greater costs for women than for men. As competitive learning environments appear to place a particularly onerous burden on women (Dowd, 1998; Sandler et al., 1996; Strenta et al., 1994), the creation of non-competitive workshops, internships, research projects, or other opportunities of this type may serve to attract women to the study of mathematics and science. Non-graded instructional programs in quantitative fields could rely on other types of assessment to provide students with an incentive to learn the material presented. Such programs would provide certification of the attainment of threshold levels of knowledge, but would not provide comparative rankings. The instructional program would be structured to allow students multiple opportunities, as needed, to acquire the skills and knowledge necessary to capitalize on their investment in the labor market. Such an approach may be less efficient than using competitive grading to identify the most able students, but may be more efficient in fostering occupational gender equity. Astin (1990) has advocated a "talent development approach" to assessment in higher education, arguing for noncompetitive assessments on the basis of both equity and efficiency.

Demand for competitive grading in verbal fields might be created by trends in student enrollment. As the human capital model indicates, both grades and the present value of lifetime earnings are part of the equation determining the best human capital investment for a particular student. If the earnings associated with verbal fields of study fell so low as to outweigh the benefits of high grading, enrollment in verbal fields would fall. In that case, colleges might seek to create better links with employers for liberal arts graduates in order to place graduates in higher paying positions and to bolster enrollments. One way to establish these links would be to take an active role in supplying the most talented students to those labor markets. Such an approach would lead to comparative grading practices that would bear more resemblance to grading practices in quantitative fields.

Alumni donors might support such developments, because the increased success of graduates in the labor market would enhance institutional prestige. As Heath (1993) observed, alumni benefit most from increases in an institution's prestige, experiencing positive benefits related to their alma mater's enhanced reputation, without having to pay the costs associated with the academic competition of a higher quality student body. Alternatively, alumni might decry the professionalism of liberal arts programs and oppose new practices. The effect of their influence would depend on whether alumni donations are of a sufficient

BEST COPY AVAILABLE

amount to motivate income-maximizing behaviors.

Liberal arts colleges and departments do not have an ethical obligation to ensure access to employment information for their students, but they may benefit themselves by enabling their students to more efficiently estimate their future utility and to make investments in course choices that will maximize their financial return. If the college's graduates are able to maximize their utility in the labor market at a higher level after having had access to employment information while in college, the graduates would be able to achieve higher levels of both income and career satisfaction. Such an outcome would increase alumni donations, as well as the demand from prospective students for a liberal arts education.

Conclusion

I have presented a theoretical model, based on various explications of a higher education production function, to explain the demand for college grades. I have described student assessment as part of the process of producing educational outputs. The practice of high grading in verbal fields and low grading in quantitative fields was placed in the context of the different levels of demand placed on those fields for the outputs of teaching and research. Low grading fields are predicted to experience high demand by preprofessional students for entry into occupations with scarce positions and/or a high demand for research. The opposite demand system would affect high grading departments. Students who are concerned with entering a lucrative and competitive profession will create a demand for rigorous grading as it contributes to the prestige of the institution and to their own "global status," or value in the labor market. Students who are less career-oriented will place greater value on the consumption benefit of a college education and be concerned with the quality of teaching and learning and the value of their own "local status," or academic standing. Evidence from prior research was presented to show that women are more influenced than men in their choice of major by local status concerns, leading them to disproportionately choose high grading verbal fields. Thus, divergent grading creates an incentive for women to under- invest in quantitative fields of study, and, thereby, contributes to occupational sex segregation and the gender pay gap.

Notes

1. See Hoenack and Weiler (1975) for a discussion of the potential impact on university administration of charging different tuition rates by field of study.
2. While this simple model refers to an either/or investment in two different kinds of study, the argument could be extended to evaluate marginal investments in quantitative and verbal subjects and to take account of the different returns to various subfields.
3. This article is based on the author's dissertation research.

References

Altonji, J. G. (1993). The Demand for and Return to Education When Education Outcomes Are Uncertain. *Journal of Labor Economics, 11*(1), 48-83.

Angle, J., & Wissmann, D. A. (1981). Gender, College Major, and Earnings. *Sociology of Education, 54*, 25- 33.

Astin, A. (1990). Educational Assessment and Educational Equity. *American Journal of Education*, (August) 458-478.

Becher, T. (1989). *Academic Tribes and Territories: Intellectual Enquiry and the Cultures of Disciplines*. The Society for Research into Higher Education and Open University Press.

Becker, G. S. (1976). *The Economic Approach to Human Behavior*. Chicago: University of Chicago Press.

Becker, G. S. (1996). *The Economic Way of Looking at Behavior: The Nobel Lecture*. Stanford University, CA: Hoover Institution on War, Revolution, and Peace.

Berger, M. C. (1992). Private Returns to Specific College Majors. In W. E. Becker & D. R. Lewis (Eds.), *The Economics of American Higher Education* (pp. 141- 171). Boston: Kluwer Academic Publishers.

Bishop, J. (1994). The Economic Consequences of Schooling and Learning (Draft report). Ithaca, NY: New York State School of Industrial and Labor Relations, Cornell University.

Breneman, D. W. (1976). The Ph.D. Production Process. In J. T. Froomkin, D. T. Jamison, & R. Radner (Eds.), *Education as an Industry* (pp. 3-52). Cambridge, Massachusetts: Ballinger Publishing Company.

Cole, N. S. (1997). The ETS Gender Study: How Females and Males Perform in Educational Settings. Princeton, NJ: Educational Testing Service.

Davis, J. A. (1966). The Campus as a Frog Pond. The *American Journal of Sociology* (July), 17-31.

Daymont, T. N., & Andrisani, P. J. (1984). Job Preferences, College Major, and the Gender Gap in Earnings. *Journal of Human Resources, 19*(3), 408-428.

Dowd, A. C. (1998). Influences on Curricular Choices in College: A Case Study of Female English Majors from a Human Capital Perspective. Unpublished Dissertation, Cornell University, Ithaca, New York.

Ehrenberg, R. G., & Smith, R. S. (1993). *Modern Labor Economics: Theory and Public Policy*. (Fifth ed.). New York: HarperCollins College Publishers.

Eide, E. (1994). College Major Choice and Changes in the Gender Wage Gap. *Contemporary Economic Policy, 12*(April).

England, P. (1982). The Failure of Human Capital Theory to Explain Occupational Sex Segregation. *Journal of Human Resources, 17*(3), 358-370.

Frank, R. H. (1985). *Choosing the Right Pond*. Oxford: Oxford University Press.

Frank, R. H. (1996). What Price the Moral High Ground? *Southern Economic Journal, 63*(1), 1-17.

Freeman, R. B. (1978). *The Overeducated American*. New York: Academic Press.

Freeman, D. G. (1999). Grade Divergence as a Market Outcome. *Journal of Economic Education* (Fall), 344- 351.

Garvin, D. A. (1980). *The Economics of University Behavior*. New York: Academic Press.

Griffin, L., & Alexander, K. (1978). Schooling and Socioeconomic Attainments: High School and College Influences. *American Journal of Sociology, 84*, 319- 47.

Heath, W. C. (1993). Choosing the Right Pond: College Choice and the Quest for Status. *Economics of Education Review, 12*(1), 81-88.

Hoenack, S. A., & Weiler, W. C. (1975). Cost-Related Tuition Policies and University Enrollments. *Journal of Human Resources, 10*(3), 332-360.

Hopkins, D. S. P. (1990). The Higher Education Production Function: Theoretical

Foundations and Empirical Findings. In S. A. Hoenack & E. L. Collins (Eds.), *The Economics of American Universities* (pp. 11-32). Albany: State University of New York Press.

Hopkins, D. S. P., & Massy, W. F. (1981). *Planning Models for Colleges and Universities*. Stanford, CA: Stanford University Press.

Jacobs, J. A. (1996). Gender Inequality and Higher Education. *Annual Review of Sociology, 22*, 153-85.

James, E. (1990). Decision Processes and Priorities in Higher Education. In S. A. Hoenack & E. L. Collins (Eds.), *The Economics of American Universities*. Albany: State University of New York Press.

James, E., & Alsalam, N. (1993). College Choice, Academic Achievement and Future Earnings. In E. P. Hoffman (Ed.), *Essays on the Economics of Education*. Kalamazoo, MI: W. E. Upjohn Institute for Employment Research.

Kuh, G. D., & Hu, S. (1999). Unraveling the Complexity of the Increase in College Grades from the Mid-1980s to the Mid-1990s. *Educational Evaluation and Policy Analysis, 21*(3), 297-320.

Loury, L. D. (1997). The Gender Earning Gap Among College-Educated Workers. *Industrial and Labor Relations Review, 50*(4), 580-593.

National Center for Education Statistics. (1997). *Women in Mathematics and Science* (NCES 97-982). Washington, DC: U.S. Department of Education.

Niskanen, Jr., William A. (1971). *Bureaucracy and Representative Government*. Chicago: Aldine Atherton.

Paglin, M., & Rufolo, A. M. (1990). Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences. *Journal of Labor Economics, 8*(1), 123-144.

Pew Higher Education Research Program. (1990). *The Lattice and the Ratchet* (Policy Perspectives Vol. 2, No. 4): University of Pennsylvania.

Polachek, S. W. (1981). Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure. *Review of Economics and Statistics, 63*(1), 60-69.

Riordan, C. (1994). The Value of Attending a Women's College: Education, Occupation, and Income Benefits. *Journal of Higher Education, 65*(4), 486-510.

Rumberger, R. W., & Thomas, S. L. (1993). The Economic Returns to College Major, Quality and Performance: A Multilevel Analysis of Recent Graduates. *Economics of Education Review, 12*(1), 1-19.

Sabot, R., & Wakeman-Linn, J. (1991). Grade Inflation and Course Choice. *Journal of Economic Perspectives, 5*(1), 159-170.

Sandler, B. R., Silverberg, L. A., & Hall, R. M. (1996). The Chilly Classroom Climate: A Guide to Improve the Education of Women. Washington, D. C.: National Association of Women in Education.

Schultz, T. W. (1961). Investment in Human Capital. *The American Economic Review, 51* (March), 1-17.

Sharp, L. M., & Weidman, J. C. (1989). Early Careers of Undergraduate Humanities Majors. *Journal of Higher Education, 60*(5), 543-64.

Smith, D. G., Wolf, L. E., & Morrison, D. E. (1995). Paths to Success: Factors Related to the Impact of Women's Colleges. *Journal of Higher Education, 66*(3), 245- 266.

Solnick, S. (1995). Changes in Women's Majors from Entrance to Graduation at Women's and Coeducational Colleges. *Industrial and Labor Relations Review, 48*(3), 505-514.

Strenta, A. C., Elliott, R., Adair, R., Matier, M., & Scott, J. (1994). Choosing and Leaving Science in Highly Selective Institutions. *Research in Higher Education, 35*(5), 513-547.

Strike, K. A. (1997). Centralized Goal Formation and Systemic Reform: Reflections on Liberty, Localism and Pluralism. *Educational Policy Analysis Archives, 5*(11).[Available online at http://epaa.asu.edu/epaa/v5n11.html]

Ware, N. C., Steckler, N. A., & Leserman, J. (1985). Undergraduate Women: Who Chooses a Science Major? *Journal of Higher Education, 56*(1), 73-84.

Zemsky, R., Massy, W. F., & Oedel, P. (1993, May/June). On Reversing the Ratchet. *Change, the Magazine of Higher Learning, 25*, 56-62.

About the Author

Alicia C. Dowd

Department of Education
415 Kennedy Hall
Cornell University
Ithaca NY 14853

Email: ad19@cornell.edu
Voice: 607 255-7755
Fax: 607 255-7905

Alicia C. Dowd is Senior Lecturer in the Department of Education, Cornell University. Her research and teaching focus on educational equity and access. Current research projects include studies of the influence of tuition and tuition subsidies on the educational attainments of community college students; student equity in community college financing in New York State; and the influence of school-college partnerships on human capital development and institutional change. She received her Ph.D. in 1998 from Cornell University, where she studied educational administration, with a focus on higher education. In 1998-99, she played a role in creating the Institute for Community College Development at Cornell University, which provides leadership development opportunities to community college administrators and faculty members. Prior to completing graduate studies, she was a staff member for over ten years at Cornell's School of Continuing Education and Summer Sessions, where she was involved in developing and marketing academic programs.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy AnalysisArchives* is http: epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (480-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary

Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William Mcinerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois- UC

Robert T. Stout
Arizona State University

BEST COPY AVAILABLE

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
aacosta@cucea.udg.mx

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Maria Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Hugo Casanova Cardiel
Universidad Nacional Autónoma de México

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

BEST COPY AVAILABLE

Este artículo ha sido consultado 384 veces desde el 14 de febrero de 2000

Education Policy Analysis Archives

Volume 8 Number 11 febrero 14, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES.**
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education.*

Globalización, Integración Macroregional y Políticas de Internacionalización en el Sistema Mexicano de Educación Superior

Sylvie Didou Aupetit
Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (México)

- INTRODUCCION: GLOBALIZACIÓN: ECONOMIA, SOCIEDAD, CULTURA Y EDUCACIÓN
- I. TLCAN Y EDUCACIÓN SUPERIOR EN MÉXICO: LA CONSTRUCCION DE UNA RELACION
- II. CONCEPCION Y GESTION DE LA RELACION EDUCACION SUPERIOR Y GLOBALIZACIÓN
- III. LA COOPERACIÓN ACADÉMICA
- IV. LOS INTERCAMBIOS DE RECURSOS HUMANOS
- V. OFERTAS DE CARRERAS O INTERNACIONALIZACIÓN DEL CURRICULUM
- VI. GLOBALIZACION MACROREGIONAL Y TRANSFRONTERIZACION
- VII. ACREDITACIÓN Y CERTIFICACIÓN
- VIII. CONCLUSIONES. EL DIAGNÓSTICO Y LAS PROPUESTAS
- BIBLIOGRAFIA
- NOTAS

Acerca del Autor

Sylvie Didou Aupetit
Departamento de Investigaciones Educativas. Centro de Investigación y de Estudios

Departamento de Investigaciones Educativas. Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional. México.

Email:didous@data.net.mx

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
aacosta@cucea.udg.mx

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doc.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umea University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

This article has been retrieved 1284 times since February 21, 2000

archives | abstracts | editors | board | submit | comment | subscribe | search | epaa.asu.edu

Education Policy Analysis Archives

Volume 8 Number 12 February 21, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Top-Down, Routinized Reform in Low-Income, Rural Schools: NSF's Appalachian Rural Systemic Initiative

Robert Bickel
Terry Tomasek
Teresa Hardman Eagle
Marshall University

This article has
● Commentary

Abstract

Since 1991, the National Science Foundation has funded fifty-nine state, urban, and rural systemic initiatives. The purpose of the initiatives is to promote achievement in math, science, and technology among all students, and to encourage schools and communities to secure the resources needed to maintain such outcomes. The Appalachian Rural Systemic Initiative (ARSI) is a six-state consortium which focuses these efforts on low-income, rural schools. The primary means of accomplishing ARSI's aims is a one-day-one-school site visit, called a Program Improvement Review, done by an ARSI math or science expert. The centrally important Program Improvement Reviews, however, seem to be premised on unsubstantiated assumptions as to the static, easy-to-understand, easy-to-evaluate nature of educational achievement in rural Appalachian schools. As a result, the Reviews resemble exercises in early-twentieth century scientific management, and are unlikely to enhance achievement in science or math. Consequently, even if there is merit to the commonsense human capital approach to economic growth and development on which systemic initiatives are tacitly premised, this first- person account makes a case that desired payoffs are unlikely to follow from the work of ARSI.

Efforts to promote economic development and eliminate poverty through investment in public education have a long history in the U.S. (See, for example, Bowles and Gintis, 1976; Kaestle, 1983; Perkinson, 1995; Spring, 1997; McMurrer and Sawhill, 1998). In recent years, such efforts have included special attention to elementary and secondary schooling in science, math, and technology (Ashton and Sung, 1997; Senate Committee on Labor and Human Resources, 1997). This emphasis is premised on the assumption that in an increasingly science-based, technology-intensive world, the economic well-being—perhaps even the simple survival—of individuals and entire societies requires ever-higher levels of pure and applied scientific and mathematical knowledge (Shapiro and Varian, 1998; National Council of Teachers of Mathematics, 1998; Reich, 1992).

The National Science Foundation's Systemic Initiatives

In line with this straightforward human capital theoretic point of view, since 1991 the National Science Foundation has funded fifty-nine state, urban, and rural systemic initiatives (National Science Foundation, 1999). The purpose of each systemic initiative is to promote education in math, science, and technology (National Science Foundation, 1994a).

Published research on the initiatives is hard to find, and evaluation reports are not available. The origin of the term "systemic initiative" remains unclear. NSF's recent request for proposals for "systemic initiative research" provides no insight as to the meaning of the concept (NSF, 1998).

The terminology may follow, however, from NSF's judgment that education involves entire communities (Shields, 1997). At its best, in this view, education in math and science focuses on everyday applications in communities where schools are located (National Science Foundation, 1994b). The communities themselves, in a reciprocal process, benefit from development of a technologically literate workforce (Consortium for Policy Research in Education, 1995).

NSF's Appalachian Rural Systemic Initiative (ARSI)

The Appalachian Rural Systemic Initiative, or ARSI, is a six-state consortium, covering all of the Appalachian region of the U.S. (Harmon and Blanton, 1997). Consistent with NSF's intent, ARSI's ambitious objective is to facilitate educational change in economically disadvantaged rural schools resulting in high achievement for all students in mathematics and science (National Science Foundation, 1997). This is to be accompanied by development of community resources to sustain educational improvements (Brown, 1996).

Program Improvement Review

The primary means of accomplishing ARSI's aims is the Program Improvement Review. Done by ARSI experts, typically retired teachers, the purpose of a Review is to identify strengths and weaknesses in schools' math and science programs and make recommendations for improvement. ARSI experts, thereby, are charged with helping low-income, rural schools make students more productively employable in a science-based, technology-intensive world. In doing this, ARSI experts aim to contribute to production of the human capital needed for the economic and social development of low-income rural areas

Will ARSI Promote Economic Development in Appalachia?

The uncomplicated human capital perspective on which ARSI is premised begs an important policy question. Specifically, can educational reform be used to drive a growth and development strategy whereby the availability of well educated prospective employees attracts employment- creating investments? A tenable alternative holds that economic development is a necessary *prerequisite* for effective educational change (Bickel and Spatig, 1999)

For present purposes, however, we will put this reservation aside and address a more manageable question: If the education-and-development assumptions on which ARSI is premised were undeniably correct, would ARSI accomplish its objectives?

A First Person Account

The following account is written from the vantage point of one who was first an ARSI expert-aspirant, then an ARSI expert writing his first Program Improvement Review, and finally an ARSI dropout. The descriptions of "shadowing," of neutral-site instruction, of report preparation, and of rejection of the ARSI model are based on work done as part of the process of bringing ARSI to West Virginia under the auspices of the regional university with which the paper's authors are affiliated. Participation in this endeavor leads to the following inferences:

> ARSI experts construe the process of educational achievement as a thoroughly understood, relatively simple mechanism manifest in static indicatiors of school effectiveness.
>
> In consequence, ARSI has standardized and accelerated its centrally important Program Improvement Review process through excessively routinized observation based on short-cut procedures and unvalidated instruments.
>
> ARSI experts show no interest in substantiation of their evaluation criteria, but, nevertheless, take them for granted as embodying the one right way to teach math and science anywhere.
>
> Student engagement and student-teacher interaction are irrelevant to ARSI evaluations. Departures from ARSI criteria, even in the presence of overwhelmingly favorable student responses, are negatively sanctioned.

The remainder of this article is devoted to clarifying these inferences based on a first-hand account of ARSI at work. Throughout, one important message seems clear: ARSI's Program Improvement Reviews in low-income, rural schools are unlikely to enhance science and math achievement or promote economic growth and development.

We attribute this unfortunate set of circumstances to specious assumptions as to the existence of a taken-for- granted, science-based rationale for the top-down routinization and streamlining of educational evaluation and practice. As a result, even if the commonsense human capital framework on which systemic initiatives are based were valid, ARSI's work would not facilitate their application.

A Checklist-Guided Audit

The Program Improvement Review takes the form of a one- day, one-expert school visit, yielding a checklist-guided audit, resulting in degree-of-compliance scores ranging from 1 to 5 on approximately seventy Likert items. The check list is called a "Consistency Rating Summary."

For example, when evaluating a math program, the first general heading is "Curriculum, " subsuming ten Likert items, the first being "1.1 The math curriculum is written and used in planning the instructional program." The remaining general headings are "Instruction," "Thinking Processes," "Equity and Diversity," "School Climate, " "Relevance" or "Connections," "Training and Development," and "Financial and Material Resources."

The total number of items varies slightly depending on the discipline, math or science, the grade level, and the state which provides the educational policy setting for the review. Minor variations in the wording of the general headings and individual items are geared to these same factors. For example, under "Instruction," the first item used in evaluating math programs in West Virginia elementary schools reads as follows: "2.1 Teachers use WV IGO's to guide their instructional practices." "WV IGO's" refers to state-mandated "Instructional Goals and Objectives," around which high-profile state achievement tests are organized.

Likert item scores are used to gauge specific strengths and weaknesses in a school's math or science program. Strengths reflect consistency with the ARSI model embedded in the "Consistency Rating Summary." Weaknesses reflect departures from the model. In practice, far more attention is given to weaknesses than to strengths.

In spite of the importance of the Consistency Rating Summary, the source of its ten headings and seventy items is not identified. Are they research-based? Are they reasonable inferences based on years of teaching experience? Are they established principles in math and science education? Is their appeal based on face validity among ARSI experts? Do they represent an identifiable educational philosophy or pedagogical model? Participants are not told. Literature is nowhere to be found.

NSF Standards

NSF has promulgated a detailed set of National Science Education Standards (National Research Council, 1996). In the course of conversation and training with ARSI experts, however, these are never mentioned. If the experts are aware of NSF Standards, they do not disclose this. If NSF Standards are a source for the Consistency Rating Summary, participants are not told. The absence of descriptive, evaluative, or any other sort of literature concerning the Summary is again conspicuous. ARSI experts occasionally make off-handed references to "constructivism," and they are fond of invoking the notion "hands-on." One might reasonably surmise, therefore, that these ideas, though they typically remain vague, are included in construction of the Consistency Rating Summary and the way it is scored. In the absence of pertinent literature, however, this remains merely plausible conjecture.

State Mandates

ARSI experts often refer to state mandates, such as West Virginia's Instructional Goals and Objectives, mentioned above, and the Kentucky Core Content for Assessment. Whatever the merit of these state-level mandates, their substance appears to have been another influence in construction of the Consistency Rating Summary, and affects the way it is applied. The heading emblazoned at the top of the Consistency Rating Summary may vary with the state in which it is being used, as in "KERActeristics of a Good Mathematics Program" used in Kentucky, or the "West Virginia Program Improvement Review Consistency Rating Summary for Mathematics."

Beyond these tentative inferences, however, no rationale for the instrument is provided. One is left with the impression that the Consistency Rating Summary may very well have been the product of brain-storming sessions. The outcome is an instrument which appears to be vaguely current and topically correct, but which, as an evaluation tool, is of uncertain value.

Consistency Rating Summary Validation

Similarly, the technical properties of the Consistency Rating Summary as a measurement tool are not reported, and may not have been investigated. Given organization of the instrument into ten sections, each subsuming six to ten items, one might reasonably surmise that a factor analysis would reveal ten identifiable subscales. If this is the case, however, results are not available. The same is true for routine reliability coefficients. In short, the psychometric properties of the instrument seem not to be known. The possibility that discussion of such properties might be pertinent, even essential, is not acknowledged by ARSI experts.

Reporting on a Program Improvement Review

The final report, usually written overnight and presented the next day, is organized around the same ten general headings and seventy Likert items. Since much more attention is given to weaknesses than to strengths, most reports do not address all general headings or all items, but only those deemed deficient.

Recommendations for change appear throughout the report. A recommendation pertaining to "Relevance", meaning "[relating] mathematical knowledge to students' goals and interests," for a middle school located in a low-income, rural district in West Virginia's southern coal fields reads as follows:

> "Make a concerted effort to display positive, engaging images of mathematics throughout the school environment, paying particular attention to highlighting student work that is creative (not just correct) . . ." [Emphasis in the original.]

Becoming an ARSI Expert

Training in doing the Program Improvement Review, including scoring the Consistency Rating Summary, usually begins with "shadowing," accompanying an ARSI math or science expert who is doing a Review. ARSI experts also provide training at neutral sites, relying heavily on videos prepared to meet their specific instructional needs. Limited role-playing is used as a means of readying prospective experts to present Program Improvement Review findings to school personnel.

Training is informal, with little or no direct instruction. Instead, the ARSI experts serve as models during shadowing, and provide illustrative opportunities to apply the ARSI model during training sessions. Total training time varies, usually ranging from two to three days. An experienced ARSI expert may also participate in the first Program Improvement Review done by a just-trained expert.

"Shadowing" in Chemistry 8-B: Deficient Instruction

To illustrate our claim that ARSI Program Improvement Reviews are unlikely to enhance achievement, we begin with a brief case study of shadowing. Two ARSI expert-aspirants, assisting in bringing ARSI to West Virginia under the auspices of the university which employs them, are observing the in-school work of an ARSI science expert at a small, rural, low-income elementary school in eastern Kentucky.

We first attend a chemistry class. The three of us open the front door to the classroom without knocking, walk to the rear without speaking, and sit in side-by-side desks, while the class goes on about us. Students seem uninterested in our intrusion. The teacher seems unconcerned, and she makes no effort to acknowledge our presence. Even though this elementary school goes through grade 8, chemistry, rather than, say, general science, seems out of place, too advanced for an elementary school. The class, moreover, is referred to as Chemistry 8-B. This, we learn, means that chemistry students are grouped or tracked, with the ostensibly more capable students located in section 8-A. Nevertheless, the approximately twenty-five students in section 8-B seem quite capable themselves. The teacher is reviewing chemical bonding, referring to positive and negative valences, what they mean with regard to the make-up of individual atoms, and how they govern the way different elements combine to form molecules. She makes occasional reference to a periodic table displayed within easy reach on the wall near the front of the room.

Desks are organized in traditional fashion, arranged in rows, all facing forward. The teacher's desk is in the front of the room in the middle, turned toward the students. The teacher stands slightly to the left of her desk facing the students and occasionally turning to the board or, less often, to the periodic table. The presentation, too, is traditional, relying largely on lecture and board work, with questions and responses to teachers' queries from students. The teacher speaks fairly rapidly. The substance of the class is in no sense trivialized to match the ostensibly limited capabilities of lower track students.

The material covered is high school chemistry, much as I remember it from the eleventh grade. The teacher, though, seems smarter and more articulate, explaining things more clearly than I remember mine doing decades ago. Her high expectations for students are genuinely taken for granted. None of the students stands out as a stellar performer or favorite. The teacher's high expectations seem to apply equally to everyone.

The truly remarkable things about the class are the students' responses. All white,

about half male and half female, they seem genuinely engaged. They attend single-mindedly to the teacher's presentation. The students, manifestly, are putting all their time on task. Not just any task, but the conceptually difficult, even esoteric task at hand.

The teacher asks questions fairly often. Answers are quickly forthcoming, spoken thoughtfully, usually confidently, without the formality of hand-raising. Students' questions are immediately acknowledged and answered in a business-like, though not unsympathetic fashion. The teacher, a woman of about thirty who seems obviously to enjoy what she is doing, tries various means of explaining the same difficult ideas, sometimes complementing her oral presentation with additional board work.

Students don't talk among themselves. Two girls on the teacher's right near the front of the room are an exception, but as they whisper, they look toward the chalkboard, and one points to a diagram that the teacher had drawn earlier, illustrating the bonding of sodium and chlorine. A male student near the rear of the room on the teacher's left has a persistent problem with understanding her explanation of positive and negative valences. He makes his difficulty conversationally evident:

"Yeah, but I still don't get it. The signs are the opposite . . ."

He makes his point, in the same conversational fashion, more than once:

"I still don't get it. Why isn't it negative . . .?"

The teacher explains again, varying her choice of words. She gives no evidence of impatience. She addresses the questioning student in a matter-of-fact, even collegial fashion. She moves on, still holding students' attention, and doing so effortlessly. She presents material with relaxed enthusiasm born of genuine interest. There is no exaggerated affect or undue dramatization as she continues with a traditional presentation of conceptually sophisticated material.

The puzzled student on the teacher's left remains confused about positive and negative valences, though the precise nature of his misunderstanding is still not quite clear. He remains engaged, however, and raises the issue yet again, without evidence of embarrassment or anxiety. The teacher stops and thinks, looks at her diagrams on the board, seems not to know what else to say.

A male student sitting to the immediate left of his confused colleague responds spontaneously and matter-off- factly:

"I think I see . . . try this."

I cannot hear what is said. After a brief exchange between the two students, the puzzled one addresses the teacher:

"If sodium is short an electron and it adds one, why isn't it negative?"

Implied in this question is a complementary query about chlorine: if chlorine has an extra electron and it gives one to sodium, why isn't chlorine positive? The source of the student's confusion is now clear. The +1 valence of sodium is determined in its free state, before it combines with chlorine to form table salt. The fact that it takes an electron from chlorine—in effect adds a negatively charged particle—does not make it negative. The fact that it has a place for an electron that it adds to its outer ring, however, does make it positive. And conversely with chlorine.

The nature of the difficulty having finally been clarified, the teacher is able to dispel the formerly puzzled students' misunderstanding. He is satisfied. The teacher and students continue in the same matter-of-fact but engaged manner which has prevailed from the beginning. One way to usefully characterize their approach and the nature of the affect which accompanies it might very well be "professional."

As an observer, I was stunned. How did the teacher manage to hold the attention and active interest of this B- level—or *any* level—eighth grade—or *any* grade—chemistry class—or *any* class—throughout an entire period in which she discussed, in traditional lecture form, chemical bonding? In a low-income, rural, K-8 elementary school in eastern Kentucky or anywhere else! Here, as best I could

determine, was science being taught and learned about as well as could be done. Since the aim of ARSI is to promote high achievement in math and science in low-income, rural schools, this, perhaps, was a model, though one that might prove difficult to codify.

NSF's National Science Education Standards, which may or may not be known to ARSI experts, are intended to enable educators to judge whether particular actions will serve the vision of a scientifically literate society (National Research Council, 1996). The actions of this teacher and her students emphatically did just this. Or so it seemed to me.

An Off-the-Cuff Evaluation of Chemistry 8-B

At the behest of the ARSI science expert, the three of us who had been observing left before the class was over. We had been in Chemistry 8-B for about twenty-five minutes. Going out the door at the front of the room, I said to the teacher:

> "We're leavin' 'cause we can't understand this stuff."

The teacher stopped in mid-lecture, looked at me while I was speaking, and an expression of uncertainty left her face as she smiled. She gestured toward her students and said with confident pride:

> "*They* can understand it!"

> "I can see that!", I replied, as I joined the other two observers in the hall.

As the three of us walked to the next class, the ARSI science expert, striding purposefully, leading the way, offered the following judgments:

> "They didn't understand a *word* she said." His tone was contemptuous. "She was way, way over their heads."

> "There was nothing to hold their interest, no manipulatives or anything."

> "The walls were just about bare. Not much about science on them . . . nothing at all about science careers."

> "She was traditional lecture the whole time. All content."

The other observer was non-committal, as if taking in what was being said but still processing it, neither concurring nor disagreeing.

There was a brief silence as we walked. Then I said, laughingly, "for what it's worth, she teaches just like I do, when I'm having a *really* good day." Neither the ARSI science expert nor the other observer acknowledged my comment. The ARSI expert led us into the next classroom. I felt sort of silly. Not because no one had acknowledged my response, but because I had felt the need to cover it with self-deprecating laughter.

It was clear that the ARSI expert had definite preconceptions as to what eighth graders could and could not handle. His conclusion that the students in 8-B chemistry had no idea what the teacher was talking about seemed wildly at odds with what I had seen and heard in the classroom. Even the puzzled student eventually understood, and he did so with the help of another student. His confusion, moreover, bespoke an understandable, even imaginative, failure to see the specific terminological conventions which were being employed. In a real sense, his confusion about terminological conventions actually reflected a clear understanding of the chemical bonding process itself.

The teacher's method of presenting the material was traditional, to be sure. The students participated freely, however, without fear and without required hand-raising. The teacher-student, and student-student exchanges were conversational and matter-of-factly animated. Students helped each other.

The Irrelevance of Students

For the ARSI expert, however, this relaxed, informal, traditional approach was inevitably ineffective. *It was abundantly clear that the living presence of students in the classroom was not essential to his judgments.* He seemed not to notice them, their engagement, or the informed nature of their exchanges with the teacher and with each other. The expert attended only to the teacher, her traditionally limited use of few instructional materials, and the dearth of wall posters.

One Best Way to Teach Science

The ARSI expert clearly judged himself to be in a position to evaluate *any* science teacher's performance without benefit of observing or otherwise evaluating student responses, to which he seemed oblivious. In this instance, he purported to know in a matter of minutes that the teacher was clue-less, and that students would not learn. Traditional lecture was bad. Absence of manipulatives was worse. "You can use them to build molecules," he assured us.

> "That's what she was trying to do, but it's something you have to get your hands on. There weren't even any [manipulatives] in the room."

Thin Description

The ARSI science experts' dismissive, almost angry assessment of the teacher's effectiveness bespoke a willingness to generalize from very limited information. His assumption, clearly, was that twenty-five minutes of haphazardly selected, barged-in-on class time enabled him to produce an accurate typification of the teacher's performance and students' consequent achievement.

His harsh judgments, moreover, seemed inconsistent with NSF's position that science teaching and inquiry can be effectively done in a variety of ways (National Research Council, 1996; also see National Science Teachers Association, 1998). But once again, the connection between NSF and ARSI may or may not entail a shared understanding about teaching science and math. NSF standards may or may not be known to ARSI experts. In any case, the experts do not mention them.

"Shadowing" in a Program Improvement Review Presentation

Three weeks after the visit to the eastern Kentucky K-8 elementary school, I was again involved as an observer. I was paired with the same ARSI expert-aspirant, shadowing another ARSI expert in another small, low-income, rural elementary school in eastern Kentucky. During an hour-long, late morning meeting, the ARSI expert presented his previous-day's findings to the school's principal and six teachers. The ARSI expert began with a weak, almost apologetic grin:

> "This isn't as bad as it looks. There are a lot of 1's, 2's, and 3's, but this can be fixed . . . a lot of it . . ." [His voice trailed off.]

Criteria used in selecting the six teachers present at the meeting were not specified. They and the principal, however, remained silent as the ARSI expert went over his largely unfavorable report.

> "There's no evidence of the importance of math. They come away thinking it's just what they do in school."

> "They don't create their own knowledge. There is a lot of mainly lecture in the classroom."

> "If you used field trips, they would be able to see math all around us."

"They don't see its importance for careers, and that it's rewarded."

The principal, in spite of the beating her school was taking, looked confident and even eager throughout, as if to say "we're professional educators sharing information. There's nothing personal about this. We're glad to hear from outside experts, and we'll benefit from it. Please go on." The teachers seemed affectively disengaged but dutifully attentive. They betrayed no emotion. They seemed to neither accept nor reject the ARSI expert's account.

Teachers' Informal Challenge

After the report was presented, with only a few words of perfunctory discussion, we went to lunch in the school's cafeteria. By chance, I stood in the serving line with two of the teachers who had attended the meeting. Female, white, in their late-forties to mid-fifties, the teachers pleasantly initiated a conversation by asking where I was from. We talked briefly about West Virginia and work I had done in a rural county there. I likened that to what was being done by ARSI in their school. This was followed by a brief what-do-we-say-now sort of silence.

By way of keeping the conversation going, I added that the West Virginia project had been a long one. One of the teachers asked how long. I replied that it had gone on for three years, relying heavily on repeated focus groups with a broad range of stakeholders, and on literally hundreds of visits to the three schools involved.

The teachers became more animated and emphatic. Speaking of the ARSI expert's report of instructional omissions and other deficiencies, they commented:

> "We do a lot of that stuff, but we don't do it all the time. He was only here for one day, for a few hours . . ."
>
> "He never came to my room. How could he know what we do?"
>
> "I never even knew he was here."

Clearly, in this low-income, rural elementary school in eastern Kentucky, teachers were challenging the assumption that ARSI experts' one-day site visits enable them to understand a school's math or science instruction. This assumption, nevertheless, tacitly under girds all ARSI Program Improvement Reviews.

In retrospect, it seems obvious that I invited this challenge from the two teachers with my mention of a three- year project in West Virginia. At the time, however, I was just awkwardly trying to hold up my end of a conversation. Moreover, the teachers' responses seemed genuine, something they were waiting for a chance to say. Perhaps I had given them a deserved rhetorical opportunity, rather than a naked invitation to engage in a defensive, self-serving polemic.

Training in Fixing Deficiencies: Conservation of Momentum

In addition to shadowing, the training of ARSI expert- aspirants includes neutral-site instruction offered by ARSI experts. As an example, ten ARSI expert-aspirants and a handful of interested onlookers met with an ARSI math and science expert at the small-city headquarters of a West Virginia regional education agency. It was the ARSI expert's aim to continue with the introduction of expert-aspirants to the ARSI approach to evaluating education in science and math.

ARSI Training Videos

A retired teacher, the expert relied largely on a series of videos intended to provide opportunities to illustrate the ARSI ethos in use. During one of the longer and more purposeful videos, a white female teacher in her late twenties is seen reviewing the concept "conservation of momentum" with her high school physics class. There are approximately twenty students, all of them are white, about evenly divided between males and females. Is this a functioning classroom, or something staged by ARSI to aid

in the production of new ARSI science and math experts? We are not told, and no one asks.

The students in the video are more or less attentive. The teacher's presentation is brief and seems to lack focus, perhaps because the video begins near the end of her explanation, immediately following an exercise with manipulatives. Oddly, there is no teacher's introduction to the video itself. It just starts. Whether or not this is an ARSI-staged video, the absence of an introduction, an explanation to students as to why the video is being shown, is disconcerting. After all, we are supposed to be engaged in the evaluation of instruction. Maybe the ARSI expert will use the teacher's failure to introduce her video as a painfully obvious illustration of the wrong way to do things, such as use audio-visual aids in explaining conservation of momentum.

The video is devoted entirely to cars crashing. Cars crashing into each other, cars crashing into telephone poles, cars careening off guard rails and rolling onto their roofs, cars going off the road and landing in ditches . . . . It is reminiscent of a demolition derby, but without a winner. It is not immediately evident to me that the video actually does illustrate conservation of momentum. The ARSI expert says nothing. The only sounds in our room, as in the classroom on the video, are made by crashing cars.

As we watch the students watching the video, they seem, for the most part, unmoved. The camera catches two male students sitting together laughing at one, seemingly unexceptional collision. The crashes, presumably, were staged. All the cars are from the middle and late '70's. The video is repetitious, it seems too long, there is no narrative, just wreck after wreck, one looking more or less like another.

Finally it occurs to me that conservation of momentum, as best I can remember from twelfth grade physics, is manifest in the cars' tendency to continue moving even after they run into something solid. Though this recollection, in retrospect, seems embarrassingly obvious, is it safe to assume that the students on the video made the same inference? After all, their teacher, much as our ARSI expert, provided no commentary. Is this an example of constructivism, of students constructing their own physical knowledge? When the conservation of momentum video is over, our video is over, too. If there was an in-class discussion of what the students had just seen, we didn't get to hear it. Employment of the video seems part of a badly disjointed instructional process.

Perhaps the point of all this has been self-evident to the other ARSI expert-aspirants. I am, however, surrounded by nine other adults, all involved in education in one way or another, some with backgrounds in science, but more from administration or higher education. I wonder how many know what conservation of momentum means. Even now, I'm not sure that I do. For all I know, my aforementioned recollection from twelfth grade physics was in error. After all, I may have confused conservation of momentum with "objects in motion tend to stay in motion . . .", or something like that.

I wonder how many of the others see the pertinence of a video of serial collisions to understanding conservation of momentum. Were they able to recall or construct their own physical knowledge? Or is this video as bad an instructional tool as it seems to be?

The ARSI expert has very little to say about the serial-collision video. For a moment, he seems at a loss. He passes up the opportunity to fault the teacher for not providing an introduction. He says nothing about the absence of a debriefing. Then, belatedly, he calls our attention to the fact that two male students had laughed:

> "You could see their interest. They weren't just being passive."

The expert says nothing more about the video. He has concluded, as far as I can tell, that it demonstrated students' engagement in the process of acquiring a clearer, deeper understanding of "conservation of momentum." Perhaps we really have seen the construction of physical knowledge. My colleagues and I are silent. In truth, the serial collision video seemed like a silly caricature of instruction with audio-visual aids, how to misuse them rather than use them. But the ARSI expert gives no evidence of sharing this view.

Training in Fixing Deficiencies: Getting "Down and Dirty"

In another instructional video, a white female teacher in her early thirties is standing in front of a class of elementary school students. We are not told the grade, but the children appear to be eight or nine years old. Once again, all the students are white. The classroom is organized in traditional fashion, with individual desks in rows and the teacher standing at the front of the room, her back to the chalkboard. The teacher has said only a few words, the point of her class has not yet become evident, when the ARSI expert interrupts while the video continues to run. He speaks emphatically and with excitement:

"Look at her! Look at her clothes! She prepared for this!"

In truth, I saw nothing distinctive about the teacher's clothing or appearance. She was dressed modestly, wearing an open jacket with lapels, a white blouse which buttoned at the neck, a just-below-the-knees skirt, and shoes with medium heels. Her clothing was well-suited to working as, say, a bank teller, a receptionist in a family dentist's office, or a casework supervisor in a state social welfare agency. Her hair was cut short, but not extremely so. It was neatly combed, but not stylishly done. She wore make- up, but there was nothing ostentatious or extraordinary about it. She looked like the girl next door, grown up and working for a modest living. But the ARSI expert did not see it that way. The fact that the teacher was presentable counted against her:

"She can't get down-and-dirty dressed like that."

"She didn't come to work."

These observations, coupled with his surmise that the teacher had come prepared to appear on a video, seemed to imbue the ARSI expert with a sense of discovery. His response to the video suggested that, perhaps, he had not seen it before. He was looking for something instructive, and quickly found it in the teacher's appearance, which still seemed unexceptional.

He judged a teacher's work as inevitably involving getting "down-and-dirty." Suitable clothes, I concluded, would have been faded jeans, a sweatshirt with holes worn in the elbows, and grass-stained tennis shoes. Why suitable attire for an elementary school teacher should take this form remained a mystery to me, just as the nature of "getting down and dirty" and why it was a pedagogical essential remained unexplained.

None of the prospective experts spoke. I saw two give obligatory grins at the "she can't get down-and-dirty dressed like that" judgment. Otherwise, the group was impenetrably difficult to read. Was the lesson clear? Did participants accept it? Did anyone find this informative? Did the ARSI expert know that NSF National Science Education Standards do not include a dress code? Was he aware that teachers' attire is often an issue in rural Appalachian schools because they sometimes dress *too* informally (Austin, 2000)? Is this what it means to become an ARSI expert?

ARSI in West Virginia

ARSI's first Program Improvement Review in West Virginia was done in mid-March of 1999. This was also the first time I worked as an ARSI expert. The same was true of my shadowing partner, who was serving as coordinator of our three-school review. Though newly-minted as an ARSI expert, he had long experience in grant writing, program development, and administration of ground-up educational change efforts. Early in his career, he had taught high school science.

Adaptation or Adoption

This Review, moreover, was to be different from those we had seen in Kentucky. It involved three schools rather than one. The schools, an elementary school, a middle school, and a high school, are in close geographical proximity to each other, situated in a low-income, rural district in the state's southern coal fields.

In addition, while the ARSI Program Improvement Review was being used as a

point of departure, it was not a governing model. The Consistency Rating Summary, replete with Likert items, was still there, but as only one source of information in preparation of a report which was to be tentative, formative, and qualitative.

Rather than one-expert school visits, as in Kentucky, there were four evaluators for each school. Most members of each team were newly-minted ARSI experts, who also had training and experience in a variety of pertinent disciplines, including assessment, math education, program evaluation, and administration.

Recommendations for improvement were to be made only *after* discussing the final report with a variety of local stakeholders from the three schools. Stakeholders would participate in the process of actually *producing* the recommendations.

Synthesizing a Final Report

My task was to synthesize a final report. The Consistency Rating Summary would have left little to synthesize, but its place was not central in West Virginia, as it had been in Kentucky. The materials for synthesis were submitted in manila folders, eleven-by-seventeen envelopes, three-ring notebooks, translucent zip-lock packets, and paper-clipped pages. Consistency Rating Summaries prepared by ARSI experts were included. The Summaries, however, were mixed in with field notes, handwritten reminders, and miscellaneous jottings on single sheets of paper. In addition, each teacher at each school had completed a Consistency Rating Summary, and these, too, had been included.

Some ARSI experts' Summaries had conspicuous marginal notes and some did not. Summaries for the same school included and excluded different headings and items. Some included experts' names and some did not. Some had a formal, finished appearance, while others looked like preliminary worksheets. In spite of our plan to make production of recommendations a collaborative effort with stakeholders, a few Summaries included recommendations. All tolled, however, the material did not resemble the output of the sort of mechanically routinized process of thin description we had seen in Kentucky.

A Formative Systemic Report

Since our Review involved three schools, a systemic report seemed in order. Furthermore, even though the schools were at three different levels, dramatic cross-school commonalities in traditional educational philosophy and old fashioned, no-nonsense practice made a single report seem fitting. The flexibly formative nature of the process was emphasized in the report's opening paragraphs under the heading "Informed Interpretation from Multiple Perspectives."

> "A good deal of what we have to say, moreover, is subject to good-faith interpretation and re- interpretation by stakeholders. . ."

Similarly, use of the Consistency Rating Summary was placed in context, subsumed by "Judicious use of a Quantitative Rating Summary":

> ". . . but one source of information for making formative judgments. Its . . scores . . . merely summarize some of the information used in making our essentially qualitative judgments."

A First Draft

The report characterized the math program in each of the schools as traditional, and noted that all adult stakeholders, teachers, administrators, and parents, preferred it that way. Parents were unaware of alternatives. Even some of the teachers were unfamiliar with current terminology and practice. When a newly-minted ARSI expert used the term "rubric," an elementary teacher asked what rubric meant.

The schools were autonomous to a fault. Though constituting a rudimentary feeder system, teachers and administrators had no cross-school contact. Insofar as their math curricula were cumulatively compatible, it was due to state and district requirements,

and adherence to the same traditional ethos and practices.

The report went on for twenty-seven double-spaced pages, addressing topics such as "Avoidance of Innovation," "Cautious Selectivity," "Exclusion of Exploration" Innovations Come and Go," "Traditional Parental Roles," "School-to-School Isolation," and "Staff Development and Teacher Traditionalism." The concluding sections re-emphasized the importance of understanding the report as interpretative and subject to legitimate challenge by stakeholders. Readers were reminded that formulation of recommendations was to be a collaborative effort.

"Their Nickel"

When I gave this determinedly formative report to my former shadowing partner, still coordinating this first West Virginia Review, his response took me by surprise. Noting the absence of a "Consistency Rating Summary," he said, "it's their nickel." In short, whatever liberties we took with the ARSI model, this remained an ARSI endeavor. ARSI was establishing itself in West Virginia under the institutional auspices of our regional university, and some ARSI expectations had to be met. In response, I used the diverse, unstandardized information which had been submitted, and tried to synthesize a set of defensible Likert item scores for the three-school system. Having attached this to the narrative, I thought the job was done. The coordinator agreed. He submitted a copy to West Virginia's first ARSI Collaborative Director, and scheduled a meeting.

Meeting with ARSI Officials

The meeting with the ARSI Collaborative Director and an associate began amicably. They had read the report, and they listened with what appeared to be friendly interest as we explained our plans to meet with stakeholders to collaboratively produce recommendations for change. I characterized the approach to Program Improvement Reviews in Kentucky as "take-it-or-leave-it," "expert-centered," "prematurely codified," "top-down," and "quick-and-dirty." The evolving West Virginia approach, by sharp contrast, was "flexible," "client-centered," "qualitatively formative," and "collaborative."

The Director responded by noting that there was only one Consistency Rating Summary for three schools. I replied:

> "Right. Like we said in the report, we took a systemic approach. It made sense, especially since the schools are so much alike."

The Director responded that there were no recommendations. I referred again to the report, noting that the recommendations were to be produced collaboratively with school-level stakeholders. The Director, still smiling, shook her head. She said:

> "The reports are standard. We need Summary scores for each school, and recommendations for each."

I responded that I had seen take-it-or-leave-it reports, loaded with misguided Likert-item claims to precision, done all too quickly during shadowing. They were the sorts of reports, I added, that later sat on shelves gathering dust, because stakeholders were not involved in their production. The Director replied:

> "I'm sorry if that was your experience."

She looked at her assistant and asked:

> "Is that the way you saw it when you made visits?"

The assistant shook her head and murmured unintelligibly. I returned to my characterization of what I had seen in Kentucky, including again "take-it-or-leave-it," "prematurely codified," and "quick-and-dirty." The Director responded:

"But that's just your opinion."

I snapped angrily:

"Of course! What else *would* it be?"

My shadowing partner intervened. He asserted that he had not expected to do Program Improvement Reviews exactly as they were done in Kentucky. He was especially concerned about formulating recommendations without collaboration with local stakeholders.

"They need to be involved in this process. They need a sense of ownership. Otherwise, the report will never be implemented."

The ARSI Collaborative Director was not persuaded. She said little, remained unflappable, and would not budge: ARSI Program Improvement Reviews were standard. I asked:

"What did you think of the text of the report?"

The Director and her assistant both nodded approval. Then the assistant added:

"It was long. People are busy . . ." (Followed by a conciliatory, partly muffled chuckle.)

I asked:

"What's missing from the report as it is now?"

The Director repeated that Consistency Rating Summaries and recommendations for each school were essential parts of any ARSI Program Improvement Review report. These, in fact, as submitted by the ARSI experts, *are* the report. I responded:

"So I just clip the three reports together? It's a clerical job?!"

The Director replied:

"Yes . . . in part."

I responded angrily:

"If I had known we were gonna do it this way, I'd never have gotten involved. This is the last one I'll do."

By this time, I had lost my composure, while the Director had retained hers. I left the room, acknowledging that ARSI would get the kind of report it wanted. That was the end of my involvement with the Appalachian Rural Systemic Initiative.

In Retrospect

It is worth noting that, until our Program Improvement Review, ARSI had kept a low profile in West Virginia. Unknown to me was an earlier series of three meetings with West Virginia educators hosted by ARSI representatives, one of whom is now the ARSI Collaborative Director. According to a participant, a former math teacher who is currently a professor of education and a co-author of this article, the meetings were held January through April of 1998. Her unsolicited invitation to attend described the first meeting as intended to explore "the development of a self-assessment instrument . . . to aid counties in:

- "Critically looking at their scien

Harmon, H. and Blanton, R. (1997) Strategies for Improving Math and Science Achievement in Rural Appalachia. Tucson, Arizona: National Rural Education Association Annual Convention.

Kaestle, C. (1983) *Pillars of the Republic*. New York: Hill and Wang.

McMurrer, D. and Sawhill, I. (1998) *Getting Ahead*. Washington, D.C.: The Urban Institute.

National Council of Teachers of Mathematics (1998) Principles and Standards for School Mathematics. Reston Virginia: National Council of Teachers of Mathematics.

National Research Council (1996) *National Science Education Standards*. Washington, D.C.: National Academy Press.

National Science Foundation (1994a) Foundation for the Future. Washington, D.C.: National Science Foundation Directorate for Education and Human Resources.

National Science Foundation (1994b) Building the System: Making Science Education Work. Washington, D.C.: National Science Foundation, Directorate for Education and Human Resources.

National Science Foundation (1997) Rural Systemic Initiatives in Science, Mathematics, and Technology Education. Washington, D.C.: National Science Foundation, Directorate for Education and Human Resources.

National Science Foundation (1998) Systemic Initiative Research. Washington, D.C.: National Science Foundation, Directorate for Education and Human Resources. URL http://www.nsf.gov/pubs/1998/nsf98140/nsf98140.htm

National Science Foundation (1999) HER Projects. Washington, D.C.: National Science Foundation, Directorate for Education and Human Resources, URL http://www.ehr.nsf.gov/project.asp

National Science Teachers Association (1998) The National Science Education Standards: A Vision for the Improvement of Science Teaching and Learning. Arlington, Virginia: National Science Teachers Association, URL http://www.natt.org/handbook/nses.htm

Perkinson, H. (1995) *The Imperfect Panacea*. New York: McGraw-Hill.

Reich, R.(1992) *The Work of Nations*. New York: Vintage Books.

Senate Committee on Labor and Human Resources (1997) Innovative Workforce Development Initiatives. Washington, D.C.: Congress of the United States, United States Senate.

Shapiro, C and Varian, H. (1998) *Information Rules*. Cambridge, Massachusetts: Harvard Business School Press.

Shields, P. (1997) Evaluation of the National Science Foundation's Statewide Systemic Initiatives (SSI) Program: First Year Report. Menlo Park, California: SRI International.

Spring, J. (1997) *The American School: 1642 to 1996*. New York: McGraw-Hill.

About the Authors

College of Education and Human Services
Marshall University
Huntington, West Virginia 25755-2480

Robert Bickel
bickel@marshall.edu

Robert Bickel is a Professor of Advanced Educational Studies at Marshall

University. His recent research is concerned with school size as a variable which moderates the relationship between social class and measured achievement, and with contextual factors which occasion the at-risk designation.

Terry Tomasek
Terry Tomasek has taught high school science, and is completing a Master of Arts in Teaching and a Master of Science in Biology at Marshall University. Her thesis research focuses on the effects of valley fill construction on aquatic environments.

Teresa Hardman Eagle
Teresa Hardman Eagle is an Assistant Professor of Educational Leadership Studies at Marshall University. A former high school math teacher, she is doing ethnographic research on women as administrators in educational institutions and social service agencies.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy AnalysisArchives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (480-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umea University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto *a* servidor unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
aacosta *a* cucea.udg.mx

Teresa Bracho (México)
Centro de Investigación y Docencia Economica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova *a* asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras *a* doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue *a* asu.edu

Maria Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb *a* orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis *a* filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Malaga
aiperez *a* uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatistica
simon *a* openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres *a* gseisucla.edu

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo *a* uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa *a* servidor.unam.mx

Hugo Casanova Cardiel
Universidad Nacional Autónoma de México

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein *a* luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent *a* gemtel.com.mx
kentr *a* data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr *a* servidor.unam.mx

Humberto Muñoz García (Mexico)
Universidad Nacional Autónoma de México
humberto *a* servidor unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky *a* oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo *a* udc.es

This article has been retrieved 1058 times since February 21, 2000

archives | abstracts | editors | board | submit | comment | subscribe | search | epaa.asu.edu

Education Policy Analysis Archives

Volume 8 Number 13 **February 21, 2000** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

A Statewide Professional Development Conference: Useful Strategy for Learning or Inefficient Use of Resources?

Paul V. Bredeson
University of Wisconsin-Madison

Jay Paredes Scribner
University of Missouri-Columbia

Abstract

In an environment increasingly skeptical of the effectiveness of large-scale professional development activities, this study examines K-12 educators' reasons for participating and beliefs in the utility in a large-scale professional development conference. Pre- and post-conference surveys revealed that while financial support played a significant role in educators' ability to participate, they were drawn to the conference by the promise to learn substantive issues related to, in this case, performance assessment—what it means, how to implement it, and how to address community concerns. In spite of the conference's utility as a means to increase awareness of critical issues and to facilitate formal and informal learning, well conceived linkages to transfer new knowledge to the school and classroom were lacking.

The professional development of teachers has increasingly been viewed as a fundamental ingredient of successful educational reform and local school improvement in the United States (Fullan, 1995; Little, 1993). For example, by the latter half of the 19th century normal schools and colleges in the US regularly offered summer workshops and institutes for teacher professional improvement. These included such opportunities as workshops, courses, in-services, training sessions, extension work, and internships designed to address the needs of teachers and implement local school, district and state education policies (Little, 1993). Both paradox and promise have helped forge the link between educational reform and training and development In

some cases the quality, training, and competence of education professionals have been viewed as a major obstacle to educational reform—one that needed to be remediated through prescribed training (National Commission on Excellence in Education, 1983; Smiley, 1996). In other instances, policy makers, researchers, and educators have argued that teachers are not the problem but rather the primary creators of solutions to the vexing problems that confront educators in a dynamic public education system serving a culturally diverse nation (Smylie, 1996; Bredeson, 1998; Corcoran, 1995; NFIE, 1996). Corcoran (1995) speaks directly to the promise of professional development in education. "It is now widely recognized that the success of these reform initiatives depends in large part on the quality and accessibility of professional development for teachers" (Corcoran, 1995, p. vi).

Even the casual reader of educational reform reports, legislative mandates, and contemporary educational literature would soon discover one common theme: professional development is critical to systemic educational reform and school improvement focussed on enhancing learning outcomes for all children in public education (Fullan & Hargreaves, 1996). Research has clearly indicated that teachers-as-learners are critical to pedagogical, social, political, and economic goals here in the US and other countries. For example, the professional development of teachers is offered as a primary educational reform strategy intended to help schools and teachers develop more rigorous curriculum standards, design meaningful educational assessments, facilitate organizational change, guide school improvement plans, and improve teachers' knowledge and skills to enhance student learning outcomes. These include calls to create stable, high quality sources of professional development (NCTAF, 1996); incorporate professional learning into the fabric of daily life in schools (NFIE, 1996; Scribner, 1999); establish professional development as a central component of state and local educational reform (Houghton & Goren, 1995); transform professional development to meet urgent educational needs (Corcoran, 1995); consider alternatives to traditional training models of staff development (Little, 1993); deal more directly with issues of racism and inequity in schools (Weissglass, 1997); and break the mold to classroom practices through new professional development practices (McLaughlin & Oberman, 1996).

Given the centrality of professional development to educational reform expressed in myriad activities, it is equally important to understand teachers' experiences with and beliefs about their own professional development (Darling-Hammond & McLaughlin, 1995; Lieberman, 1995). To address part of this larger question, this study examines educators' (including teachers, principals and specialists) experiences and beliefs as they pertain to one vehicle for professional development—professional conferences. While the limitations of conferences as a delivery mechanism for professional growth have long been extolled (e.g., Joyce, 1990; Little 1993), we examined participants experiences in one statewide (Wisconsin) professional development conference to more fully understand (1) the potential benefits of large scale professional conferences and (2) the influence these conferences may have on professional learning and the school change process. Specifically, we sought to answer the following questions: (1) what motivated participants to attend a large scale conference and what were their expectations; (2) what types of knowledge did participants acquire at the conference; and (3) what role may the knowledge acquired play in participant and/or school improvement?

Conceptual Organizers

We use two conceptual lenses to shed light on this study. First, we borrow Schlechty and Whitford's (1983) useful typology of professional development to examine the intended and unintended purposes and expectations inherent in large scale conferences. Second, we employ a professional knowledge framework to make sense of what types of knowledge these educators may (or may not) have learned in this setting.

A Professional Development Typology

Students of teacher learning have categorized professional development activities in different ways. Perhaps one of the most useful and enduring frameworks to examine

specific activities is Schlechty and Whitford's (1983). They described professional development activities as serving one or more of three functions: (1) an *establishment* function (e.g., increasing awareness) when the purpose is to promote organizational change through the implementation of programs, technologies, or procedures in schools and school districts; (2) an *enhancement* function (e.g., apply to and improve practice) to improve teacher effectiveness; or (3) a *maintenance* function (e.g., continued practice) to ensure compliance with administrative and organizational goals and objectives. Viewed through this lens a large scale conference such as the one examined here would be expected to best serve an establishing function.

Professional Knowledge

Implicit in most professional development endeavors is an expectation that knowledge acquired will be used in some fashion at a later time. In this realm, Eraut (1994) provided important frameworks to investigate and understand knowledge acquisition and use. Concerned not only with the relevance of the knowledge acquired, Eraut's work focuses on how knowledge is acquired and the relationship between knowledge acquisition and knowledge use. He argues that most professionals learn continuously, but he warns routine experiences do not necessarily add to the professional's knowledge base. Rather special circumstances or unique occurrences offer the most fertile grounds for adding to the professional's knowledge base. Furthermore, Eraut embeds the concept of the professional knowledge acquisition within the work context. Put differently, the nature of the professional's work plays a major part in determining what knowledge is learned, how it is learned, and how that knowledge is (or is not) used (see also, Scribner, 1999). On the surface, these ideas would seem to seriously limit the utility of large scale conferences conducted beyond the contexts of classrooms, schools, and districts.

Eraut and others (e.g., Marsick & Watkins, 1990) have also attempted to describe various types or classes of knowledge. Generally speaking, Eraut frames professional knowledge as a triad of propositional, procedural, and personal knowledge. Propositional knowledge includes academic knowledge, typically discipline-based, and theoretical knowledge. Propositional knowledge is concerned with describing actions and is often of little use to practitioners with immediate needs. Limitations placed on professionals by the context of their work often relegate theories (propositional knowledge) learned in the classroom to the mind's attic never to be retrieved. Procedural knowledge is "how-to" knowledge professionals develop that is needed to perform job tasks. Finally, personal knowledge includes "notes and memories of cases and problems which have been encountered, reflected upon and theorized to varying extents and with varying significance for current practice" (Eraut, 1994, p. 17). We kept these knowledge types in mind as we analyzed our data. By overlaying these two frameworks, we hope to shed new light on both the promise and persistent pitfalls of large scale conferences.

Methods

This evaluative study takes a utilization-focused approach (Patton, 1997) to address the research questions outlined above. Working closely with the Wisconsin Education Association Council (WEAC), a major sponsor and financial contributor to the conference, we designed an evaluation that would summatively show the merit (i.e., strengths and weaknesses) of such a large scale endeavor.

Event and Participant Selection

Due to the evaluative nature of the study and our close working relationship with sponsoring agencies (i.e., WEAC, Wisconsin Department of Public Instruction, and the University of Oshkosh), both the case (i.e., the conference) and the participants represent convenience samples (Bogdan & Biklen, 1998). The three-day conference on student performance assessment consisted of an array of workshops, round table discussions, work groups, consultation time with assessment experts, opportunities to work with school teams, presentations by invited speakers, as well as informal times for teachers to socialize and network with colleagues.

Conference participants represented the gamut of public education in Wisconsin including K-12 teachers, administrators, and other specialists, and higher education administrators. However, to answer the questions and concerns raised by the sponsoring organizations we only surveyed full-time, public school K-12 teachers, administrators and specialists. Furthermore, due to the nature of the population sampled two respondent "cohorts" were used. The first cohort represents all full-time teachers, administrators, and specialists (N=301). Surveys from cohort I inform the first research question and were administered during the opening session of the conference. Cohort II (N=101; a subset of the cohort I) represents those participants supported by WEAC to attend the conference. Along with financial support to attend the conference, professionals in cohort II were obligated to attend a post-conference meeting lasting two hours. At this meeting participants responded to our survey that addressed the second and third research questions. (Note: throughout the narrative below we will refer to our findings by cohort to avoid confusion).

Data Collection and Analysis

To address our research purposes we designed two written questionnaires to collect survey data from conference participants (See appendix). The first survey, which cohort I (N=301) completed at the outset of the conference, consisted of 1) demographic items, 2) information about whether or not they were part of a school team, and 3) how their expenses for the conference were covered to gain an understanding of participants' expectations of the conference and reasons for attending. We also asked a series of open-ended queries, including: 1) how they found out about the conference; 2) why the conference was of interest; 3) what they hoped to gain from the conference; and 4) what activities in the area of performance assessment were currently going on in their schools. Cohort II (i.e., participants whose conference fees were paid by WEAC) completed a post-conference survey that sought participant perspectives on actual conference benefits, the role WEAC sponsorship played in their attendance, and how the topics and activities were connected to assessment issues and activities in their schools.

We used two primary methods for data analysis. First, we completed descriptive and statistical analyses of all quantitative data. Next, narrative responses were transcribed and organized into text files by question. We then analyzed narrative data using a constant comparative method (Strauss & Corbin, 1990) in which we coded data, developed categories, and identified themes in the open-ended responses.

Limitations

This study focuses on the experiences of educators attending a three-day professional development conference; further the participants surveyed represent convenience samples. As such, the findings from this study are limited in their generalizability. Nevertheless, we believe these data do provide an understanding of several issues including but not limited to the following: (1) participants' beliefs about their own professional learning and the linkage between that learning and their work; (2) the expected and unexpected outcomes of large scale conferences; and (3) the ongoing tension between efficiency and effectiveness as they relate to professional development.

Findings

We begin our discussion of findings by describing the conference, its participants (cohort I), and their responses to the pre-conference written survey. Data analysis on cohort I responses led to the formation of two categories: *motivation* for attending a large scale conference and *utility* of a large-scale conference. Data analysis on cohort II responses informed our understanding of the possibilities (or improbabilities) of connecting knowledge acquired at the conference to school and classroom practice.

Description of Setting and Participants

During the summer of 1996 WEAC, the Wisconsin Department of Public Instruction (DPI), and the University of Wisconsin-Oshkosh sponsored the Wisconsin State Assessment Institute. These organizations viewed the conference as an opportunity to support professional development in what one WEAC official described as a "hot topic area." WEAC earmarked money to support participation in this conference by covering the conference registration fee ($150) and per diem costs. Participants had to cover the cost of transportation.

Of the total possible number of respondents (N=301), we analyzed 299 usable surveys. Two hundred and thirty-seven females (76%) and 62 males (21%) completed the pre-conference survey (about 3% of the respondents did not indicate gender). Slightly over 69% of the respondents were classroom teachers, 9% principals, 4% Directors of Instruction, and 14% other (e.g. school and district administrators and specialists). Elementary teachers represented the largest category of participants (52%). Overall, the sample represented a very experienced group of educators with an average of 17 years in education, nine of those years in their current positions. Seventy-eight percent were attending this conference as a member of a school team.

Reasons for Attending a Large-scale Conference

Financial support was a major inducement to attend the institute, especially for teachers. Most participants received financial support from their school, district, and/or professional association to attend the three-day conference. Only about 3% of respondents reported spending their own resources to attend the institute. WEAC, for example, covered over 90 percent of conference costs for approximately one-third of the participants (n=101). Among all respondents, 62% reported paying less than 25% of the cost for the 3-day institute. In the post-conference survey we asked cohort II respondents (N=101) whether or not they would have participated in the assessment institute if WEAC had not covered the majority of the cost. Twenty-four percent indicated they would not have attended and four percent of respondents were unsure. Only five percent said they would have attended even if their costs had not been covered. Educators were willing to give their time, but clearly they needed financial support. As an additional incentive, the Institute also offered all participants continuing education or graduate credit. Since all K-12 educational professionals in Wisconsin must earn an additional 6 graduate credits, or its equivalent, every five years to retain their license, the financial support provided by school districts and by WEAC to its members coupled with credits toward license renewal were particularly attractive incentives.

While financial incentives and credits for license renewal were important factors influencing participants' decisions to attend the three-day institute, the issues and topics addressed at the institute on student performance assessment also provided a strong incentive to attend. According to cohort I respondents, recent adoption of performance based standards by DPI, mandated state-wide competency tests for students at grades 4, 8, and 10, and increased public scrutiny of student learning outcomes, especially those that demonstrate proficiencies through performance, made the content and activities in this institute especially attractive and timely. Specifically, our analysis of cohort I open-ended responses indicated that there were four primary reasons respondents attended the institute. First, the topics addressed and the varied professional development opportunities were relevant to current state-mandated performance assessment activities in their schools and/or districts. Second, participants wanted to know more about student performance assessments. Third, they believed that learning more about performance assessment would enhance their classroom teaching and student assessment skills. Finally, the three-day institute provided participants, especially those who came as members of school teams (78%), a chance to work with colleagues for an extended, uninterrupted period of time. For example, when asked to explain why the conference on performance assessment interested them, they provided responses such as "It provides time to work as a team;" or "Time to work with my colleagues from my school;" "A chance to work with the school team to learn and grow;" and "Time to work with colleagues on our projects."

Expectations of Conference Utility

We found that respondents primarily hoped to gain an awareness of the concepts and theories, how-to knowledge, and political knowledge from the conference activities In most cases, respondents reported that they were in the early stages of performance assessment implementation in their schools. As a result, most respondents simply wanted to know more about performance assessment (i.e., Schlechty & Whitford's establishment function). What was it exactly? What are the key ideas, theories, concepts, and language they needed in order to be able to consider its application to their current work? For many respondents, such terms as "rubrics", "portfolios", and "performance-based assessment" remained fuzzy abstractions, not part of their current thinking, language, or practice.

Respondents also hoped to gain other forms of useful knowledge. For instance, respondents wanted to know (1) how new forms of student performance assessment work; (2) how to put together portfolios; (3) how to communicate to parents clearly and confidently information about performance indicators during conferences; and (4) how to integrate performance assessment into their current teaching practices. For example one respondent commented on her interest in acquiring both procedural and propositional knowledge:

> I hope to walk out of here with a clearer understanding of what performance assessment is and [what are] its components. I hope to have some concrete ideas, which can be, employed "day one" of this school year. I hope to bring back some recommended strategies in which to diversify our testing methods currently being used.

These same respondents also hoped to gain insight into the dynamics and politics associated with changes in student performance assessment. That is, they wanted to learn more about how to disseminate information to their colleagues and communicate clearly the purposes and importance of new forms of performance assessment to parents, school board members, and others in the community. For instance, one person's comments reflected a common hope among respondents that the conference would provide her school's performance assessment team with the know-how to influence others about the potential for new ways of assessing student work: "The development of camaraderie among our team to work together to carry the message of performance assessment back to our district."

Connecting Professional Learning to Work in Schools

At the conclusion of the three-day institute, we asked cohort II respondents whose participation had been supported by WEAC to complete a second written survey. Of particular importance to us were these respondents' views on the enhancement and maintenance functions of conferences as a professional development activity (i.e., the connection between what they had learned and how this might influence their professional practice). When respondents were asked, "Do you plan to implement changes in the way you assess performance as a result of information obtained at this conference?," 30% of the respondents indicated they were planning to make such changes. Only three percent of respondents said they would not be making any changes. However, when we asked if their team or school was planning to implement changes in performance assessment, the number of affirmative responses dropped. Only 23% of respondents believed their school would be implementing changes in performance assessment, while 11% did not believe any changes would be made in assessment practices in their schools. These findings may reflect the predisposition of attendees to reconsider their assessment practices while their colleagues, not in attendance, were less likely to be making significant changes in their practices in the near future. Whether or not significant changes in teachers' performance assessment practices will be successfully implemented in schools remains an empirical question Regardless of the outcome, like any innovation, successful implementation of performance assessment requires careful planning, adequate resources, and purposeful strategies for the dissemination and diffusion of the innovation.

We were also interested in knowing what these respondents would do with the

knowledge and skills they had acquired during this three-day institute. When asked how they intended to share what they had learned when they returned to their schools, most participants (86%) indicated that they would share what they had learned with their colleagues. As encouraging as this appears, most strategies mentioned for sharing information with colleagues were informal. In other words, few respondents described systematic ways in which newly acquired information on performance assessment and knowledge about assessment practices would be disseminated in their schools. The most frequently cited format was in meetings—faculty, team, curriculum, and departmental (32%). Other strategies for sharing information included working with colleagues and modeling particular uses of assessment practices in their schools (24%), staff in-services and workshops (18%), distributing printed materials (11%), and working in teacher study groups (7%). Sixteen percent of respondents indicated they "did not know" how information would be disseminated in their schools.

The professional development of teachers and change processes in schools require sufficient resources for optimal impact on the lives of teachers and students. We asked respondents to describe the types of resources, if any, that were available to support the implementation of new forms of student performance assessment in their schools. The most frequently listed resources revealed that these educators primarily looked inward. For example, according to 42 percent of respondents the quality and professional expertise of their own staff, teachers, and administrators were the most important resources available to support teachers' continued learning and successful implementation of performance assessment changes in their schools. This suggests that the respondents believed the richest (or perhaps the only) possibilities to support teacher learning and substantive change, in this case performance assessment practices, were already in place in schools, not externally in some remote bureaucracy, corporation, or private benefactor. School level capacity was important, but individual will and commitment were essential to successful change. In addition on- site professional expertise, existing staff development funds, and school/district sponsored professional development activities were cited as key sources of support (24%). Another important resource was printed materials/literature (17%). External funds (10%) and outside experts (7%) were also listed as resources available. Twenty-seven percent of respondents indicated that there were no resources available, or if there were, they did not know how to access them.

Discussion

As noted earlier, we remain cautious about the generalizability of our findings because the survey respondents represented a convenience sample of educators at one professional development institute. Despite this limitation to external validity, we believe our findings highlight several important issues related to the role of large scale conferences and workshops in the larger context of professional development. We organize our discussion according to the following topics: (1) this conference's place in Schlechty and Whitford's typology; (2) the types of knowledge educators sought (and perhaps acquired); and (3) factors that facilitated or impeded the usefulness of this often maligned professional development activity.

Functions of a Large Scale Conference

Clearly, this large scale conference served an *establishment* function. That is, the purpose of the conference (according to its organizers) was to introduce the latest concepts of and approaches to student performance assessment. Our data support that most educators expected to have basic questions about student performance assessment answered at the conference. Most respondents indicated they simply needed to know more about performance assessment. These findings are consistent with others' perspectives on adult and professional learning. For example, these findings parallel Hall and Hord's (1987) stages of concern model for educational innovations. According to Hall and Hord, at the early stages of any innovation, teacher interests center on awareness and informational concerns. Once dealt with adequately, then teachers' concerns shift to task and impact concerns. Our data from open-ended responses also indicate that teachers' stages of learning and levels of concern are similar to the sequence of stages in teacher career development: survival, exploration and bridging.

adaptation, conceptual change, and invention (Huberman, 1989). During the initial stages of this innovation, i.e., changes in teachers' assessment practices, the survival stage is intertwined with what Huberman (1989) calls "discovery." "Empirical studies show that these two aspects occur in parallel, and that the excitement and challenge of 'discovery' is what brings many teachers through the attrition of day-to-day 'survival'" (Huberman, 1989, p. 349).

To a lesser degree, but worth mentioning, is the *maintenance* function the conference served. Participants attended the conference during a time of great debate and legislative change in Wisconsin's education landscape. New state requirements were beginning to emerge and these teachers and administrators wanted to not only increase their awareness of the initiative, but also, ensure that they were in compliance with the new legal requirements. However, this conference did not show promise according to Schlechty and Whitford's *enhancement* function. Our findings suggested that while respondents (i.e., cohort II) did learn valuable information at the conference, it was not clear how that knowledge would be transferred to their own classroom practice or to their colleagues.

Finally, we would be remiss if we did not mention the opportunity to gain continuing education credits required by the Wisconsin DPI as an important purpose of the three-day conference. However, as we describe in the preceding two paragraphs, our skepticism (or perhaps cynicism) that participants were probably motivated more by the chance to "knock out" continuing education hours than by intrinsic interest in learning an important topic on the state's education landscape was tempered by our findings.

Knowledge Acquired

Giving up three days of their summer break was strong evidence that these participants were interested in knowing more about assessment. The types of knowledge discussed by Eraut and others that we outlined above provide insights into the kinds of learning respondents claimed to have experienced. In particular they wanted three types of professional knowledge: *propositional, procedural*, and, what we call *political knowledge*. That is, participants expected to learn the concepts, theories, and language—or propositional knowledge (i.e., how to talk about performance assessment), how to actually implement new performance assessment models such as portfolios in practice (procedural knowledge), and how to learn how others had successfully implement these new performance assessment models in the face of potentially skeptical parents, the business community, and their own colleagues (political knowledge).

Factors that Facilitate or Impede Usefulness

Inherent in this conference were several factors that respondents believed facilitated its usefulness in spite of popular criticisms of this professional development vehicle. First, the large scale nature of the conference provided the almost 300 respondents with numerous learning activities from which to choose. The availability of choices was important to participants given their varying degrees of awareness of student performance assessment. For instance, we found that elementary teachers (accounting for slightly over two-thirds of the respondents) demonstrated a better understanding of issues around student performance assessment, its link to teaching and the curriculum, and how various types of student performance measures (e.g., portfolios, demonstrations, and projects) would be implemented in their classrooms than did their secondary school counterparts.

Incentives and resources to support professional development are important. Time, money, and graduate credits for license renewal influenced respondents' motivation to participate in this institute. Without financial support from WEAC, from local school districts, or other agencies, respondents stated overwhelmingly that they would not have attended this conference. The financial support from WEAC, the largest teacher union in the state, also suggests that unions are beginning to reexamine ways in which they can support their members beyond contract bargaining and the protection of members' rights and due process. This resonates with a recent statement on the role of teachers'

unions by Bob Chase (1997), President of the National Education Association:

> Membership polls tell us that most teachers want their union to match its traditional emphasis on decent salaries, benefits, and working conditions with a more aggressive commitment to professionalism and quality.
>
> And I agree. Our sights are set on tougher academic standards, stricter discipline, less bureaucracy, higher quality schools. These goals, shared by teachers and school boards alike, compel us to transform collective bargaining into a collaborative process - negotiations focusing not only on traditional bread-and-butter issues, and also on issues of employee involvement and school quality (Chase, 1997, not paginated).

Furthermore, two other important themes emerge from respondents' preferences and descriptions of what they hoped to gain in this three-day institute. The first is finding time to work with school colleagues. Given the reputation of large scale professional development conferences in recent literature, we found it somewhat ironic that respondents viewed this conference as a place that provided the time and place for colleagues to collaborate. Although this finding was somewhat surprising, it makes sense given traditional school structures that often result in teachers' career-long isolation from their professional colleagues. Teachers' self-reliance as practitioners and as l arners is evident in these survey data. In part, this is a legacy of the one-room school where teachers were isolated from their professional colleagues and thus developed a powerful sense of individualism. Ironically, sometimes the only way teachers and principals can find the time to work together is to leave their schools. A second theme was the importance of social interaction in professional learning that cut across structured sessions and informal exchanges among these educators.

Conclusion

As we stated earlier, professional development has risen in status to become one of the principal mechanisms to achieve the 1990's reform agenda. As professional development has become a primary strategy for reform implementation, so has it gained the attention of not only school and district educational practitioners and policy makers, but state level policy makers as well. The results of this study of a state wide conference to educate practitioners about performance assessment underscore at least a few important points. For instance, from these educators' perspectives workshops and professional conferences serve an important purpose by (1) introducing and demystifying often abstract reform concepts; (2) deprivatizing teacher practice in ways that foster the "cross-pollination" of practical ideas; and (3) providing a venue for teachers and other educators—committed to addressing daily moral imperatives of their work—to explore pressing issues that can broaden the professional frames through which they approach their profession.

However, as professional development takes on increased significance at the state and even federal levels, this study also highlights the need to strengthen linkages between schools, school districts, and state level education agencies (e.g., state departments of education and state teachers unions). While a majority of participants in this study attended the conference as part of a school team, and many were supported by WEAC and/or their school districts, alarmingly few participants were confident that they could disseminate their newly acquired knowledge to colleagues in their schools So, while large scale professional development conferences may have their place in overall professional development programs, coordination between the various levels of our educational system must occur to ensure that the professional knowledge gained is internalized by teachers, principals, and others into their respective practices.

References

Bogdan, R. C . & Biklen, S. K. (1998). *Qualitative research in education: An introduction to theory and methods*. Boston: Allyn and Bacon

Bredeson, P. V. (1998) Paradox and possibilty: Professional development and

organizational learning in education. A paper presented at Exploring New Horizons in School Leadership Conference at Umea University, Umea, Sweden.

Chase, R. (1997). The new NEA: Reinventing teacher unions for a new era. Presentation by Bob Chase to the National Press Club.

Corcoran, T. C. (1995). *Transforming professional development for teachers: A guide of state policymakers*. Washington, DC: National Governors' Association.

Darling-Hammond, L. & McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan*, April 597-604.

Eraut, M. (1994). *Developing professional knowledge and competence*. London: Falmer Press.

Fullan, M. (1995). The limits and the potential of professional development. In T. R. Guskey & M. Huberman (Eds.), *Professional development in education: New paradigms and practices*. New York: Teachers College Press.

Fullan, M. & Hargreaves, A. (1996). *What's worth fighting for in your school?* New York: Teachers College Press.

Hall, G. & Hord, S. (1987). *Change in schools: Facilitating the process*. New York: SUNY Press.

Houghton, M & Goren, P. (1995). *Professional development for educators: New state priorities and models*. Washington, DC: National Governors' Association.

Huberman, M. (1989). On teachers' careers: Once over lightly, with a broad brush. *International Journal of Educational Research, 13*, 347-362.

Joyce, B. (Ed.). (1990). *Changing school culture through staff development*. Alexandria, VA: ASCD.

Lieberman, A. (1995). Practices that support teacher development. *Phi Delta Kappan*, April, 591-596.

Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis, 15* (2), 129-151.

Marsick, V. J., & Watkins, K. E. (1990). *Informal and incidental learning in the workplace*. New York: Routledge.

McLaughlin, M. W. & Oberman, I. (1996). *Teacher learning: New policies and practices*. New York: Teachers College Press.

National Commission on Teaching and America's Future (1996). *What matters most: Teaching for America's future*. New York. The National Commission on Teaching and America's Future.

National Commission on Excellence in Education (1983). *A nation at risk*. Washington, DC. Government Printing Office.

National Foundation for the Improvement of Education (1996). *Teachers take charge of their learning*. Washington, DC: The National Foundation for the Improvement of Education.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text*. Thousand Oaks, CA: SAGE.

Schlechty, P. C., & Whitford, B. L. (1983). The organizational context of school systems and the functions of staff development. In G. A. Griffin (Ed.), Staff development: *82nd yearbook of the NSSE* (pp. 62-91). Chicago: University of Chicago Press.

Scribner, J. P. (1999). Professional development: Untangling the influence of work context on teacher learning. *Educational Administration Quarterly, 35* (2), 238-266

Smylie, M. A. (1996). From bureaucratic control to building human capital: The importance of teacher learning in education reform. *Educational Researcher, 25* (9), 9-11.

Strauss, A. L. & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: SAGE Publications.

Weissglass, J. (1997). Deepening our dialogue about equity. *Educational Leadership, 54* (7), 78-81.

About the Authors

Paul V. Bredeson

University of Wisconsin-Madison
Department of Educational Administration
1282J Educational Sciences Building
1025 West Johnson Street
Madison, Wisconsin 53706

Email: bredeson *a* education.wisc.edu
Tel: (608) 262-3886
Fax: (608) 265-3135

Over the past 19 years, Paul Bredeson's research has centered on alternative conceptions of leadership, especially in regard to school principals. This interest is grounded in his professional work experiences as a high school Spanish teacher, high school principal, director of a bilingual administrator training program, and as Executive Director of the Pennsylvania School Study Council located at Penn State University. More recently, his teaching and scholarly interests have focused on the professional development and learning in educational organizations. Publications include Hart, A. W. and Bredeson, P.V. (1996). *The principalship: A theory of professional learning and practice*. New York: McGraw-Hill, Inc. and Bredeson, P. V. (1996). New directions in the preparation of educational leaders. Chapter in *The International Handbook for Educational Leadership and Administration*. Eds. K. Leithwood and A.W. Hart. Amsterdam: Kluwer.

Jay Paredes Scribner

University of Missouri-Columbia
211 Hill Hall
Columbia, MO 65211

Email: scribnerj *a* missouri.edu

Phone: 573-884-1708
Fax: 573-884-5714

Jay Paredes Scribner is Assistant Professor in the Department of Educational Leadership and Policy Analysis at the University of Missouri-Columbia. His most recent research has focused on teacher professional learning, organizational learning, and professional community.

Appendix
Survey Items

1. Cohort I Survey
   - Gender:
   - Indicate current position (official title):
   - Level (e.g., elementary, middle, or high school):
   - Number of years in current position:
   - Total number of years employed in education:
   - School size (approximate number of students):
   - Are you attending this workshop as part of a school team? (If yes, how many people are in your team?)
   - What percentage of total costs (including registration, lodging, per diem, travel) of attending this three day conference is being paid by the following? (Indicate percentages)
     - Personal funds
     - School and/or district support
     - Professional association (e.g., WEAC) support
     - Other (If "other" please specify)
   - How did you find out about this workshop?
   - In the area of performance assessment, what activities are currently going on in your school?
   - Why did this conference on performance assessment interest you?
   - What are the three most important things you hope to gain from this conference?
2. Cohort II Survey
   - Name:
   - District:
   - School:
   - School Address:
   - Did you attend this conference individually or as part of a school (or district) team?
   - What did you learn about performance assessment that you believe would benefit you and your school? Please list up to 3 examples.
   - How did you find out about this conference on performance assessment? Please specify.
   - Would you have attended this conference if WEAC had not covered the cost of attending? Why or why not?
   - What is your school currently doing in the area of performance assessment?
   - Do you plan to implement changes in the way you assess performance as a result of information obtained at this conference? (Check one).
   - Do you (or does your team) plan to implement changes in the way teachers in your school assess student performance? (Check one).

     If so, how do you plan to share what you have learned in the three-day conference in your school? Please explain.

   - What resources, if any, are available to support the implementation of performance assessment in your school? Please explain

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (480-965-9644) The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu The Commentary Editor is Casey D. Cobb

casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeá University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
aacosta@cucea.udg.mx

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Hugo Casanova Cardiel
Universidad Nacional Autónoma de México

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE CINVESTAV
rkent@gemtel.com.mx kentr@data.net.mx

Maria Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gomez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon @ openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 2396 times since February 24, 2000

archives | abstracts | editors | board | submit | comment | subscribe | search | epaa.asu.edu

Education Policy Analysis Archives

Volume 8 Number 14 February 24, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Teachers and Tests: Exploring Teachers' Perceptions of Changes in the New York State Testing Program

S. G. Grant
State University of New York at Buffalo

Abstract

How do teachers change their pedagogical practices? While many current initiatives seek to raise educational standards and improve student academic performance, there is a curious gap in national and state reforms. Considerable attention is given to defining higher expectations for what students will know and be able to do, yet little attention is given to how teachers should learn new pedagogical ideas and practices. This exploratory study uses focus group interview data collected over two years to examine how cross-subject matter groups of elementary and secondary New York State teachers respond to one way of learning to change their classroom practices: state-level testing. Analysis of the data highlights three issues: the nature and substance of the tests, the professional development opportunities available to teachers, and the rationales for and consequences of the state exams

Many current initiatives seek to raise educational standards and improve student academic performance. Yet, there is a curious gap in the recent talk about national and state reforms. While much attention focuses on defining higher expectations for what students will know and be able to do, little attention is given to how teachers should learn new pedagogical ideas and practices. Such policies as the federal *Goals 2000: Educate America Act* and the New York *New Compact for Learning* focus on the resources, conditions, and practices necessary for all students to learn. None of these efforts, however, seriously addresses how experienced teachers will learn the intended innovations.

How do teachers change their pedagogical practices? Some suggest change comes through new subject matter standards proposed by professional organizations (National Council for Social Studies, 1994), by national groups (National Center for History in the Schools, 1994), or by state education departments (New York State Education Department, 1996). Others believe teachers change their practices in response to organizational restructuring (e.g., smaller classes, block scheduling). Still others assert that real change in the classroom lives of teachers and students depends on changes in state-level assessments (Comfort, 1991; Smith & O'Day, 1991). The assumption in this last case is that testing drives much of what teachers do, and so curricular and instructional change will occur if and when state tests change.

This last idea is intriguing for, if true, it suggests the potential for big pedagogical changes with a modicum of policy effort: Change the test and one changes teachers' practices. New York state policymakers seem taken with this approach, for although they have developed new curriculum standards, it is revision of the state testing program which gets most of the attention (Grant, 1997a). The scope of that revision is wide. One piece is the change from program evaluation tests at the elementary level to high-stakes individual student testing. A second piece is the phase-out of the less demanding high school Regents Competency Tests and the requirement that all students pass the more demanding Regents tests. A third piece is a change in the content and format of all state tests presumably to reflect the higher expectations expressed in the state's new standards documents.

What sense do teachers make of these new state tests and how, if at all, do the tests influence their classroom practices? Strange as it seems, there is little empirical evidence to suggest how teachers, especially teachers at different grade levels, respond to changes in state tests. Assessment is a particularly hot topic in educational circles today, yet there is surprisingly little research which digs deeply into teachers' understandings of the import of standardized tests (Cohen & Barnes, 1993; Grant, in press). Corbett and Wilson's (1991) study of teachers' reactions to a new Maryland testing program is well-known as is the on-going work of Mary Lee Smith and her colleagues in Arizona (Noble & Smith, 1994; Smith, 1991; Smith, Heinecke, & Noble, 1999), but these are few studies in a field that is more prone to study students' responses than teachers'

In this article, I use the data collected through focus group interviews over two years to explore the relationships between teachers and tests. My findings suggest that teachers need to be much more involved in the process of changing state assessments, and that professional development needs to be more attuned to the different needs teachers have

The Study

The Teacher Learning and Assessment (TLA) research project (Note 1) is designed to look generally at the intersection of teachers and assessments. The research team is a cross-subject matter group of faculty and students (English, mathematics, science, and social studies) who are interested in exploring the relationship between teacher learning and state-level testing. Our study questions include: a) In what ways are tests and test results used in classrooms, schools, and the districts? b) What do the proposed changes in state-level tests mean for teachers and learners? c) How are teachers being prepared to respond to the new state assessments? and d) What challenges do teachers and administrators anticipate in moving toward new state assessments? In each case, we are interested in the extent to which these issues differ across school subject matters and grade levels.

Data Collection

In the first year of data collection, we organized two focus groups, one composed of 7 elementary school teachers and counselors and one composed of 12 high school teachers. The participants represented a cross-section of urban, suburban, and rural school districts in western New York state, a breadth of teaching experience (2-25 years), and a range of school subjects (language arts, mathematics, science, and social studies). Each of the two-hour focus group interviews was tape-recorded and transcribed.

During the second year of data collection, we again organized separate elementary and secondary focus groups. We debated whether to: a) reconstitute the original groups only; b) develop new groups of teachers separate from those involved in the first year's interviews; or c) call together groups that mixed teachers new to the project with those who had participated during the previous year. We rejected the first option, fearing that attrition might leave us with groups that were too small. We also rejected the second option, though largely because of timing· We did not think we could hold four focus groups near the end of the school year. In the end, we decided to constitute mixed groups for two reasons. One reason was that we wanted to expand the number of teachers we were talking with; the second reason is that we were interested in how the two groups might interact. The secondary focus group consisted of 8 teachers representing mathematics, science, English, and social studies; 5 of the 8 were in the original sample. The elementary focus group consisted of 5 teachers, 3 of whom were in the original sample. (Note 2)

The data consist of interview transcripts of the focus group sessions and post-interview evaluations completed by the participants. The focus group interviews followed a semi- structured interview protocol (see Appendix). Questions used during the first year asked participants to construct a metaphor to represent their sense of the changes in state-level testing, what the new tests mean for teaching and learning across school subjects, how teachers are being prepared for new standards and new assessments, and what challenges teachers believe they face. The post-interview questions asked the participants to reflect on the issues raised around the relationship between state-level assessment and classroom practice. The interview protocol was largely the same during year two. Changes consisted of replacing the metaphor task with a fill-in-the-blank exercise ("I used to think of the state assessment as _____. now I [still] think of it as _____.") and the addition of probes that asked participants if they sensed a change from last year to the present. There were no changes to the post- interview evaluation

Data Analysis

All data were analyzed inductively from an interpretivist stance (Bogdan & Biklen, 1982; LeCompte, Preissle, & Tesch, 1993). That stance emphasizes the importance of context, and the multiple ways individuals construct meaning. All data were also analyzed using a constant comparative method (Bogdan & Biklen, 1982; Glaser, 1978). That method assumes that data collection and analysis are recursive, one informing the other throughout the course of the study. After coding the data both within and across grade levels and subject matters, I began seeking patterns in the informants' responses. The themes which emerged reflect the full data set, but in each case I highlight the implications for social studies

Although this data can be considered largely exploratory, patterns and themes surfaced as the interview and evaluation data were analyzed related to the research questions. In the analysis of the focus group interviews, I focused on: how teachers make sense of, and make different sense of, the state curriculum and assessment documents they encounter; the kinds of learning opportunities they attend, and how, if at all, these reforms and opportunities influence what teachers think about and do in their classrooms. Looking across the interviews, I saw patterns which help explain the teachers' responses in a social context and the nature of their learning in an array of social settings The three preliminary patterns I synthesized from the data and report on in this paper relate to the nature and substance of the tests, the professional

development opportunities available to teachers, and the rationales for and the consequences of the state exams.

On Tests and Teaching

Standardized tests matter. The professional literature is replete with debates about tests as a means of accountability, as measures of performance, and as levers of change (Corbett & Wilson, 1991; Editors, 1994; Feltovich, Spiro, & Coulson, 1993; Finn, 1995; Fuhrman, Clune, & Elmore, 1988; Koretz, 1988; Ravitch, 1995; Resnick & Resnick, 1985). These concerns become elevated when situations like CTB/McGraw-Hill's mis-scoring of almost 9000 New York City students' tests occur. In all of the talk about tests, however, one area gets scant regard: What teachers learn from tests, and if and how that knowledge affects their instructional practice. Common sense holds that tests drive classroom instruction. Evidence for that opinion is thin, however. Much research focuses on the relationship between students and tests (see, for example, Natriello & Pallas, 1998; Stiggins & Conklin, 1992; Wolf, 1998), but relatively few empirical studies explore the relationship between teachers and the tests they administer (Corbett & Wilson, 1991; Firestone, Mayrowetz, & Fairman, 1998; Grant, in press; Noble & Smith, 1994; Smith, 1991). The research that is available presents a mixed picture at best.

Those advocates of tests as a vehicle for driving educational change tend to cite general positive effects rather than specifics. Some (Feltovich et al., 1993; Popham, 1998; Shanker, 1995) simply argue that good tests will inevitably drive good instruction. Lacking any more specificity, Popham, Cruse, Rankin, Sandifer, and Williams (1985) claim that tests measure important learning, and that good tests results equal good education. Systemic reformers (Fuhrman, 1993; Smith & O'Day, 1991) advocate for testing as part of an overall strategy aimed at fundamental school change. Others (English, 1980; Glatthorn, 1987; Heubert & Hauser, 1999) argue that because standardized tests are a reality in most school districts, they should be used as a fundamental part of curriculum planning.

Critics of standardized testing are more direct in their assessment of the impact of testing on teaching. Madaus (1988) claims, among other things, that teachers will teach to the test, that they will adjust their instruction to follow the form of the questions asked (e.g., multiple-choice, essay), and that tests transfer control over the curriculum to whomever controls the test (Note 3). Claims by LeMahieu (1984) and Koretz (1995) are more tentative, but they too conclude that teachers may tailor their curricula to the content covered on the test. Recent empirical work supports some of these claims. Smith (1991) argues that many teachers respond overtly to test pressures and she offers a typology of eight orientations toward test preparation: ordinary curriculum with no special preparation, teaching test-taking skills, exhortation, teaching content known to be covered by the test, teaching to the test in format and content, stress inoculation, practicing test or parallel test items, and cheating. Firestone, Mayrowetz, and Fairman (1998) assert that testing programs in Maine and Maryland seem to influence teachers' content decisions, although they conclude that such influences are weaker than expected. Corbett and Wilson (1991) argue that testing, especially minimum-competency testing, has a pernicious effect on teachers in that it causes them to narrow their sense of educational purposes and to focus on activities designed to raise test scores whether or not they think those activities are good for students. They conclude that squeezing teachers in this fashion encourages them to rebel against reform measures good and bad. "Statewide testing programs do control activity at the local level, but the subsequent activity is not reform" (p. 1)

Other researchers are less sure that a direct relationship exists between standardized testing and teachers' classroom practices. Freeman, Kuhs, Porter, Knappen, Floden, Schmidt, & Schwille (1980), Kellaghan, Madaus, and Airasian (1982), and Salmon-Cox (1981) found little direct impact of standardized testing on teachers' daily instruction. Firestone, Mayrowetz, and Fairman (1998) claim that, while tests may have influenced teachers' decisions about what to teach, there was virtually no influence on their decisions about how to teach. In a cross-case comparison of two high school teachers' civil rights units (Grant, in press), I found little direct influence of testing on either teacher's content or pedagogical decision-making.

This brief review suggests two points. First, we need to know more about the

relationship between teachers and tests. While the impact of tests on students has been much explored, research that inquires into if and how teachers are influenced by standardized tests is lacking. Second, that research around teachers and tests fails to show a clear or consistent pattern of influence. Tests matter, but how and to what extent is unclear.

State-Level Curriculum and Assessment in New York State

State-level influence over curriculum and assessment is a well-established tradition in New York State. The Regents test has been administered continually for over 100 years. These tests are administered in all academic subjects and are tied to school courses. For example, in social studies, students take the Global Studies test at the end of a two-year Global Studies course sequence in ninth and tenth grades; eleventh graders take the U. S. History and Government test after completing a course of the same name. Elementary and middle school teachers also follow a state curriculum in all school subjects and students take state-developed tests.

Recent State-Level Curriculum Changes

As is the case in most states, educational reform has been steady work since the 1980s. Begun during the tenure of former Commissioner of Education, Thomas Sobol, state-level focus on and activity around school curriculum hit full stride in the mid-1990s under current Commissioner Richard Mills.

Since 1994, working groups of state policymakers, teachers, and administrators have produced new curriculum and learning standards and scope and sequences for all school subjects. Social studies teachers, for example, may now consult the *Learning Standards for Social Studies* (New York State Education Department, 1996) and the *Resource Guide for Social Studies* (New York State Education Department, 1998). Compared with the previous round of curricular revisions in the mid-to-late 1980s, the changes represented in these documents vary from virtually no changes in the K-5 grades curricula, which follow an expanding horizons model, in the seventh and eighth grade U.S. and New York State history, or in the twelfth grade Participation in Government and Economics courses. Modest changes are evident in other curricula, such as the emphasis on geography in the eleventh grade U.S. history and government course. Major changes seem localized at sixth grade, where the course of study expanded from Western and Eastern Europe and the Middle East to the entire Eastern hemisphere, and at ninth and tenth grades, where the emphasis has changed from a cultural approach as represented in Global Studies to a chronological study as expressed as Global History and Geography.

Recent State-Level Assessment Changes

The state-level testing program is also changing. Although the scope of the changes varies (Note 4), the net effect appears to be a general ratcheting up of the stakes for both teacher and students.

State tests of language arts, mathematics, and science have undergone radical transformations which include reducing the number of multiple-choice items and increasing the number and range of performance tasks. For example, new science tests call for students to actually perform experiments. By contrast, the social studies assessments will apparently change little: Multiple-choice questions will still dominate the tests, accounting for 55% of a student's score (Note 5). The major change seems to be in the writing portion of the exam. Unlike many minimum competency tests, New York students have always had to answer essay questions on state exams. The new tests are different primarily in the fact that a) students will no longer have a range of essay prompts to choose from, and b) a new kind of essay question, a document-based question (DBQ), is being introduced on each of the fifth, eighth, tenth, and eleventh grade tests. A DBQ asks students to write an essay synthesizing a number of primary source documents (e.g., short quotes from government documents and famous individuals, political cartoons, poems, charts and graphs) (Note 6). Plans call for students to answer a main idea-type question about each of the documents before writing their essay. High school students will also write a second, "thematic" essay

BEST COPY AVAILABLE

based on a single prompt (Note 7) The inclusion of the DBQ is the primary change in the structure of the social studies exams. One might argue that such a question represents a major shift away from traditional testing, but given the scope of the test (and the fact that students can easily pass the test without a single DBQ point), adding a DBQ could be read as a minor revision, or an instance of what Tyack and Cuban (1995) call "tinkering toward utopia."

Three other changes seem more dramatic. One is that the new fifth and eighth grade tests will produce individual student scores. Tests at those levels, termed "Program Evaluation Tests," have aimed at helping teachers understand the effectiveness of their content and pedagogical decisions (Note 8). The shift of emphasis to individual students is apparently intended to raise the stakes of these tests and tie them more directly to the high school Regents exams. The function of the Regents test is also being fundamentally changed. In the past, passing Regents tests in all academic subjects meant that a student earned a Regents diploma. Students could opt to take the less rigorous Regents Competency Exam (RCT) and earn a local diploma. Ninth graders beginning in 2001 will no longer have these options. The RCT will no longer be administered, and all students will have to pass five Regents examinations (English, mathematics, global history, U.S. history, and science) in order to graduate.

Given these changes, state-level tests are no less high-stakes for teachers than they are for students. Since the mid-1990s, state policymakers have introduced a number of curriculum reforms, such as new state standards for social studies, yet it is a concern about the state tests which surfaces most regularly in teachers' talk (Grant, 1997a). This makes sense for two reasons. First, the curriculum documents produced thus far offer teachers little assistance in making concrete instructional decisions (Grant, 1997b). Second, the messages teachers receive often promote the view that tests are intended to drive change (Grant, 1996). For example, during sessions devoted to new state social studies standards, one representative from the New York State Education Department (NYSED) said that new tests will "help grow change in the system." During another session, a different SED representative said, "New assessments will represent a change in instruction....Kids won't perform well until (teachers') instruction reflects this." And at yet a third meeting, NYSED Commissioner Richard Mills added, "Instruction won't change until the tests change." The message that tests matter was echoed during local school and district meetings. A suburban district social studies supervisor, for example, told teachers that "change in content will come if we change the tests." An urban district supervisor observed, "If we change the assessments, we'll change instruction" (p. 271). One might question the focus of test influence--instruction, curriculum, or the "system" in general--but it is hard to miss the larger point: tests matter.

The Prospects and Problems of State-Level Testing In New York State

The tendency of advocates and critics to cast standardized testing in black and white images is not supported here. My analysis suggests that teachers see the new NYS tests as a mixed bag. The prospects of tests which more closely mirror and support thoughtful instruction and closer collaboration with colleagues are mitigated by the problems of, among other things, uncertainty about the rationale for and consequences of the new tests and the unevenness of the opportunities to learn about and respond to changes in the tests. In short, teachers across grade levels and subject matters express an uneasy combination of hope and fear, anticipation and dread. I explore those poles by looking at teachers' perceptions of the new tests in terms of their nature and substance, the professional development opportunities available, and the rationales and consequences.

The Nature and Substance of the New NYS Tests

The NYSED is phasing in the new state tests over a period of four years, beginning with the English language arts tests at grade 4 in January, 1999. Consequently most of the teachers interviewed have not seen final versions of the tests they will administer. All have, however, received preliminary materials from state, district, and professional organization sources and so most assume that they have a fair

sense of what the new exams will be like. Most believe the tests will be an improvement over past assessments, but questions about the nature and substance arise.

Both elementary and secondary teachers expressed at least modest support for the general direction taken in the new tests. A middle school science teacher suggested simply that the NYSED was "changing what assessment means." An elementary school teacher was more specific. "I think there was a lot of change going on and then they changed the assessment," she said, "I remember giving that CTBS (a basic skills test) and teaching a literature-based program, and we were all complaining that it wasn't reflective [of our teaching]." Another elementary school teacher was more specific: "The new assessments test the same way we teach reading, and where we want kids to be in math."

Social studies teachers approved of the move to include primary sources within the DBQ. A high school teacher cited the real world relevance of questions which employ political cartoons. "You give them a cartoon and you say, 'Interpret this cartoon,'" she said, "That's interpretation, you know? If you open a paper and you look at a picture in the newspaper and you go, 'What's that mean?' That's something you would do in real life." A middle school teacher noted she now uses DBQ kinds of questions as a regular part of her instruction:

> I was working on a social studies test today for grade seven where they have to look at a document and think about some stuff like, what was the theme about the Revolutionary war, and they've got to write notes based on the picture. And it looks-the test is a lesson. It's a lesson in analyzing documents and taking notes from the document so you're not looking to see if they're right or wrong. You're looking to see can they look and think about what's on there.

This teacher and most others praised state efforts to bring standardized assessments into closer alignment with the kind of ambitious instruction they believe is important, such as analyzing primary sources and understanding that such texts can be interpreted in multiple ways. Social studies teachers worry about the continued strong emphasis on multiple-choice questions, but in questions like the DBQ, they see potential for pushing their students toward richer understandings.

But not all teachers held this view. Some focused on the continuing heavy presence of generally low-level multiple-choice questions, arguing that the test has changed little overall. As one middle school teacher explained:

> From my perspective, the social studies assessment doesn't seem like it's a change at all. Seems like it's kind of repackaged, kind of dressed up a little differently, but not really different and to me, there is something broken in [teachers' instruction] and we need to fix it This new assessment to me isn't fixing it.

One might argue about whether teachers' practices are "broken," but the sentiment that some state tests, like social studies, seem less changed than others emerged throughout the focus group sessions The English language arts and science tests, in particular, were cited as moving away from a heavy reliance on objective-style questions and toward questions with more real world and practical applications. For example, the English language arts tests asks students to write a range of pieces including technical, literary, and literary analysis essays. The science tests include performance tasks which ask students, for example, to set up a lab experiment. Teachers in these areas had questions about the nature of their respective exams, but there was a general sense that these exams push in more ambitious directions than the social studies tests do.

Social studies teachers see the prospective new state assessments as a mix of old and new. While most applaud the presence of primary sources and questions like the DBQ that ask students to analyze and synthesize information, they wonder if that emphasis won't be undercut by the continuing heavy weight of the multiple-choice section and questions which teachers generally perceive of as asking for low-level knowledge

Opportunities to Learn About the New State Tests

New state tests, like many other educational policies, can be viewed as an occasion to learn about the craft of teaching (Cohen & Barnes, 1993; Grant, in press). The focus group teachers nodded in agreement when participants raised questions such as, "Do I have the skills that I need?" and made assertions such as, "We have not been taught the way we're being asked to teach.... And I think that's really difficult without a lot of staff development to get people to think differently and to teach differently."

If the need for professional development was widely expressed, the teachers' experiences suggested that they may not be getting all that they want. Studies of professional development activities suggest that what session leaders think they are "teaching" and what participating teachers think they are "learning" during professional development activities can vary dramatically (Darling-Hammond & McLaughlin, 1996; Grant, 1997a, Smylie, 1995). Consequently understanding what kinds of professional development opportunities teachers had available to them and what sense they made of those opportunities was a major element of the focus group interviews.

Three patterns emerged from analysis of the interview transcripts. One was that all teachers seemed to have had access to a wide range of professional development opportunities both around the new curriculum standards and around the new tests. A second pattern was that they found those opportunities of uncertain value. Teachers reported that the state, and occasionally district, activities often resulted in incomplete and mixed messages. The frustration many teachers expressed about the more formal professional development opportunities was mitigated, however, by their sense that working more directly with colleagues was a more profitable use of their time. The third pattern, reform by "rumor," began to emerge in the first year of interviews, but was full-blown by the second year. Despite the wide array of professional development opportunities, the teachers clearly felt that there was still much indecision about how tests would ultimately look, how they would be scored, and the like. In a context of increasing pressure to respond, but little solid information, several teachers reported the sense that rumors were driving much of their responses.

The professional development opportunities available. Asked to describe the professional development opportunities available to them, the teachers constructed a long and varied list. Some NYSED-led sessions occurred in several venues (e.g., stand-alone sessions, part of district-level in-services, sessions during professional organization conferences) and focused alternately on the new tests alone or on how the tests reflected the new state curriculum standards. Representatives from local Board of Cooperative Extension Services (BOCES) programs also led professional development activities as stand-alone and district sessions. Some district-level sessions featured state and BOCES representatives, but others utilized the talents of district personnel, while still others brought in local and national experts. School-level professional development opportunities were also varied in that some called all teachers together, while others asked teachers to meet in grade or department-level activities. The focus group teachers also mentioned state teachers' union sessions, college and university course work, professional literature, informal networks, and colleagues as additional sources of information on tests and testing.

The uncertain value of professional development. Of these many sources, teachers were most critical of the state-led sessions. Some felt that cuts in the NYSED have left the agency woefully understaffed. Most others, especially the high school teachers, were less generous. An English teacher said, "I'm not going to break a sweat trying to reformulate what I do when their people (NYSED) don't know what they're doing." A social studies teacher was more blunt: "Do they have a clue as to what's going on?"

District-level sessions received more mixed reviews. A high school mathematics teacher praised her district's efforts to develop professional development activities that would meet teachers' perceived needs:

> My district is real supportive. If I say to them we need an inservice on blah, they will say we'll do it. They're wonderful that way. It's very teacher driven. Our school district is wonderful as far as them involving teachers

and listening to the teachers and valuing what the teachers say.

This comment stood largely alone, however, as most other teachers suggested that district-led professional development was lacking in usefulness A high school social studies teacher noted:

> We've had two district wide superintendent's conference days and we've talked about [the tests] and gone over some things, but not into the detail that needs to be done to get a good feel for the types of questions and changes. I think in our building many people would still be hard pressed to give an accurate reflection of what the assessment is all about.

A middle school science teacher attended a district-sponsored inservice led by a district teacher. She reported that while the session could have been valuable, she left frustrated because the teacher who led the session came from a magnet school where resources are plentiful, whereas she teaches in a resource-starved neighborhood school. Not all the blame for weak district-sponsored professional development was laid at the feet of the leaders, however. A secondary social studies teacher panned the district-level sessions she attended, but she assigned much of that responsibility to her colleagues:

> We went to the district-wide [in-services]. They (the in- service leaders) always tried to be very positive, but the overwhelming number of teachers who are so negative about this assessment always wins out. It basically becomes a complaining session and you really aren't focusing on what the whole meeting was about anyway

The focus group teachers reported that school-, grade-, and/or department-level professional development activities were generally more useful than state or district efforts. An elementary school teacher, for example, praised the work her grade-level colleagues were doing:

> We have grade-level meetings. They're very positive, you know, even though we all don't want to test, we all feel like we shouldn't have to do it They're (her colleagues) always very positive, always very friendly approaching it. Every time we go to a grade level meeting, [the team leader] always is handing us stacks and stacks of information materials. Things that we might need or might be able to use to help the kids get ready, whether it's for the science or the math or the English [tests]. There's always something positive going on

A high school mathematics teacher explained that not only has the amount of conversation increased in her department, but that it is becoming increasingly acceptable to say, "I don't know how to do this " She went on to describe how her colleagues, both veteran and novice, were creating a new ethic whereby the traditional norms of isolation and "doing your own thing" were fading.

Not all teachers are similarly situated, however, and more than any other group, the high school social studies teachers present described their departmental interactions as less than optimal. Several nodded in agreement when an untenured teacher portrayed her colleagues as being obsessed with talk about "how to beat the test, or change the test, or fight the state, or fix the state or ..how is the administration wrong, how are we right " Potentially useful discussions of teaching, learning, and assessment, she explained, get lost in the mix

If teachers found formal state, district, and school-level professional development of uncertain value, all reported instances where informal networks and relationships had proven valuable. A high school social studies teacher said that, while she appreciated some elements of her district staff development days, "it is a lot easier to bounce off the ideas with somebody. And I just wrote [a DBQ] a few weeks ago with a colleague. We have now the same planning period so that worked out " A high school teacher reported that she and her colleagues have met informally after school to consider assessment issues "There were a handful of us that got together after school on a voluntary basis."

BEST COPY AVAILABLE

she said, ". . It makes my life a lot easier when I talk to other English teachers." In addition to these unstructured activities, several elementary school and high school mathematics teachers described informal networks of educators who meet regularly to discuss a range of issues, including those related to testing. A mathematics teacher described the benefits she has appreciated from her involvement:

> We have each other (she laughs) We have a network through (a local state university),where there have to be what-about 70 teachers, maybe 100 maybe that-we have meetings four times a year, and so now I don't feel isolated anymore. I mean I can always call [a colleague in a neighboring district]. I have friends [in another district]. Friends just about anywhere. I know what's going on at what school and I can pool resources, and so that helps a lot.

The power of such informal relationships is apparent: These teachers sense that they are working with peers who hold similar goals and concerns, who are willing to share ideas and practices, and who offer a sense of belonging. Such relationships, then, have an immediacy and a specificity that seems missing from the more formal professional development opportunities teachers typically experience. That these teachers have sought out and participated in these relationships is admirable; that they have felt compelled to do so in order to meet their needs is ironic, however, given the seeming wealth of structured opportunities.

Reform by rumor. Having informal sources of information and support may help teachers navigate some of the challenges the new state tests posed, but they do little to help teachers with the problems of mixed messages and unanswered questions. In fact, the more sources of information teachers encounter, the greater the incidence of reform by rumor.

Common across teachers of all grade levels and subject matters was a frustration with incomplete and conflicting information about the new tests. An elementary school teacher noted, "If we just had more information and if we knew what was expected of us and how to do it, possibly,we could do what was expected of us." A high school mathematics teacher added:

> If they're (NYSED) going to give us information, they have to give it more structured backing. Not this haphazard changing the rules daily.... Our math department head has said [at an in-service led by an NYSED representative], "Tell us what you want. We will do it. We will change the way we teach.... But you can't keep changing the messages you're giving us."

To be sure, state leaders seem to recognize that they are sending multiple and, at times, confusing messages. A high school mathematics teacher reported the following experience during a state-sponsored in-service:

> When we go to state meetings, (the NYSED representative) who's in the math ed department always prefaces his remarks with, "What I'm going to tell you is true at May 13th at 4 whatever. It's true right now. When I go back to my office, it might not be true." And we get to go to a lot of state meetings and everything and find out what's going on. And we always find out the latest stuff, but then it changes.

As this quote suggests, teachers do not necessary blame the state education representatives, but they are frustrated with the uncertainty of the situation. A high school social studies teacher's experience summed up some of the anxiety mixed and multiple messages can induce:

> I don't know if this geography thing (i.e., that the state curriculum and test for tenth grade were changed from Global Studies to Global History and Geography) is true or not. But somebody in my department had been in the state conference the week before and said, "I didn't hear any of this." And

> then we started frantically calling-I think we called the (local state university) Social Studies department, and they were calling all over to find if this was true. And I think the final verdict was that, "yes (geography has been added), but geography the way we've always taught it, so don't be nervous. They (NYSED) are not asking to name which direction the Danube River flows or anything like that." But, I don't know. It's crazy.

This teacher went on to remark, "I see it as just lots of rumors. It's like every other day we're coming in, 'Did you hear they're cutting out the constructed response? Oh, now the new course is Global History and Geography?'"

A cynical interpretation of the above is that teachers are merely pawns in a game that is being transacted all around them. This view asserts that while changing teachers' practices is the target, teachers' ideas and voices are largely ignored as those above them-state and district-level actors-do the real work of policy change. Teachers, through their professional development opportunities, may listen in. But as listeners rather than as full participants, they hear only bits and pieces, and rumors rule the day.

A more generous interpretation has two elements. One is that reforming education is simply hard work, especially when done in midstream, or what a policy maker in another state termed, "rebuilding the airplane while you're flying it" (Lusi, 1997, p. 91). The second element is that, given the sheer number of teachers and the wide range of circumstances in which they work, policy makers face a daunting task in attempting to change pedagogical practices. Whether they should try to or not, the parameters of the NYSED operation are intimidating: thousands of teachers, in thousands of schools, in close to 700 districts, and an agency with little more than a handful of employees. Clearly, then, NYSED must rely on the efforts of proxies-BOCES educators, professional organizations, district and school-level leaders, college and university academics-who may or may not understand and/or support the state agenda. In such a situation, the potential grows for mixed and confusing messages, and for reform by rumor.

The Rationales for and the Consequences of the New NYS Tests

The notion of "reform by rumor" functioned as a proxy for a number of comments where focus group teachers talked about feeling left out of the conversation about changing state assessments. Teachers across grade levels and school subjects expressed frustration that, while they are the professionals on whom the tests will have the most impact, their voices are not well reflected in important discussions about the nature, import, and design of new state tests. As one teacher said, "I really fear that unless there's open communication...this whole thing would be just kind of a charade." Another added, "I just feel that I've been talked at."

These teachers remain uncertain about the rationales for and the consequences of the state assessments, but seek to question rather than condemn. Most said they have attended meetings designed to inform them about the tests, but none said they were satisfied: Their questions either went unaddressed or, if they were addressed, the information they received did not always jive with information circulated previously. While numerous questions arose during the focus group interviews, two dominated questions about the rationales for changing the assessments and questions about the intended and unintended consequences of the tests.

Questioning the rationales for the tests. Whether the NYSED hopes to induce changes in teachers' curriculum decisions, their instructional practices, or both has been unclear for some time (Grant, 1997a). The focus group teachers echoed this confusion. They also discussed their uncertainty about whether the state's intention was to change their behavior or the students'. As a middle school social studies teacher said, "Are they (NYSED) doing this to better students' education, or are they doing it so they can say, 'Look, we changed something.'"

On the question of whose behavior NYSED is targeting, teachers expressed considerable frustration. For instance, an elementary teacher asked, "Who is it assessing? Is it really assessing the students? Or is it assessing the teachers?" Another elementary teacher echoed this point: "What is the purpose of the state exams? Is it actually to assess the students or to push the teachers in a direction?" A secondary social

studies teacher spoke directly to the issue of whose life is changing the most as a result of the new state tests:

> I think it's ironic that the state came out with all of these decisions in order to improve student learning and to make students better students and.I feel like I am doing so much work this year. When I do essays, I try to fix things and give them lots of responses and they just-I feel like I'm doing more work than the kids sometimes.... The last couple weeks it's like "I'm not taking this test! I took this test!" This is you. Not me. But it seems like the teachers are on the chopping block. And it's just ironic that it's no longer the student anymore. And it's the kids who are taking the test. And it seems like the kids are almost less and less responsible....

The last part of the quote above suggests that the issue of whether teachers or students are targeted is important, in part, because teachers are unsure where the blame is going to come down should test scores not rise. Many suspect, however, that teachers will take the brunt of the criticism. A high school mathematics teacher said, "They're (local administrators) are going to be pointing their finger if your kids don't do well. They're going to be pointing their finger at those teachers and that's unfortunate because they're (the teachers) going to be a scapegoat because of it." A secondary English teacher talked about the unfairness of holding the teachers whose students are taking the tests entirely responsible for the outcomes:

> I think that whole culture needs to change because you are not the sole responsible party for that student's abilities.... If someone did a lousy job last year, then you're getting a group of students without the proper foundation. And is there going to be some kind of mechanism that will address that if you realize that the child did not get proper foundation?...There's no way I solely am responsible for that child's [test scores]. I've had students who are functioning very very low and you're asking me to.. bring that child further along. Is that child going to pass that test? No. So you're going to come to me and say, "Well, only 55% of your students passed this test. You're lousy!" I'm going to say, "Well, what did you give me?"

This quote raises a number of thorny issues, not the least of which is a seeming deficit view of children. This view implies that students come to a teacher with a set of deficiencies, resulting from poor parenting, poor schooling, and the like, which the teacher must then "correct." The problems with this view are several, but in this case, they serve to amplify the dilemma this teacher faces: She feels the twin burdens of preparing students to take the exam and of being held accountable for their performance. Although it seems unfair to make the child the pawn, this teacher rightly points out that she alone can not be responsible for test scores

Teacher frustration was also apparent around the question of whether NYSED's intent was to change curriculum, instruction, or both. The focus group teachers assumed the tests were meant to induce changes, but they were unsure what sort of change was expected

A secondary social studies teacher saw the state's aim as primarily directed toward curriculum

> But it looks like -- the more I hear about it it's as if the state through its tests is controlling what gets taught in the classroom By saying that the test is going to be done this way, all of a sudden it's going in and saying well you can't teach this, this, and this when you want to You have to teach this. You have to teach this.

An elementary teacher, by contrast, suspected that the state's intention is to influence teachers' instructional practices

> Is this a way of making teachers look at their practice and alter their

> teaching techniques because they see a certain topic being covered on an exam and so they'll say, "Oh, I didn't do that so well that time. I guess I have to spend more time on that next year." So if you see the focus on the exams, then you've got to go back and make sure that you include that type of instruction the next year. And so I think-are the tests pushing-is the state using the test to push teachers in a certain direction with their instruction?

While most of the focus groups sensed that the state tests were being used to leverage change of one sort or another, not all did. A high school English teacher reported that she had been told, "We've been doing this all along. That this is no big deal...all we have to do is get kids accustomed to the format [of the test]." A secondary science teacher added to this notion, by reciting a familiar teacher expression, that is, "this too shall pass." "In our science department," he said, "they feel because science is the last assessment [to be introduced] that this is all going to blow over." The notion that whatever NYSED introduces is likely to fade in importance over time was not the dominant view among the focus group teachers. But its expression should warn state-level reformers that whatever leverage they believe tests hold for changing instruction and/or curriculum may be illusory. This is not because teachers do not sense that problems exist: None of the focus group teachers was willing to suggest that all is right with public education. But several supported the following sentiments of an elementary school teacher who questioned the reliance on tests as a lever of real instructional change:

> I understand that certainly there are places in American education that are in dire need of shaping up somehow....It (the test) just seems to me a misdirection of resources. We're spending how much--thousands of dollars on training, on writing these tests or whatever they're doing to when the real issue is what's happening in the classroom. What kind of preparation are teachers getting? What kind of preparation are they getting before they even get a classroom? What kind of thinking is going on here? And are those questions even being asked? Or were they ever asked before this happened? It was just suddenly that we had this massive assessment. And I don't remember any sort of input from teachers. I don't remember any state education people coming to us and saying, "What do you think?" Or, "What's going on in your classroom?" It was just this kind of mandated attempt to reform. And maybe it will work. I mean, I don't know whether it will work or not. But it seems to me there's so much more that could be done that hasn't been attempted in terms of helping teachers.

To be fair, NYSED officials and the state Board of Regents have proposed a range of reforms that push changes in curriculum and in teacher education. The primacy of the state testing program, however, weighs heavily. The focus group teachers are not opposed to improving teaching and learning, but they are uncertain about the rationale for standardized tests as a vehicle.

Predicting the consequences of the new tests. The idea that the new tests may yield no real consequences for teachers' practices was one of several predictions the focus group teachers made. Most of those predicted consequences were negative, but not all. For example, several teachers in the first year focus groups expressed the hope that the tests would mean greater collaboration with their colleagues. A high school English teacher summed up the feeling: "If there were more opportunities to get more people together, that would help." While it was far from unanimous, a number of the year two teachers reported that, in fact, they had found their peers receptive to and interested in working together.

The overwhelming sentiment, however, was that the new tests could produce undesirable effects. Those effects grouped loosely around issues of pedagogy, students, and teachers.

Two related consequences of tests for pedagogy arose. One is that, rather than promote more ambitious teaching and learning, the state tests may actually push more reductive forms of teaching and learning. The most common expression was that teachers felt increased pressure to tailor one's teaching to the test parameters. As a

secondary social studies teacher noted, "You've got people in high places just saying 'teach to the test.'" A middle school English teacher complained that he felt pressure to "teach them (students) test terminology when I could be teaching them other things." This teacher went on to describe the kind of support his district provides as little more than practice exercises. "The only thing I've gotten from my district," he said, "is lots of practices. Every week there's, 'Thank so and so for giving this practice material. Here's another listening practice that you may want to use.' I could have spent my whole year doing practices."

The sense that teachers feel pressed to adopt direct teaching approaches as a means of bolstering short-term test performance was in direct competition with the sentiments expressed earlier that the new state tests could be viewed as supportive of more ambitious instruction. During the interviews, however, no teacher commented on this seeming contradiction. One explanation is that they were simply unaware of its emergence. A more interesting possibility is that these teachers can read multiple messages in the tests. Take social studies as an example. Teachers thinking about the multiple-choice questions could reasonably assume that a more traditional, direct instruction approach was being encouraged. If those same teachers were thinking instead about the DBQ questions, it seems equally reasonable to assume that richer forms of pedagogy were intended. This ambivalence, which has surfaced in a number of places already, underscores the difficulty in understanding teachers' perceptions of state tests and it suggests that their classroom responses may be more complex and textured than reformers may want or expect.

A second potentially negative consequence of the new tests was an increased emphasis on remediation as a way to deal with low test scores. The teachers, especially those in the second year interviews, described a wide array of remedial approaches taken in their schools. Those approaches included additional classes designed for students presumably at risk of failing, summer and Saturday test review courses, hiring additional teachers and aides to staff learning labs where students could either come voluntarily or by teacher assignment, and reassigning teachers to classes of students based on their perceived ability to help those students pass the exam.

The teachers offering these examples generally seemed supportive of them. The seeming contradiction that ratcheting up remedial efforts would occur at the same time teachers were being pushed to change their pedagogy went unremarked upon. Again, however, this contradiction may be less apparent than one might suspect. Empirical evidence is surprisingly thin on the question of which instructional approaches lead directly to high test scores (Cohen & Barnes, 1993; Grant, in press). Consequently, a reasonable response to a new testing situation might be both to make changes in "regular" classes and to begin planning for remedial instruction at the same time.

The real danger, however, is that these remedial opportunities will become little more than drill sessions, a point that was recognized by several teachers. For example, a high school mathematics teacher observed.

> If the students do not pass, they're going to be remedied with questions that will make them pass. So eventually every student will pass. Doesn't matter the categories, they're going to do component retesting, so if the student doesn't do well in these three areas, they'll be grilled in those three areas with a bank of questions, and then the student will have another test from the bank that he was drilled in. So eventually they'll get it

Such an approach may work for low-level skills, but is of dubious use in areas like social studies where conceptual knowledge is central. As VanSledright & Brophy (1992) observed, "naive but imaginative accounts persisted in some children even after direct instruction designed to change them" (p. 854). Without any definitive research supporting one means of improving test performance over another, drill and practice remediation is as likely to flourish as any other approach.

A second area of negative consequences anticipated by the focus group teachers concerned students. An elementary teacher worried generally that the net effect of a high profile, high- stakes testing program would be a "nation of test-takers".

> Something that I've been thinking about more is the effect this has on the

> children, on the student. What kind of learners is this going to shape? Are we producing a nation of test-takers, and if so, are those test-taking techniques or skills what we need to produce life long learners that we talked about before?

Other teachers expressed more focused concern about the anticipated consequences for urban students. Wiles (1996) argues that test performance is clearly distributed along socio- economic lines with upscale, white suburban children consistently outscoring their urban and minority peers. The focus group teachers, both urban- and suburban-based, recognized the inherent threat that high-stakes testing poses for some children. An elementary school teacher said, "I'm very concerned about some of the larger populations in the bigger urban areas. I don't understand how this is going to positively affect these kids." A high school teacher, commenting on the anticipated testing of special education students, asked, "How do we accommodate the non-standard kids on a standardized test?"

No teachers thought their students' scores on the new tests would improve immediately over past test scores. A couple of teachers did express, however, the hope that their students' scores would increase over time. A middle school English teacher said, "I think, naive though it may be, that our kids are going to do better ultimately on these exams. Maybe not this year, but ultimately."

This hopefulness stood in stark contrast with the prevailing view that teachers anticipated problems for their students. Underlying both these sentiments is a harsh truth: These teachers simply do not know how their students will perform on the new tests. Given the general tendency for a correlation between test scores and students' social capital, it is difficult to understand why suburban teachers would be worried. And yet, analysis of the relative concern expressed by suburban vs. urban teachers suggested that suburban teachers and administrators may be even more concerned about potentially low scores than their urban peers. One proxy for this finding is the observation that the overwhelming number of remedial efforts planned are being developed in suburban schools.

As noted above, no teacher feels s/he has an inside track on what approaches will insure high scores. Left to follow one's hunches, it is no particular surprise to find concern among all teachers, both suburban and urban. But what explains the fact that suburban teachers seem to be more concerned about their students' performance than their urban peers? Part of an explanation must consider the notion that not all suburban districts are created equal. The suburban teachers in focus group teachers represented first-, second-, and third-ring suburbs. First-ring suburbs tend to include a range of working to middle class students. Second-ring suburbs are more upscale; most students come from middle to upper-middle class homes. Finally, the third-ring suburbs are rural areas that recently have attracted a large number of middle and high SES families. With the exception of one or two urban magnet schools, it is the schools in the second- and third-ring suburbs that consistently rank in the top quartile according to a highly publicized local business magazine. Top quartile spots on this list have real consequences for real estate values, bragging rights, and the like, and so the scramble to move up can be intense. New tests, then, represent a potential threat to schools' past standings. School people in high performing schools want to maintain their positions, educators in middle and low performing schools hope to at least avoid dropping further.

The competition for high test scores plays out as a third set of consequences. Here, the focus is on the pressure and uncertainty teachers feel as they decide if and how to modify their teaching based on their perceptions of the state test. A couple of these pressures have already been described. One is the feeling of uncertainty teachers have about which approaches will ensure higher scores. A second pressure surfaces as teachers report being made to feel entirely responsible for their students' results. Putting the point on this feeling is a secondary social studies teacher:

> Just this week I was called down to the office and we were comparing some of the *Business First* statistics that were out just recently....So according to our administration [if we get low test scores] people come out to vote and decide they don't want to vote on the budget, therefore the whole

> community goes down. So, I left the office thinking the weight of this town...is on my shoulders. Whether or not, you know, my kids pass. And we had like a 70% last year and we're expected to have at least a 90 if not higher. So, in terms of administration, testing is a pretty big deal.

Not all principals apply pressure so directly, but many apparently do. This is more likely to happen in high schools than elementary schools, however. According to several of the focus group elementary school teachers, their principals are more likely to talk about test scores as part of a bigger picture of how students are progressing. These teachers do not necessarily feel any less pressure than their high school peers, but one source of pressure, the school administrator, seems to be less of a factor.

The new elementary school exams are more high-stakes than they used to be: recall that now individual student scores will be reported rather than group scores. The stakes are even higher in the high schools, however, as passing the Regents exams will be necessary in order to graduate. Consequently, it is not hard to understand why high school administrators might be more likely than their elementary peers to put pressure on their teachers. Whether that tactic will pay off ultimately or not is hard to predict. But one manifestation of that pressure is to cause teachers to consider issues that they probably have not had to think about in the past. One particularly compelling story came from a high school social studies teacher who said she now wonders about each new student who comes into her classes:

> I never--it never crossed my mind before that a certain kid was going to lower my passing rate or not, and I actually started thinking about that this year. And I was so ashamed of myself about that. And one of the girls I had transferred from a general track. She stayed in my class. I didn't want to just dump her. But she can now take the RCT at the end of the year. But I had a girl a couple years ago who transferred from another state. She never had Global 9. And I was just happy to work with her and she was going to try it. And if you go to look at an individual kid and say they're not going to do it, it's horrible to think that--to individualize it like that. Because I guess every couple kids knocks you down a little bit. And our--I know that our department chairs had our results individualized and our principal keeps coming into meetings saying, "How can we raise this up? How can we do this better?"

This teacher concluded her story with a nervous laugh, saying, "But I'm glad I have tenure, right?" Yet, having tenure seems little consolation for this thoughtful and dedicated teacher now confronted with the dilemma of wanting to work with all students, but recognizing that doing so may cause her teaching to be called into question should her students' scores not measure up.

Not all the consequences described were negative, however. Several teachers cited greater collaboration with their peers as a key benefit of the new tests. Elementary teachers and high school mathematics and English teachers were most vocal on this point. "I think we have so much to learn from each other," one elementary teacher said. Another echoed this point, commenting, "We're really trying to deal with this [new tests] and trying to work as a faculty to help each other." A high school English teacher noted that information is vital and that colleagues are an important source. "What's most important to me is being able to communicate with other people so I can get some information." A high school mathematics teacher concurred, but pointed out that that the new exams were forcing teachers to rely on each other:

> I think the nature of the testing--it certainly sets the situation up for teachers to talk. Because the types of questions that happen to be asked. They don't have the stockpile of old Regents questions. So [teachers say] "I came up with this. You know, I'm going to use this." We can share, and the nature of the beast is forcing the issue.

Social studies teachers reported some positive collaborations with peers, but they also cited more instances than the other teachers of situations where friction had developed

A high school teacher described the tension that arose over course assignments:

> We have attempted to get together and work, but what we have found out has been happening is just been a lot of back- stabbing and a lot of animosity because there are a couple of teachers who just adamantly refuse to teach 10th grade (when the Global exam is administered). So the feeling is, well, they can do the ninth grade program. But where is their accountability? Because they just will not do that 10th grade when their kids take the Regents at the end of the year.

This teacher's experience points, again, to the variability in the way consequences of the test are playing out. This variation is explained, in part, by the development of as many unintended as intended consequences. State-level reformers may have hoped, for example, that teachers would see the test as an impetus for more ambitious instruction, closer collaboration, and the like. And this seems to be occurring. But reformers probably did not predict the more negative consequences these teachers are seeing. That these outcomes are unintended is little solace, for they may be just as real to the teachers as the intended outcomes. Actually, these unintended consequences may ultimately be more important because they seem to receive scant attention from state and district-level actors. State and district leaders may be unaware of these issues, they may be ignoring them, or they may not see them as problems. In any event, it seems interesting that no teacher mentioned that s/he had participated in any explicit conversations about the problems they anticipated. As noted above, teachers did see positive possibilities arising from the new state tests and there was no particular sense of gloom during the interviews. How teachers will manage the more negative consequences is unclear, but the supposition that they will have no effect seems naive.

Implications

Substantive change is always unsettling. So reform on the scale that New York state is attempting, in all grades and in all school subjects, is bound to generate some frustration, anxiety, and uncertainty. The findings above tell us that while teachers are not adverse to change, they have real concerns about the nature of the changes proposed, the professional development opportunities available to learn about these changes, and the rationales for and consequences of the new state tests.

Given the complexities of teaching and policy (Grant, 1998), it is not surprising to learn that teachers see both prospects and problems in the new NYS tests. State-level policymakers in New York, like most of their peers, are attempting reform on a massive level (Elmore, 1997) and are doing so with relatively few levers for change. What this study suggests is that teachers are not passive participants and must not be designed around. The dream of teacher-proof curriculum as a means of changing teachers' practices has proven to be a myth (see, for example, Dow, 1991; Schwille, Porter, Belli, Floden, Freeman, & Knappen, 1983). Faith in tests as a means of corralling teachers' practices may ultimately prove just as chimerical as long as teachers are left out of the loop. If any of the changes state reformers propose are to stick, then these teachers are saying they need to be more actively involved in the formulation of those changes. But there is something else. These findings also suggest that there are real and important differences in the ways teachers perceive reforms across grade levels. Among other things, this means that reformers can not take a one-size- fits-all stance and that professional development needs to be sensitive to the differences in the perceived needs of teachers.

Notes

The author wishes to acknowledge Bob Stevenson's thoughtful comments on an earlier draft of this article.

1 The TLA study is funded by the Collaborative Research Network, sponsored by the Graduate School of Education at SUNY-Buffalo. The faculty and students who worked on this study include Suzanne Miller, Robert Stevenson, Mark

BEST COPY AVAILABLE

Templin, Meg Callahan, Diana Lawrence-Brown, and Gina Trzyna.

2. The small number of elementary school teachers was due partly to design and partly to exigencies that prevented the other invitees from attending on that date.
3. Corbett and Wilson (1991) point out, however, that Madaus's claims are based on limited data: "anecdotes, testimony from public hearings, historical accounts, and an occasional international study" (p. 26).
4. Revisions of state tests is still in progress so some of what follows is based on SED reports of changes they expect will occur.
5. The first administrations of new social studies tests will begin in the fall of 2000.
6. For example, in the test sampler for the Global History and Geography exam (New York State Education Department, 1999), students would be given documents that range from a poem by Lao Tzu; portions from Pericles' "Funeral Oration," the English Bill of Rights, the Japanese Constitution, a speech by Benito Mussolini; and a political cartoon about the monarchy in France during the 1600-1700s. They are then directed to write an essay in which they "compare and contrast the different viewpoints societies have held about the process of governmental decision making and about the role of citizens in the political decision-making process" and to "discuss the advantages and disadvantages of a political system that is under the absolute control of a single individual or a few individuals, or a political system that is a democracy" (p. 25).

   A test sampler in NYS consists of a description the types of test items, sample questions, a breakdown of the number of questions by curriculum standard and topic, rubrics for essay questions, and sample student responses.

   At present, the only test sampler available is that for tenth grade Global History and Geography. The first administration of that test is scheduled for June 2000. Test samplers for the grades 5 and 8 tests are to be available this fall with administration of the grade 5 test scheduled in November 200 and the grade 8 test in June 2001. The test sampler for the grade 11 test is due out in spring 2000 and the new test is scheduled for June 2001.
7. From the Global History test sampler (New York State Education Department, 1999), students are given this theme on belief systems: "At various times in global history, members of different religions have acted to bring people together. Members of these same religions have also acted to divide people and have caused conflict." Students are then directed to this task: "Choose two religions from your study of global history and geography. For each religion: Describe two basic beliefs of the religion; Explain how members of the religion, at a specific time and place, acted either to unify society or to cause conflict in society" (p. 29)
8. The PET tests were given at grades 6 and 8. The new tests will be administered at grades 5 and 8.

References

Bogdan, R., & Biklen, S. (1982). *Qualitative research for education: An introduction to theory and methods*. Boston: Allyn and Bacon.

Cohen, D., & Barnes, C. (1993). Pedagogy and policy. In D. Cohen, M. McLaughlin, & J. Talbert (Eds.), *Teaching for understanding: Challenges for policy and practice* (pp. 207-239). San Francisco: Jossey-Bass.

Comfort, K. (1991). A national standing ovation for the new performance testing. In G. Kulm & S. Malcolm (Eds.), *Science assessment in the service of reform* (pp. 149-161). Washington, DC: American Association for the Advancement of Science.

Corbett, H. D., & Wilson, B. (1991). *Testing, reform, and rebellion*. Norwood, NJ: Ablex.

Darling-Hammond, L., & McLaughlin, M. (1996). Policies and support professional development in an era of reform. In M. McLaughlin & I. Oberman (Eds.), *Teacher learning: New policies new practices* (pp. 202-219). New York: Teachers College Press.

Dow, P. (1991). *Schoolhouse politics: Lessons from the Sputnik era*. Cambridge, MA: Harvard University Press.

Editors. (1994). Introduction. Symposium: Equity and educational assessment. *Harvard Educational Review, 64*, 1-3.

English, F. (1980). *Improving curriculum management in the schools*. Washington, DC: Council for Basic Education.

Feltovich, P., Spiro, R., & Coulson, R. (1993). Learning, teaching, and testing for complex conceptual understanding. In N. Frederikson, R. Mislevy, & I. Bejarq (Eds.), *Test theory for a new generation of tests* (pp. 181-217). Hillsdale, NJ: Lawrence Erlbaum.

Finn, C. (1995). Who's afraid of the big, bad test? In D. Ravitch (Ed.), *Debating the future of American education: Do we need national standards and assessments?* (pp. 120-144). Washington, DC: Brookings Institution.

Firestone, W., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*(2), 95-113.

Freeman, D., Kuhs, T., Porter, A., Knappen, L., Floden, R., Schmidt, W., & Schwille, J. (1980). *The fourth grade mathematics curriculum as inferred from textbooks and tests*. East Lansing, MI: Institute for Research on Teaching, Michigan State University.

Fuhrman, S. (1993). The politics of coherence. In S. Fuhrman (Ed.), *Designing coherent education policy: Improving the system* (pp. 1-34). San Francisco: Jossey Bass.

Fuhrman, S., Clune, W., & Elmore, R. (1988). Research on educational reforms: Lessons on the implementation of policy. *Teachers College Record, 90*, 237-258.

Glaser, B. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press.

Glatthorn, A. (1987). *Curriculum leadership*. Glenview, IL: Scott, Foresman.

Grant, S. G. (1996). Locating authority over content and pedagogy: Cross-current influences on teachers' thinking and practice. *Theory and Research in Social Education, 24*(3), 237-272.

Grant, S. G. (1997a). Opportunities lost: Teachers learning about the New York state social studies framework. *Theory and Research in Social Education, 25*(3), 259-287.

Grant, S. G. (1997b). A policy at odds with itself: The tension between constructivist and traditional views in the New York state social studies framework. *Journal of Curriculum and Supervision, 13*(1), 92-113.

Grant, S. G. (1998). *Reforming reading, writing, and mathematics: Teachers' responses and the prospects for systemic reform*. Mahwah, NJ: Lawrence Erlbaum.

Grant, S. G. (in press). An uncertain lever: The influence of state-level testing in New York on teaching social studies. *Teachers College Record*.

Heubert, J., & Hauser, R. (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press.

Kellaghan, T., Madaus, G., & Airasian, P. (1982). *The effects of standardized testing*

Boston: Kluwer-Nijhoff.

Koretz, D. (1988). Arriving in Lake Woebegon: Are standardized tests exaggerating achievement and distorting instruction? *American Educator, 12*(2), 8-15; 46-52.

Koretz, D. (1995). Sometimes a cigar is only a cigar, and often a test is only a test. In D. Ravitch (Ed.), *Debating the future of American education: Do we need national standards and assessments?* (pp. 154-166). Washington, DC: Brookings Institution.

LeCompte, M., Preissle, J., & Tesch, R. (1993). *Ethnography and qualitative design in educational research.* (2nd ed.). New York: Academic Press.

LeMahieu, P. (1984). The effects on achievement and instructional content of a program of student monitoring through frequent testing. *Educational Evaluation and Policy Analysis, 6*(2), 175-187.

Lusi, S. F. (1997). *The role of state departments of education in complex school reform.* New York: Teachers College Press.

Madaus, G. (1988). The influence of testing on the curriculum. In L. Tanner (Ed.), *Critical issues in curriculum: 87th yearbook of the NSSE, Part 1*. Chicago: University of Chicago Press.

National Center for History in the Schools. (1994). *National standards for United States history.* Los Angeles: Author.

National Council for Social Studies. (1994). *Expectations of excellence.* Washington, DC: Author

Natriello, G., & Pallas, A. (1998). The development and impact of high stakes testing. Available on-line. http://www.columbia edu/~gjn6/histake.html

New York State Education Department. (1996) *Learning standards for social studies* Albany, NY: Author.

New York State Education Department. (1998). *Social studies resource guide.* Albany, NY: Author.

New York State Education Department. (1999). *Global history and geography Regents examination: Test sampler draft.* Albany, NY: Author.

Noble, A., & Smith, M. L. (1994). Old and new beliefs about measurement-driven reform: "Build it and they will come." *Educational Policy, 8*(2), 111-136.

Popham, J., Cruse, K., Rankin, S., Sandifer, P., & Williams, P. (1985). Measurement-driven instruction: It's on the road *Phi Delta Kappan, 66*(9), 628-634.

Popham, W. J. (1998). Farewell, curriculum: Confessions of an assessment convert *Phi Delta Kappan, 79*, 380-384.

Ravitch, D. (Ed.). (1995). *Debating the future of American education: Do we need national standards and assessments?* Washington, DC: Brookings Institution.

Resnick, D., & Resnick, L. (1985). Standards, curriculum, and performance. A historical and comparative perspective. *Educational Researcher, 14*(4), 5-20

Salmon-Cox, L. (1981). Teachers and standardized achievement tests. What's really happening? *Phi Delta Kappan, 62*, 631- 634.

Schwille, J., Porter, A., Belli, G., Floden, R., Freeman, D., & Knappen, L. (1983)

Teachers as policy brokers in the content of elementary school mathematics. In L. Shulman & G. Sykes (Eds.), *Handbook of teaching and policy* (pp. 370-391). New York: Longman.

Shanker, A. (1995). The case for high stakes and real consequences. In D. Ravitch (Ed.), *Debating the future of American education: Do we need national standards and assessments?* (pp. 145-153). Washington, DC: Brookings Institution.

Smith, M., & O'Day, J. (1991). Systemic school reform. In S. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing* (pp. 233-267). New York: Falmer.

Smith, M. L. (1991). Meanings of test preparation. *American Educational Research Journal, 28*(3), 521-542.

Smith, M. L., Heinecke, W., & Noble, A. (1999). Assessment policy and political spectacle. *Teachers College Record, 101*(2), 157-191.

Smylie, M. (1995). Teacher learning in the workplace: Implications for school reform. In T. Guskey & M. Huberman (Eds.), *Professional development in education: New paradigms and practices* (pp. 92-113). New York: Teachers College Press.

Stiggins, R., & Conklin, N. (1992). *In teachers' hands: Investigating the practices of classroom assessment.* Albany, NY: SUNY Press.

Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia.* Cambridge: Harvard University Press.

Wiles, D. (1996). *Networking high performance in New York's secondary education: The Regents curriculum story.* Lanham, MD: University Press of America.

Wolf, R. (1998). National standards: Do we need them? *Educational Researcher, 27*, 22-5.

About the Author

S. G. Grant
517 Baldy Hall
State University of New York at Buffalo
Buffalo, New York 14260

Email: sggrant@acsu.buffalo.edu
Telephone. 716-645-6493

S. G. Grant is an assistant professor of Social Studies Education in the Department of Learning and Instruction. He has published papers in both social studies and general education journals. His most recent journal publications have been in *Theory and Research in Social Education* and the American Educational Research Journal. In the fall of 1998, he published his first book, *Reforming Reading, Writing, and Mathematics: Teacher's Responses and the Prospects for Systemic Change*. An article on the influence of state-level tests on teachers' classroom practices is forthcoming in *Teachers College Record*.

Appendix

FOCUS GROUP PROTOCOL

Spring, 1998

- Introduction: Why we are here. Guidelines and ground rules.

- METAPHORS

  Moderators and participants introduce themselves to group.

  To get started, introduce yourself to someone next to you and describe an image or metaphor that characterizes your thinking and/or feelings about the new state assessments.

  After they have shared in pairs, have them share their metaphors with the group.

  Have participants discuss and elaborate on the metaphors. Lead a discussion of the metaphors. What do they say about our thinking? Common features? Significant differences.

  Direct the discussion toward the next question-what do these assessments mean to you?.

- MEANING OF ASSESSMENTS

  What do/will these assessments mean to you? Your school? Your students?

  Transition to next question-are you prepared to deal with these implications?

- BEING PREPARED

  How prepared to deal with these assessments do you feel? How are you being prepared? What are you being prepared for? What opportunities do you have to talk about the assessments and related issues?

  Build on these expressions to move toward a discussion of needs.

  What help do you need?

  This discussion should lead naturally to talk of challenges.

- CHALLENGES

  What challenges/concerns do you anticipate? How will you deal with these challenges/concerns? Who do you expect will help you?

- CLOSURE

  What has this conversation made you think about concerning teaching and testing (e.g., issue, question, new image)?

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (480-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina -- Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois -- UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
aacosta@cucea.udg.mx

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Hugo Casanova Cardiel
Universidad Nacional Autónoma de México

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

This article has been retrieved 2396 times since February 24, 2000

archives | abstracts | editors | board | submit | comment | subscribe | search | epaa.asu.edu

Education Policy Analysis Archives

Volume 8 Number 14 February 24, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Teachers and Tests: Exploring Teachers' Perceptions of Changes in the New York State Testing Program

S. G. Grant
State University of New York at Buffalo

Abstract

How do teachers change their pedagogical practices? While many current initiatives seek to raise educational standards and improve student academic performance, there is a curious gap in national and state reforms. Considerable attention is given to defining higher expectations for what students will know and be able to do, yet little attention is given to how teachers should learn new pedagogical ideas and practices. This exploratory study uses focus group interview data collected over two years to examine how cross-subject matter groups of elementary and secondary New York State teachers respond to one way of learning to change their classroom practices: state-level testing. Analysis of the data highlights three issues: the nature and substance of the tests, the professional development opportunities available to teachers, and the rationales for and consequences of the state exams

Many current initiatives seek to raise educational standards and improve student academic performance. Yet, there is a curious gap in the recent talk about national and state reforms. While much attention focuses on defining higher expectations for what students will know and be able to do, little attention is given to how teachers should learn new pedagogical ideas and practices. Such policies as the federal *Goals 2000: Educate America Act* and the New York *New Compact for Learning* focus on the resources, conditions, and practices necessary for all students to learn. None of these efforts, however, seriously addresses how experienced teachers will learn the intended innovations.

How do teachers change their pedagogical practices? Some suggest change comes through new subject matter standards proposed by professional organizations (National Council for Social Studies, 1994), by national groups (National Center for History in the Schools, 1994), or by state education departments (New York State Education Department, 1996). Others believe teachers change their practices in response to organizational restructuring (e.g., smaller classes, block scheduling). Still others assert that real change in the classroom lives of teachers and students depends on changes in state-level assessments (Comfort, 1991; Smith & O'Day, 1991). The assumption in this last case is that testing drives much of what teachers do, and so curricular and instructional change will occur if and when state tests change.

This last idea is intriguing for, if true, it suggests the potential for big pedagogical changes with a modicum of policy effort: Change the test and one changes teachers' practices. New York state policymakers seem taken with this approach, for although they have developed new curriculum standards, it is revision of the state testing program which gets most of the attention (Grant, 1997a). The scope of that revision is wide. One piece is the change from program evaluation tests at the elementary level to high-stakes individual student testing. A second piece is the phase-out of the less demanding high school Regents Competency Tests and the requirement that all students pass the more demanding Regents tests. A third piece is a change in the content and format of all state tests presumably to reflect the higher expectations expressed in the state's new standards documents.

What sense do teachers make of these new state tests and how, if at all, do the tests influence their classroom practices? Strange as it seems, there is little empirical evidence to suggest how teachers, especially teachers at different grade levels, respond to changes in state tests. Assessment is a particularly hot topic in educational circles today, yet there is surprisingly little research which digs deeply into teachers' understandings of the import of standardized tests (Cohen & Barnes, 1993; Grant, in press). Corbett and Wilson's (1991) study of teachers' reactions to a new Maryland testing program is well-known as is the on-going work of Mary Lee Smith and her colleagues in Arizona (Noble & Smith, 1994; Smith, 1991; Smith, Heinecke, & Noble, 1999), but these are few studies in a field that is more prone to study students' responses than teachers'.

In this article, I use the data collected through focus group interviews over two years to explore the relationships between teachers and tests. My findings suggest that teachers need to be much more involved in the process of changing state assessments, and that professional development needs to be more attuned to the different needs teachers have

The Study

The Teacher Learning and Assessment (TLA) research project (Note 1) is designed to look generally at the intersection of teachers and assessments. The research team is a cross-subject matter group of faculty and students (English, mathematics, science, and social studies) who are interested in exploring the relationship between teacher learning and state-level testing. Our study questions include: a) In what ways are tests and test results used in classrooms, schools, and the districts? b) What do the proposed changes in state-level tests mean for teachers and learners? c) How are teachers being prepared to respond to the new state assessments? and d) What challenges do teachers and administrators anticipate in moving toward new state assessments? In each case, we are interested in the extent to which these issues differ across school subject matters and grade levels.

Data Collection

In the first year of data collection, we organized two focus groups, one composed of 7 elementary school teachers and counselors and one composed of 12 high school teachers. The participants represented a cross-section of urban, suburban, and rural school districts in western New York state, a breadth of teaching experience (2-25 years), and a range of school subjects (language arts, mathematics, science, and social studies). Each of the two-hour focus group interviews was tape-recorded and transcribed.

During the second year of data collection, we again organized separate elementary and secondary focus groups. We debated whether to: a) reconstitute the original groups only; b) develop new groups of teachers separate from those involved in the first year's interviews; or c) call together groups that mixed teachers new to the project with those who had participated during the previous year. We rejected the first option, fearing that attrition might leave us with groups that were too small. We also rejected the second option, though largely because of timing: We did not think we could hold four focus groups near the end of the school year. In the end, we decided to constitute mixed groups for two reasons. One reason was that we wanted to expand the number of teachers we were talking with; the second reason is that we were interested in how the two groups might interact. The secondary focus group consisted of 8 teachers representing mathematics, science, English, and social studies; 5 of the 8 were in the original sample. The elementary focus group consisted of 5 teachers, 3 of whom were in the original sample. (Note 2)

The data consist of interview transcripts of the focus group sessions and post-interview evaluations completed by the participants. The focus group interviews followed a semi- structured interview protocol (see Appendix). Questions used during the first year asked participants to construct a metaphor to represent their sense of the changes in state-level testing, what the new tests mean for teaching and learning across school subjects, how teachers are being prepared for new standards and new assessments, and what challenges teachers believe they face. The post-interview questions asked the participants to reflect on the issues raised around the relationship between state-level assessment and classroom practice. The interview protocol was largely the same during year two. Changes consisted of replacing the metaphor task with a fill-in-the-blank exercise ("I used to think of the state assessment as _____. now I [still] think of it as _____.") and the addition of probes that asked participants if they sensed a change from last year to the present. There were no changes to the post- interview evaluation.

Data Analysis

All data were analyzed inductively from an interpretivist stance (Bogdan & Biklen, 1982; LeCompte, Preissle, & Tesch, 1993). That stance emphasizes the importance of context, and the multiple ways individuals construct meaning. All data were also analyzed using a constant comparative method (Bogdan & Biklen, 1982; Glaser, 1978). That method assumes that data collection and analysis are recursive, one informing the other throughout the course of the study. After coding the data both within and across grade levels and subject matters, I began seeking patterns in the informants' responses. The themes which emerged reflect the full data set, but in each case I highlight the implications for social studies.

Although this data can be considered largely exploratory, patterns and themes surfaced as the interview and evaluation data were analyzed related to the research questions. In the analysis of the focus group interviews, I focused on: how teachers make sense of, and make different sense of, the state curriculum and assessment documents they encounter; the kinds of learning opportunities they attend, and how, if at all, these reforms and opportunities influence what teachers think about and do in their classrooms. Looking across the interviews, I saw patterns which help explain the teachers' responses in a social context and the nature of their learning in an array of social settings. The three preliminary patterns I synthesized from the data and report on in this paper relate to the nature and substance of the tests, the professional

development opportunities available to teachers, and the rationales for and the consequences of the state exams.

On Tests and Teaching

Standardized tests matter. The professional literature is replete with debates about tests as a means of accountability, as measures of performance, and as levers of change (Corbett & Wilson, 1991; Editors, 1994; Feltovich, Spiro, & Coulson, 1993; Finn, 1995; Fuhrman, Clune, & Elmore, 1988; Koretz, 1988; Ravitch, 1995; Resnick & Resnick, 1985). These concerns become elevated when situations like CTB/McGraw-Hill's mis-scoring of almost 9000 New York City students' tests occur. In all of the talk about tests, however, one area gets scant regard: What teachers learn from tests, and if and how that knowledge affects their instructional practice. Common sense holds that tests drive classroom instruction. Evidence for that opinion is thin, however. Much research focuses on the relationship between students and tests (see, for example, Natriello & Pallas, 1998; Stiggins & Conklin, 1992; Wolf, 1998), but relatively few empirical studies explore the relationship between teachers and the tests they administer (Corbett & Wilson, 1991; Firestone, Mayrowetz, & Fairman, 1998; Grant, in press; Noble & Smith, 1994; Smith, 1991). The research that is available presents a mixed picture at best.

Those advocates of tests as a vehicle for driving educational change tend to cite general positive effects rather than specifics. Some (Feltovich et al., 1993; Popham, 1998; Shanker, 1995) simply argue that good tests will inevitably drive good instruction. Lacking any more specificity, Popham, Cruse, Rankin, Sandifer, and Williams (1985) claim that tests measure important learning, and that good tests results equal good education. Systemic reformers (Fuhrman, 1993; Smith & O'Day, 1991) advocate for testing as part of an overall strategy aimed at fundamental school change. Others (English, 1980; Glatthorn, 1987; Heubert & Hauser, 1999) argue that because standardized tests are a reality in most school districts, they should be used as a fundamental part of curriculum planning.

Critics of standardized testing are more direct in their assessment of the impact of testing on teaching. Madaus (1988) claims, among other things, that teachers will teach to the test, that they will adjust their instruction to follow the form of the questions asked (e.g., multiple-choice, essay), and that tests transfer control over the curriculum to whomever controls the test (Note 3). Claims by LeMahieu (1984) and Koretz (1995) are more tentative, but they too conclude that teachers may tailor their curricula to the content covered on the test. Recent empirical work supports some of these claims. Smith (1991) argues that many teachers respond overtly to test pressures and she offers a typology of eight orientations toward test preparation: ordinary curriculum with no special preparation, teaching test-taking skills, exhortation, teaching content known to be covered by the test, teaching to the test in format and content, stress inoculation, practicing test or parallel test items, and cheating. Firestone, Mayrowetz, and Fairman (1998) assert that testing programs in Maine and Maryland seem to influence teachers' content decisions, although they conclude that such influences are weaker than expected. Corbett and Wilson (1991) argue that testing, especially minimum-competency testing, has a pernicious effect on teachers in that it causes them to narrow their sense of educational purposes and to focus on activities designed to raise test scores whether or not they think those activities are good for students. They conclude that squeezing teachers in this fashion encourages them to rebel against reform measures good and bad. "Statewide testing programs do control activity at the local level, but the subsequent activity is not reform" (p. 1)

Other researchers are less sure that a direct relationship exists between standardized testing and teachers' classroom practices. Freeman, Kuhs, Porter, Knappen, Floden, Schmidt, & Schwille (1980), Kellaghan, Madaus, and Airasian (1982), and Salmon-Cox (1981) found little direct impact of standardized testing on teachers' daily instruction. Firestone, Mayrowetz, and Fairman (1998) claim that, while tests may have influenced teachers' decisions about what to teach, there was virtually no influence on their decisions about how to teach. In a cross- case comparison of two high school teachers' civil rights units (Grant, in press), I found little direct influence of testing on either teacher's content or pedagogical decision-making

This brief review suggests two points. First, we need to know more about the

relationship between teachers and tests. While the impact of tests on students has been much explored, research that inquires into if and how teachers are influenced by standardized tests is lacking. Second, that research around teachers and tests fails to show a clear or consistent pattern of influence. Tests matter, but how and to what extent is unclear.

State-Level Curriculum and Assessment in New York State

State-level influence over curriculum and assessment is a well-established tradition in New York State. The Regents test has been administered continually for over 100 years. These tests are administered in all academic subjects and are tied to school courses. For example, in social studies, students take the Global Studies test at the end of a two-year Global Studies course sequence in ninth and tenth grades; eleventh graders take the U. S. History and Government test after completing a course of the same name. Elementary and middle school teachers also follow a state curriculum in all school subjects and students take state-developed tests.

Recent State-Level Curriculum Changes

As is the case in most states, educational reform has been steady work since the 1980s. Begun during the tenure of former Commissioner of Education, Thomas Sobol, state-level focus on and activity around school curriculum hit full stride in the mid-1990s under current Commissioner Richard Mills.

Since 1994, working groups of state policymakers, teachers, and administrators have produced new curriculum and learning standards and scope and sequences for all school subjects. Social studies teachers, for example, may now consult the *Learning Standards for Social Studies* (New York State Education Department, 1996) and the *Resource Guide for Social Studies* (New York State Education Department, 1998). Compared with the previous round of curricular revisions in the mid-to-late 1980s, the changes represented in these documents vary from virtually no changes in the K-5 grades curricula, which follow an expanding horizons model, in the seventh and eighth grade U.S. and New York State history, or in the twelfth grade Participation in Government and Economics courses. Modest changes are evident in other curricula, such as the emphasis on geography in the eleventh grade U.S. history and government course. Major changes seem localized at sixth grade, where the course of study expanded from Western and Eastern Europe and the Middle East to the entire Eastern hemisphere, and at ninth and tenth grades, where the emphasis has changed from a cultural approach as represented in Global Studies to a chronological study as expressed as Global History and Geography.

Recent State-Level Assessment Changes

The state-level testing program is also changing. Although the scope of the changes varies (Note 4) , the net effect appears to be a general ratcheting up of the stakes for both teacher and students.

State tests of language arts, mathematics, and science have undergone radical transformations which include reducing the number of multiple-choice items and increasing the number and range of performance tasks. For example, new science tests call for students to actually perform experiments. By contrast, the social studies assessments will apparently change little: Multiple-choice questions will still dominate the tests, accounting for 55% of a student's score (Note 5) The major change seems to be in the writing portion of the exam. Unlike many minimum competency tests, New York students have always had to answer essay questions on state exams. The new tests are different primarily in the fact that a) students will no longer have a range of essay prompts to choose from, and b) a new kind of essay question, a document-based question (DBQ), is being introduced on each of the fifth, eighth, tenth, and eleventh grade tests. A DBQ asks students to write an essay synthesizing a number of primary source documents (e.g , short quotes from government documents and famous individuals, political cartoons, poems, charts and graphs) (Note 6) Plans call for students to answer a main idea-type question about each of the documents before writing their essay. High school students will also write a second, "thematic" essay

based on a single prompt (Note 7). The inclusion of the DBQ is the primary change in the structure of the social studies exams. One might argue that such a question represents a major shift away from traditional testing, but given the scope of the test (and the fact that students can easily pass the test without a single DBQ point), adding a DBQ could be read as a minor revision, or an instance of what Tyack and Cuban (1995) call "tinkering toward utopia."

Three other changes seem more dramatic. One is that the new fifth and eighth grade tests will produce individual student scores. Tests at those levels, termed "Program Evaluation Tests," have aimed at helping teachers understand the effectiveness of their content and pedagogical decisions (Note 8). The shift of emphasis to individual students is apparently intended to raise the stakes of these tests and tie them more directly to the high school Regents exams. The function of the Regents test is also being fundamentally changed. In the past, passing Regents tests in all academic subjects meant that a student earned a Regents diploma. Students could opt to take the less rigorous Regents Competency Exam (RCT) and earn a local diploma. Ninth graders beginning in 2001 will no longer have these options. The RCT will no longer be administered, and all students will have to pass five Regents examinations (English, mathematics, global history, U.S. history, and science) in order to graduate.

Given these changes, state-level tests are no less high-stakes for teachers than they are for students. Since the mid-1990s, state policymakers have introduced a number of curriculum reforms, such as new state standards for social studies, yet it is a concern about the state tests which surfaces most regularly in teachers' talk (Grant, 1997a). This makes sense for two reasons. First, the curriculum documents produced thus far offer teachers little assistance in making concrete instructional decisions (Grant, 1997b). Second, the messages teachers receive often promote the view that tests are intended to drive change (Grant, 1996). For example, during sessions devoted to new state social studies standards, one representative from the New York State Education Department (NYSED) said that new tests will "help grow change in the system." During another session, a different SED representative said, "New assessments will represent a change in instruction....Kids won't perform well until (teachers') instruction reflects this." And at yet a third meeting, NYSED Commissioner Richard Mills added, "Instruction won't change until the tests change." The message that tests matter was echoed during local school and district meetings. A suburban district social studies supervisor, for example, told teachers that "change in content will come if we change the tests." An urban district supervisor observed, "If we change the assessments, we'll change instruction" (p. 271). One might question the focus of test influence--instruction, curriculum, or the "system" in general--but it is hard to miss the larger point: tests matter.

The Prospects and Problems of State-Level Testing In New York State

The tendency of advocates and critics to cast standardized testing in black and white images is not supported here. My analysis suggests that teachers see the new NYS tests as a mixed bag. The prospects of tests which more closely mirror and support thoughtful instruction and closer collaboration with colleagues are mitigated by the problems of, among other things, uncertainty about the rationale for and consequences of the new tests and the unevenness of the opportunities to learn about and respond to changes in the tests. In short, teachers across grade levels and subject matters express an uneasy combination of hope and fear, anticipation and dread. I explore those poles by looking at teachers' perceptions of the new tests in terms of their nature and substance, the professional development opportunities available, and the rationales and consequences.

The Nature and Substance of the New NYS Tests

The NYSED is phasing in the new state tests over a period of four years, beginning with the English language arts tests at grade 4 in January, 1999. Consequently most of the teachers interviewed have not seen final versions of the tests they will administer. All have, however, received preliminary materials from state, district, and professional organization sources and so most assume that they have a fair

sense of what the new exams will be like. Most believe the tests will be an improvement over past assessments, but questions about the nature and substance arise.

Both elementary and secondary teachers expressed at least modest support for the general direction taken in the new tests. A middle school science teacher suggested simply that the NYSED was "changing what assessment means." An elementary school teacher was more specific. "I think there was a lot of change going on and then they changed the assessment," she said, "I remember giving that CTBS (a basic skills test) and teaching a literature-based program, and we were all complaining that it wasn't reflective [of our teaching]." Another elementary school teacher was more specific: "The new assessments test the same way we teach reading, and where we want kids to be in math."

Social studies teachers approved of the move to include primary sources within the DBQ. A high school teacher cited the real world relevance of questions which employ political cartoons. "You give them a cartoon and you say, 'Interpret this cartoon,'" she said. "That's interpretation, you know? If you open a paper and you look at a picture in the newspaper and you go, 'What's that mean?' That's something you would do in real life." A middle school teacher noted she now uses DBQ kinds of questions as a regular part of her instruction:

> I was working on a social studies test today for grade seven where they have to look at a document and think about some stuff like, what was the theme about the Revolutionary war, and they've got to write notes based on the picture. And it looks-the test is a lesson. It's a lesson in analyzing documents and taking notes from the document so you're not looking to see if they're right or wrong. You're looking to see can they look and think about what's on there.

This teacher and most others praised state efforts to bring standardized assessments into closer alignment with the kind of ambitious instruction they believe is important, such as analyzing primary sources and understanding that such texts can be interpreted in multiple ways. Social studies teachers worry about the continued strong emphasis on multiple-choice questions, but in questions like the DBQ, they see potential for pushing their students toward richer understandings.

But not all teachers held this view. Some focused on the continuing heavy presence of generally low-level multiple-choice questions, arguing that the test has changed little overall. As one middle school teacher explained:

> From my perspective, the social studies assessment doesn't seem like it's a change at all. Seems like it's kind of repackaged, kind of dressed up a little differently, but not really different and to me, there is something broken in [teachers' instruction] and we need to fix it. This new assessment to me isn't fixing it.

One might argue about whether teachers' practices are "broken," but the sentiment that some state tests, like social studies, seem less changed than others emerged throughout the focus group sessions. The English language arts and science tests, in particular, were cited as moving away from a heavy reliance on objective-style questions and toward questions with more real world and practical applications. For example, the English language arts tests asks students to write a range of pieces including technical, literary, and literary analysis essays. The science tests include performance tasks which ask students, for example, to set up a lab experiment. Teachers in these areas had questions about the nature of their respective exams, but there was a general sense that these exams push in more ambitious directions than the social studies tests do.

Social studies teachers see the prospective new state assessments as a mix of old and new. While most applaud the presence of primary sources and questions like the DBQ that ask students to analyze and synthesize information, they wonder if that emphasis won't be undercut by the continuing heavy weight of the multiple-choice section and questions which teachers generally perceive of as asking for low-level knowledge

Opportunities to Learn About the New State Tests

New state tests, like many other educational policies, can be viewed as an occasion to learn about the craft of teaching (Cohen & Barnes, 1993; Grant, in press). The focus group teachers nodded in agreement when participants raised questions such as, "Do I have the skills that I need?" and made assertions such as, "We have not been taught the way we're being asked to teach.... And I think that's really difficult without a lot of staff development to get people to think differently and to teach differently."

If the need for professional development was widely expressed, the teachers' experiences suggested that they may not be getting all that they want. Studies of professional development activities suggest that what session leaders think they are "teaching" and what participating teachers think they are "learning" during professional development activities can vary dramatically (Darling-Hammond & McLaughlin, 1996; Grant, 1997a; Smylie, 1995). Consequently understanding what kinds of professional development opportunities teachers had available to them and what sense they made of those opportunities was a major element of the focus group interviews.

Three patterns emerged from analysis of the interview transcripts. One was that all teachers seemed to have had access to a wide range of professional development opportunities both around the new curriculum standards and around the new tests. A second pattern was that they found those opportunities of uncertain value. Teachers reported that the state, and occasionally district, activities often resulted in incomplete and mixed messages. The frustration many teachers expressed about the more formal professional development opportunities was mitigated, however, by their sense that working more directly with colleagues was a more profitable use of their time. The third pattern, reform by "rumor," began to emerge in the first year of interviews, but was full-blown by the second year. Despite the wide array of professional development opportunities, the teachers clearly felt that there was still much indecision about how tests would ultimately look, how they would be scored, and the like. In a context of increasing pressure to respond, but little solid information, several teachers reported the sense that rumors were driving much of their responses.

The professional development opportunities available. Asked to describe the professional development opportunities available to them, the teachers constructed a long and varied list. Some NYSED-led sessions occurred in several venues (e.g., stand-alone sessions, part of district-level in-services, sessions during professional organization conferences) and focused alternately on the new tests alone or on how the tests reflected the new state curriculum standards. Representatives from local Board of Cooperative Extension Services (BOCES) programs also led professional development activities as stand-alone and district sessions. Some district-level sessions featured state and BOCES representatives, but others utilized the talents of district personnel, while still others brought in local and national experts. School-level professional development opportunities were also varied in that some called all teachers together, while others asked teachers to meet in grade or department-level activities. The focus group teachers also mentioned state teachers' union sessions, college and university course work, professional literature, informal networks, and colleagues as additional sources of information on tests and testing.

The uncertain value of professional development. Of these many sources, teachers were most critical of the state-led sessions. Some felt that cuts in the NYSED have left the agency woefully understaffed. Most others, especially the high school teachers, were less generous. An English teacher said, "I'm not going to break a sweat trying to reformulate what I do when their people (NYSED) don't know what they're doing." A social studies teacher was more blunt: "Do they have a clue as to what's going on?"

District-level sessions received more mixed reviews. A high school mathematics teacher praised her district's efforts to develop professional development activities that would meet teachers' perceived needs:

> My district is real supportive. If I say to them we need an inservice on blah, they will say we'll do it. They're wonderful that way. It's very teacher driven. Our school district is wonderful as far as them involving teachers

and listening to the teachers and valuing what the teachers say.

This comment stood largely alone, however, as most other teachers suggested that district-led professional development was lacking in usefulness. A high school social studies teacher noted:

> We've had two district wide superintendent's conference days and we've talked about [the tests] and gone over some things, but not into the detail that needs to be done to get a good feel for the types of questions and changes. I think in our building many people would still be hard pressed to give an accurate reflection of what the assessment is all about.

A middle school science teacher attended a district-sponsored inservice led by a district teacher. She reported that while the session could have been valuable, she left frustrated because the teacher who led the session came from a magnet school where resources are plentiful, whereas she teaches in a resource-starved neighborhood school. Not all the blame for weak district-sponsored professional development was laid at the feet of the leaders, however. A secondary social studies teacher panned the district-level sessions she attended, but she assigned much of that responsibility to her colleagues:

> We went to the district-wide [in-services]. They (the in- service leaders) always tried to be very positive, but the overwhelming number of teachers who are so negative about this assessment always wins out. It basically becomes a complaining session and you really aren't focusing on what the whole meeting was about anyway.

The focus group teachers reported that school-, grade-, and/or department-level professional development activities were generally more useful than state or district efforts An elementary school teacher, for example, praised the work her grade-level colleagues were doing:

> We have grade-level meetings. They're very positive, you know, even though we all don't want to test, we all feel like we shouldn't have to do it. They're (her colleagues) always very positive, always very friendly approaching it. Every time we go to a grade level meeting, [the team leader] always is handing us stacks and stacks of information materials. Things that we might need or might be able to use to help the kids get ready, whether it's for the science or the math or the English [tests]. There's always something positive going on.

A high school mathematics teacher explained that not only has the amount of conversation increased in her department, but that it is becoming increasingly acceptable to say, "I don't know how to do this." She went on to describe how her colleagues, both veteran and novice, were creating a new ethic whereby the traditional norms of isolation and "doing your own thing" were fading.

Not all teachers are similarly situated, however, and more than any other group, the high school social studies teachers present described their departmental interactions as less than optimal. Several nodded in agreement when an untenured teacher portrayed her colleagues as being obsessed with talk about "how to beat the test, or change the test, or fight the state, or fix the state or...how is the administration wrong, how are we right." Potentially useful discussions of teaching, learning, and assessment, she explained, get lost in the mix.

If teachers found formal state, district, and school-level professional development of uncertain value, all reported instances where informal networks and relationships had proven valuable. A high school social studies teacher said that, while she appreciated some elements of her district staff development days, "it is a lot easier to bounce off the ideas with somebody. And I just wrote [a DBQ] a few weeks ago with a colleague. We have now the same planning period so that worked out." A high school teacher reported that she and her colleagues have met informally after school to consider assessment issues. "There were a handful of us that got together after school on a voluntary basis,"

she said, ".... It makes my life a lot easier when I talk to other English teachers." In addition to these unstructured activities, several elementary school and high school mathematics teachers described informal networks of educators who meet regularly to discuss a range of issues, including those related to testing. A mathematics teacher described the benefits she has appreciated from her involvement:

> We have each other (she laughs). We have a network through (a local state university).where there have to be what-about 70 teachers, maybe 100 maybe that-we have meetings four times a year, and so now I don't feel isolated anymore. I mean I can always call [a colleague in a neighboring district]. I have friends [in another district]. Friends just about anywhere. I know what's going on at what school and I can pool resources, and so that helps a lot.

The power of such informal relationships is apparent: These teachers sense that they are working with peers who hold similar goals and concerns, who are willing to share ideas and practices, and who offer a sense of belonging. Such relationships, then, have an immediacy and a specificity that seems missing from the more formal professional development opportunities teachers typically experience. That these teachers have sought out and participated in these relationships is admirable; that they have felt compelled to do so in order to meet their needs is ironic, however, given the seeming wealth of structured opportunities.

Reform by rumor. Having informal sources of information and support may help teachers navigate some of the challenges the new state tests posed, but they do little to help teachers with the problems of mixed messages and unanswered questions. In fact, the more sources of information teachers encounter, the greater the incidence of reform by rumor.

Common across teachers of all grade levels and subject matters was a frustration with incomplete and conflicting information about the new tests. An elementary school teacher noted, "If we just had more information and if we knew what was expected of us and how to do it, possibly.we could do what was expected of us." A high school mathematics teacher added:

> If they're (NYSED) going to give us information, they have to give it more structured backing. Not this haphazard changing the rules daily.... Our math department head has said [at an in-service led by an NYSED representative], "Tell us what you want. We will do it. We will change the way we teach.... But you can't keep changing the messages you're giving us."

To be sure, state leaders seem to recognize that they are sending multiple and, at times, confusing messages. A high school mathematics teacher reported the following experience during a state-sponsored in-service:

> When we go to state meetings, (the NYSED representative) who's in the math ed department always prefaces his remarks with, "What I'm going to tell you is true at May 13th at 4 whatever. It's true right now. When I go back to my office, it might not be true." And we get to go to a lot of state meetings and everything and find out what's going on. And we always find out the latest stuff, but then it changes.

As this quote suggests, teachers do not necessary blame the state education representatives, but they are frustrated with the uncertainty of the situation. A high school social studies teacher's experience summed up some of the anxiety mixed and multiple messages can induce:

> I don't know if this geography thing (i.e., that the state curriculum and test for tenth grade were changed from Global Studies to Global History and Geography) is true or not. But somebody in my department had been in the state conference the week before and said, "I didn't hear any of this." And

> then we started frantically calling-I think we called the (local state university) Social Studies department, and they were calling all over to find if this was true. And I think the final verdict was that, "yes (geography has been added), but geography the way we've always taught it, so don't be nervous. They (NYSED) are not asking to name which direction the Danube River flows or anything like that." But, I don't know. It's crazy.

This teacher went on to remark, "I see it as just lots of rumors. It's like every other day we're coming in, 'Did you hear they're cutting out the constructed response? Oh, now the new course is Global History and Geography?'"

A cynical interpretation of the above is that teachers are merely pawns in a game that is being transacted all around them. This view asserts that while changing teachers' practices is the target, teachers' ideas and voices are largely ignored as those above them-state and district-level actors-do the real work of policy change. Teachers, through their professional development opportunities, may listen in. But as listeners rather than as full participants, they hear only bits and pieces, and rumors rule the day.

A more generous interpretation has two elements. One is that reforming education is simply hard work, especially when done in midstream, or what a policy maker in another state termed, "rebuilding the airplane while you're flying it" (Lusi, 1997, p. 91). The second element is that, given the sheer number of teachers and the wide range of circumstances in which they work, policy makers face a daunting task in attempting to change pedagogical practices. Whether they should try to or not, the parameters of the NYSED operation are intimidating: thousands of teachers, in thousands of schools, in close to 700 districts, and an agency with little more than a handful of employees. Clearly, then, NYSED must rely on the efforts of proxies-BOCES educators, professional organizations, district and school-level leaders, college and university academics-who may or may not understand and/or support the state agenda. In such a situation, the potential grows for mixed and confusing messages, and for reform by rumor.

The Rationales for and the Consequences of the New NYS Tests

The notion of "reform by rumor" functioned as a proxy for a number of comments where focus group teachers talked about feeling left out of the conversation about changing state assessments. Teachers across grade levels and school subjects expressed frustration that, while they are the professionals on whom the tests will have the most impact, their voices are not well reflected in important discussions about the nature, import, and design of new state tests. As one teacher said, "I really fear that unless there's open communication...this whole thing would be just kind of a charade." Another added, "I just feel that I've been talked at."

These teachers remain uncertain about the rationales for and the consequences of the state assessments, but seek to question rather than condemn. Most said they have attended meetings designed to inform them about the tests, but none said they were satisfied: Their questions either went unaddressed or, if they were addressed, the information they received did not always jive with information circulated previously While numerous questions arose during the focus group interviews, two dominated: questions about the rationales for changing the assessments and questions about the intended and unintended consequences of the tests.

Questioning the rationales for the tests. Whether the NYSED hopes to induce changes in teachers' curriculum decisions, their instructional practices, or both has been unclear for some time (Grant, 1997a). The focus group teachers echoed this confusion. They also discussed their uncertainty about whether the state's intention was to change their behavior or the students'. As a middle school social studies teacher said, "Are they (NYSED) doing this to better students' education, or are they doing it so they can say, 'Look, we changed something.'"

On the question of whose behavior NYSED is targeting, teachers expressed considerable frustration. For instance, an elementary teacher asked, "Who is it assessing? Is it really assessing the students? Or is it assessing the teachers?" Another elementary teacher echoed this point: "What is the purpose of the state exams? Is it actually to assess the students or to push the teachers in a direction?" A secondary social

studies teacher spoke directly to the issue of whose life is changing the most as a result of the new state tests:

> I think it's ironic that the state came out with all of these decisions in order to improve student learning and to make students better students and.I feel like I am doing so much work this year. When I do essays, I try to fix things and give them lots of responses and they just-I feel like I'm doing more work than the kids sometimes.... The last couple weeks it's like "I'm not taking this test! I took this test!" This is you. Not me. But it seems like the teachers are on the chopping block. And it's just ironic that it's no longer the student anymore. And it's the kids who are taking the test. And it seems like the kids are almost less and less responsible....

The last part of the quote above suggests that the issue of whether teachers or students are targeted is important, in part, because teachers are unsure where the blame is going to come down should test scores not rise. Many suspect, however, that teachers will take the brunt of the criticism. A high school mathematics teacher said, "They're (local administrators) are going to be pointing their finger if your kids don't do well. They're going to be pointing their finger at those teachers and that's unfortunate because they're (the teachers) going to be a scapegoat because of it." A secondary English teacher talked about the unfairness of holding the teachers whose students are taking the tests entirely responsible for the outcomes:

> I think that whole culture needs to change because you are not the sole responsible party for that student's abilities.... If someone did a lousy job last year, then you're getting a group of students without the proper foundation. And is there going to be some kind of mechanism that will address that if you realize that the child did not get proper foundation?...There's no way I solely am responsible for that child's [test scores]. I've had students who are functioning very very low and you're asking me to...bring that child further along. Is that child going to pass that test? No. So you're going to come to me and say, "Well, only 55% of your students passed this test. You're lousy!" I'm going to say, "Well, what did you give me?"

This quote raises a number of thorny issues, not the least of which is a seeming deficit view of children. This view implies that students come to a teacher with a set of deficiencies, resulting from poor parenting, poor schooling, and the like, which the teacher must then "correct." The problems with this view are several, but in this case, they serve to amplify the dilemma this teacher faces: She feels the twin burdens of preparing students to take the exam and of being held accountable for their performance. Although it seems unfair to make the child the pawn, this teacher rightly points out that she alone can not be responsible for test scores.

Teacher frustration was also apparent around the question of whether NYSED's intent was to change curriculum, instruction, or both. The focus group teachers assumed the tests were meant to induce changes, but they were unsure what sort of change was expected.

A secondary social studies teacher saw the state's aim as primarily directed toward curriculum.

> But it looks like -- the more I hear about it it's as if the state through its tests is controlling what gets taught in the classroom. By saying that the test is going to be done this way, all of a sudden it's going in and saying well you can't teach this, this, and this when you want to. You have to teach this. You have to teach this.

An elementary teacher, by contrast, suspected that the state's intention is to influence teachers' instructional practices:

> Is this a way of making teachers look at their practice and alter their

> teaching techniques because they see a certain topic being covered on an exam and so they'll say, "Oh, I didn't do that so well that time. I guess I have to spend more time on that next year." So if you see the focus on the exams, then you've got to go back and make sure that you include that type of instruction the next year. And so I think-are the tests pushing-is the state using the test to push teachers in a certain direction with their instruction?

While most of the focus groups sensed that the state tests were being used to leverage change of one sort or another, not all did. A high school English teacher reported that she had been told, "We've been doing this all along. That this is no big deal...all we have to do is get kids accustomed to the format [of the test]." A secondary science teacher added to this notion, by reciting a familiar teacher expression, that is, "this too shall pass." "In our science department," he said, "they feel because science is the last assessment [to be introduced] that this is all going to blow over." The notion that whatever NYSED introduces is likely to fade in importance over time was not the dominant view among the focus group teachers. But its expression should warn state-level reformers that whatever leverage they believe tests hold for changing instruction and/or curriculum may be illusory. This is not because teachers do not sense that problems exist: None of the focus group teachers was willing to suggest that all is right with public education. But several supported the following sentiments of an elementary school teacher who questioned the reliance on tests as a lever of real instructional change:

> I understand that certainly there are places in American education that are in dire need of shaping up somehow....It (the test) just seems to me a misdirection of resources. We're spending how much--thousands of dollars on training, on writing these tests or whatever they're doing to when the real issue is what's happening in the classroom. What kind of preparation are teachers getting? What kind of preparation are they getting before they even get a classroom? What kind of thinking is going on here? And are those questions even being asked? Or were they ever asked before this happened? It was just suddenly that we had this massive assessment. And I don't remember any sort of input from teachers. I don't remember any state education people coming to us and saying, "What do you think?" Or, "What's going on in your classroom?" It was just this kind of mandated attempt to reform. And maybe it will work. I mean, I don't know whether it will work or not. But it seems to me there's so much more that could be done that hasn't been attempted in terms of helping teachers.

To be fair, NYSED officials and the state Board of Regents have proposed a range of reforms that push changes in curriculum and in teacher education. The primacy of the state testing program, however, weighs heavily. The focus group teachers are not opposed to improving teaching and learning, but they are uncertain about the rationale for standardized tests as a vehicle.

Predicting the consequences of the new tests. The idea that the new tests may yield no real consequences for teachers' practices was one of several predictions the focus group teachers made. Most of those predicted consequences were negative, but not all. For example, several teachers in the first year focus groups expressed the hope that the tests would mean greater collaboration with their colleagues. A high school English teacher summed up the feeling: "If there were more opportunities to get more people together, that would help." While it was far from unanimous, a number of the year two teachers reported that, in fact, they had found their peers receptive to and interested in working together.

The overwhelming sentiment, however, was that the new tests could produce undesirable effects. Those effects grouped loosely around issues of pedagogy, students, and teachers.

Two related consequences of tests for pedagogy arose. One is that, rather than promote more ambitious teaching and learning, the state tests may actually push more reductive forms of teaching and learning. The most common expression was that teachers felt increased pressure to tailor one's teaching to the test parameters. As a

secondary social studies teacher noted, "You've got people in high places just saying 'teach to the test.'" A middle school English teacher complained that he felt pressure to "teach them (students) test terminology when I could be teaching them other things." This teacher went on to describe the kind of support his district provides as little more than practice exercises. "The only thing I've gotten from my district," he said, "is lots of practices. Every week there's, 'Thank so and so for giving this practice material. Here's another listening practice that you may want to use.' I could have spent my whole year doing practices."

The sense that teachers feel pressed to adopt direct teaching approaches as a means of bolstering short-term test performance was in direct competition with the sentiments expressed earlier that the new state tests could be viewed as supportive of more ambitious instruction. During the interviews, however, no teacher commented on this seeming contradiction. One explanation is that they were simply unaware of its emergence. A more interesting possibility is that these teachers can read multiple messages in the tests. Take social studies as an example. Teachers thinking about the multiple-choice questions could reasonably assume that a more traditional, direct instruction approach was being encouraged. If those same teachers were thinking instead about the DBQ questions, it seems equally reasonable to assume that richer forms of pedagogy were intended. This ambivalence, which has surfaced in a number of places already, underscores the difficulty in understanding teachers' perceptions of state tests and it suggests that their classroom responses may be more complex and textured than reformers may want or expect.

A second potentially negative consequence of the new tests was an increased emphasis on remediation as a way to deal with low test scores. The teachers, especially those in the second year interviews, described a wide array of remedial approaches taken in their schools. Those approaches included additional classes designed for students presumably at risk of failing, summer and Saturday test review courses, hiring additional teachers and aides to staff learning labs where students could either come voluntarily or by teacher assignment, and reassigning teachers to classes of students based on their perceived ability to help those students pass the exam.

The teachers offering these examples generally seemed supportive of them. The seeming contradiction that ratcheting up remedial efforts would occur at the same time teachers were being pushed to change their pedagogy went unremarked upon. Again, however, this contradiction may be less apparent than one might suspect. Empirical evidence is surprisingly thin on the question of which instructional approaches lead directly to high test scores (Cohen & Barnes, 1993; Grant, in press). Consequently, a reasonable response to a new testing situation might be both to make changes in "regular" classes and to begin planning for remedial instruction at the same time.

The real danger, however, is that these remedial opportunities will become little more than drill sessions, a point that was recognized by several teachers. For example, a high school mathematics teacher observed:

> If the students do not pass, they're going to be remedied with questions that will make them pass. So eventually every student will pass. Doesn't matter the categories, they're going to do component retesting, so if the student doesn't do well in these three areas, they'll be grilled in those three areas with a bank of questions, and then the student will have another test from the bank that he was drilled in. So eventually they'll get it.

Such an approach may work for low-level skills, but is of dubious use in areas like social studies where conceptual knowledge is central. As VanSledright & Brophy (1992) observed, "naive but imaginative accounts persisted in some children even after direct instruction designed to change them" (p. 854). Without any definitive research supporting one means of improving test performance over another, drill and practice remediation is as likely to flourish as any other approach.

A second area of negative consequences anticipated by the focus group teachers concerned students. An elementary teacher worried generally that the net effect of a high profile, high- stakes testing program would be a "nation of test-takers":

> Something that I've been thinking about more is the effect this has on the

> children, on the student. What kind of learners is this going to shape? Are we producing a nation of test-takers, and if so, are those test-taking techniques or skills what we need to produce life long learners that we talked about before?

Other teachers expressed more focused concern about the anticipated consequences for urban students. Wiles (1996) argues that test performance is clearly distributed along socio- economic lines with upscale, white suburban children consistently outscoring their urban and minority peers. The focus group teachers, both urban- and suburban-based, recognized the inherent threat that high-stakes testing poses for some children. An elementary school teacher said, "I'm very concerned about some of the larger populations in the bigger urban areas. I don't understand how this is going to positively affect these kids." A high school teacher, commenting on the anticipated testing of special education students, asked, "How do we accommodate the non-standard kids on a standardized test?"

No teachers thought their students' scores on the new tests would improve immediately over past test scores. A couple of teachers did express, however, the hope that their students' scores would increase over time. A middle school English teacher said, "I think, naive though it may be, that our kids are going to do better ultimately on these exams. Maybe not this year, but ultimately."

This hopefulness stood in stark contrast with the prevailing view that teachers anticipated problems for their students. Underlying both these sentiments is a harsh truth: These teachers simply do not know how their students will perform on the new tests. Given the general tendency for a correlation between test scores and students' social capital, it is difficult to understand why suburban teachers would be worried. And yet, analysis of the relative concern expressed by suburban vs. urban teachers suggested that suburban teachers and administrators may be even more concerned about potentially low scores than their urban peers. One proxy for this finding is the observation that the overwhelming number of remedial efforts planned are being developed in suburban schools.

As noted above, no teacher feels s/he has an inside track on what approaches will insure high scores. Left to follow one's hunches, it is no particular surprise to find concern among all teachers, both suburban and urban. But what explains the fact that suburban teachers seem to be more concerned about their students' performance than their urban peers? Part of an explanation must consider the notion that not all suburban districts are created equal. The suburban teachers in focus group teachers represented first-, second-, and third-ring suburbs. First-ring suburbs tend to include a range of working to middle class students. Second-ring suburbs are more upscale; most students come from middle to upper-middle class homes. Finally, the third-ring suburbs are rural areas that recently have attracted a large number of middle and high SES families. With the exception of one or two urban magnet schools, it is the schools in the second- and third-ring suburbs that consistently rank in the top quartile according to a highly publicized local business magazine. Top quartile spots on this list have real consequences for real estate values, bragging rights, and the like, and so the scramble to move up can be intense. New tests, then, represent a potential threat to schools' past standings. School people in high performing schools want to maintain their positions; educators in middle and low performing schools hope to at least avoid dropping further

The competition for high test scores plays out as a third set of consequences. Here, the focus is on the pressure and uncertainty teachers feel as they decide if and how to modify their teaching based on their perceptions of the state test. A couple of these pressures have already been described. One is the feeling of uncertainty teachers have about which approaches will ensure higher scores. A second pressure surfaces as teachers report being made to feel entirely responsible for their students' results. Putting the point on this feeling is a secondary social studies teacher:

> Just this week I was called down to the office and we were comparing some of the *Business First* statistics that were out just recently....So according to our administration [if we get low test scores]...people come out to vote and decide they don't want to vote on the budget, therefore the whole

> community goes down. So, I left the office thinking the weight of this town...is on my shoulders. Whether or not, you know, my kids pass. And we had like a 70% last year and we're expected to have at least a 90 if not higher. So, in terms of administration, testing is a pretty big deal.

Not all principals apply pressure so directly, but many apparently do. This is more likely to happen in high schools than elementary schools, however. According to several of the focus group elementary school teachers, their principals are more likely to talk about test scores as part of a bigger picture of how students are progressing. These teachers do not necessarily feel any less pressure than their high school peers, but one source of pressure, the school administrator, seems to be less of a factor.

The new elementary school exams are more high-stakes than they used to be; recall that now individual student scores will be reported rather than group scores. The stakes are even higher in the high schools, however, as passing the Regents exams will be necessary in order to graduate. Consequently, it is not hard to understand why high school administrators might be more likely than their elementary peers to put pressure on their teachers. Whether that tactic will pay off ultimately or not is hard to predict. But one manifestation of that pressure is to cause teachers to consider issues that they probably have not had to think about in the past. One particularly compelling story came from a high school social studies teacher who said she now wonders about each new student who comes into her classes:

> I never--it never crossed my mind before that a certain kid was going to lower my passing rate or not, and I actually started thinking about that this year. And I was so ashamed of myself about that. And one of the girls I had transferred from a general track. She stayed in my class. I didn't want to just dump her. But she can now take the RCT at the end of the year. But I had a girl a couple years ago who transferred from another state. She never had Global 9. And I was just happy to work with her and she was going to try it. And if you go to look at an individual kid and say they're not going to do it, it's horrible to think that--to individualize it like that. Because I guess every couple kids knocks you down a little bit. And our--I know that our department chairs had our results individualized and our principal keeps coming into meetings saying, "How can we raise this up? How can we do this better?"

This teacher concluded her story with a nervous laugh, saying, "But I'm glad I have tenure, right?" Yet, having tenure seems little consolation for this thoughtful and dedicated teacher now confronted with the dilemma of wanting to work with all students, but recognizing that doing so may cause her teaching to be called into question should her students' scores not measure up.

Not all the consequences described were negative, however. Several teachers cited greater collaboration with their peers as a key benefit of the new tests. Elementary teachers and high school mathematics and English teachers were most vocal on this point. "I think we have so much to learn from each other," one elementary teacher said. Another echoed this point, commenting, "We're really trying to deal with this [new tests] and trying to work as a faculty to help each other." A high school English teacher noted that information is vital and that colleagues are an important source. "What's most important to me is being able to communicate with other people so I can get some information." A high school mathematics teacher concurred, but pointed out that that the new exams were forcing teachers to rely on each other:

> I think the nature of the testing--it certainly sets the situation up for teachers to talk. Because the types of questions that happen to be asked. They don't have the stockpile of old Regents questions. So [teachers say] "I came up with this. You know, I'm going to use this." We can share, and the nature of the beast is forcing the issue.

Social studies teachers reported some positive collaborations with peers, but they also cited more instances than the other teachers of situations where friction had developed.

A high school teacher described the tension that arose over course assignments:

> We have attempted to get together and work, but what we have found out has been happening is just been a lot of back- stabbing and a lot of animosity because there are a couple of teachers who just adamantly refuse to teach 10th grade (when the Global exam is administered). So the feeling is, well, they can do the ninth grade program. But where is their accountability? Because they just will not do that 10th grade when their kids take the Regents at the end of the year.

This teacher's experience points, again, to the variability in the way consequences of the test are playing out. This variation is explained, in part, by the development of as many unintended as intended consequences. State-level reformers may have hoped, for example, that teachers would see the test as an impetus for more ambitious instruction, closer collaboration, and the like. And this seems to be occurring. But reformers probably did not predict the more negative consequences these teachers are seeing. That these outcomes are unintended is little solace, for they may be just as real to the teachers as the intended outcomes. Actually, these unintended consequences may ultimately be more important because they seem to receive scant attention from state and district-level actors. State and district leaders may be unaware of these issues, they may be ignoring them, or they may not see them as problems. In any event, it seems interesting that no teacher mentioned that s/he had participated in any explicit conversations about the problems they anticipated. As noted above, teachers did see positive possibilities arising from the new state tests and there was no particular sense of gloom during the interviews. How teachers will manage the more negative consequences is unclear, but the supposition that they will have no effect seems naive.

Implications

Substantive change is always unsetthng. So reform on the scale that New York state is attempting, in all grades and in all school subjects, is bound to generate some frustration, anxiety, and uncertainty. The findings above tell us that while teachers are not adverse to change, they have real concerns about the nature of the changes proposed, the professional development opportunities available to learn about these changes, and the rationales for and consequences of the new state tests.

Given the complexities of teaching and policy (Grant, 1998), it is not surprising to learn that teachers see both prospects and problems in the new NYS tests. State-level policymakers in New York, like most of their peers, are attempting reform on a massive level (Lusi, 1997) and are doing so with relatively few levers for change. What this study suggests is that teachers are not passive participants and must not be designed around. The dream of teacher-proof curriculum as a means of changing teachers' practices has proven to be a myth (see, for example, Dow, 1991; Schwille, Porter, Belli, Floden, Freeman, & Knappen, 1983). Faith in tests as a means of corralling teachers' practices may ultimately prove just as chimerical as long as teachers are left out of the loop. If any of the changes state reformers propose are to stick, then these teachers are saying they need to be more actively involved in the formulation of those changes. But there is something else. These findings also suggest that there are real and important differences in the ways teachers perceive reforms across grade levels. Among other things, this means that reformers can not take a one-size- fits-all stance and that professional development needs to be sensitive to the differences in the perceived needs of teachers.

Notes

The author wishes to acknowledge Bob Stevenson's thoughtful comments on an earlier draft of this article.

1. The TLA study is funded by the Collaborative Research Network, sponsored by the Graduate School of Education at SUNY-Buffalo. The faculty and students who worked on this study include Suzanne Miller, Robert Stevenson, Mark

Templin, Meg Callahan, Diana Lawrence-Brown, and Gina Trzyna.

2. The small number of elementary school teachers was due partly to design and partly to exigencies that prevented the other invitees from attending on that date
3. Corbett and Wilson (1991) point out, however, that Madaus's claims are based on limited data: "anecdotes, testimony from public hearings, historical accounts, and an occasional international study" (p. 26).
4. Revisions of state tests is still in progress so some of what follows is based on SED reports of changes they expect will occur.
5. The first administrations of new social studies tests will begin in the fall of 2000.
6. For example, in the test sampler for the Global History and Geography exam (New York State Education Department, 1999), students would be given documents that range from a poem by Lao Tzu; portions from Pericles' "Funeral Oration," the English Bill of Rights, the Japanese Constitution, a speech by Benito Mussolini; and a political cartoon about the monarchy in France during the 1600-1700s. They are then directed to write an essay in which they "compare and contrast the different viewpoints societies have held about the process of governmental decision making and about the role of citizens in the political decision-making process" and to "discuss the advantages and disadvantages of a political system that is under the absolute control of a single individual or a few individuals, or a political system that is a democracy" (p. 25).

   A test sampler in NYS consists of a description the types of test items, sample questions, a breakdown of the number of questions by curriculum standard and topic, rubrics for essay questions, and sample student responses.

   At present, the only test sampler available is that for tenth grade Global History and Geography. The first administration of that test is scheduled for June 2000. Test samplers for the grades 5 and 8 tests are to be available this fall with administration of the grade 5 test scheduled in November 200 and the grade 8 test in June 2001. The test sampler for the grade 11 test is due out in spring 2000 and the new test is scheduled for June 2001.
7. From the Global History test sampler (New York State Education Department, 1999), students are given this theme on belief systems: "At various times in global history, members of different religions have acted to bring people together. Members of these same religions have also acted to divide people and have caused conflict." Students are then directed to this task: "Choose two religions from your study of global history and geography. For each religion: Describe two basic beliefs of the religion; Explain how members of the religion, at a specific time and place, acted either to unify society or to cause conflict in society" (p. 29).
8. The PET tests were given at grades 6 and 8. The new tests will be administered at grades 5 and 8.

References

Bogdan, R., & Biklen, S. (1982). *Qualitative research for education: An introduction to theory and methods*. Boston: Allyn and Bacon.

Cohen, D., & Barnes, C. (1993). Pedagogy and policy. In D. Cohen, M. McLaughlin, & J. Talbert (Eds.), *Teaching for understanding: Challenges for policy and practice* (pp. 207-239). San Francisco: Jossey-Bass.

Comfort, K. (1991). A national standing ovation for the new performance testing. In G. Kulm & S. Malcolm (Eds.), *Science assessment in the service of reform* (pp. 149-161). Washington, DC: American Association for the Advancement of Science.

Corbett, H. D., & Wilson, B. (1991). *Testing, reform, and rebellion*. Norwood, NJ: Ablex.

Darling-Hammond, L., & McLaughlin, M. (1996). Policies and support professional development in an era of reform. In M. McLaughlin & I. Oberman (Eds.), *Teacher learning: New policies new practices* (pp. 202-219). New York: Teachers College Press.

Dow, P. (1991). *Schoolhouse politics: Lessons from the Sputnik era*. Cambridge, MA: Harvard University Press.

Editors. (1994). Introduction. Symposium: Equity and educational assessment. *Harvard Educational Review, 64*, 1- 3.

English, F. (1980). *Improving curriculum management in the schools*. Washington, DC: Council for Basic Education.

Feltovich, P., Spiro, R., & Coulson, R. (1993). Learning, teaching, and testing for complex conceptual understanding. In N. Frederikson, R. Mislevy, & I. Bejarq (Eds.), *Test theory for a new generation of tests* (pp. 181-217). Hillsdale, NJ: Lawrence Erlbaum.

Finn, C. (1995). Who's afraid of the big, bad test? In D. Ravitch (Ed.), *Debating the future of American education: Do we need national standards and assessments?* (pp. 120-144). Washington, DC: Brookings Institution.

Firestone, W., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*(2), 95-113.

Freeman, D., Kuhs, T., Porter, A., Knappen, L., Floden, R., Schmidt, W., & Schwille, J. (1980). *The fourth grade mathematics curriculum as inferred from textbooks and tests* . East Lansing, MI: Institute for Research on Teaching, Michigan State University.

Fuhrman, S. (1993). The politics of coherence. In S. Fuhrman (Ed.), *Designing coherent education policy: Improving the system* (pp. 1-34). San Francisco: Jossey Bass.

Fuhrman, S., Clune, W., & Elmore, R. (1988). Research on educational reforms: Lessons on the implementation of policy. *Teachers College Record, 90*, 237-258.

Glaser, B. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press.

Glatthorn, A. (1987). *Curriculum leadership*. Glenview, IL: Scott, Foresman.

Grant, S. G. (1996). Locating authority over content and pedagogy: Cross-current influences on teachers' thinking and practice. *Theory and Research in Social Education, 24*(3), 237-272.

Grant, S. G. (1997a). Opportunities lost: Teachers learning about the New York state social studies framework. *Theory and Research in Social Education, 25*(3), 259-287.

Grant, S. G. (1997b). A policy at odds with itself: The tension between constructivist and traditional views in the New York state social studies framework. *Journal of Curriculum and Supervision, 13*(1), 92-113.

Grant, S. G. (1998). *Reforming reading, writing, and mathematics: Teachers' responses and the prospects for systemic reform*. Mahwah, NJ: Lawrence Erlbaum.

Grant, S. G. (in press). An uncertain lever: The influence of state-level testing in New York on teaching social studies. *Teachers College Record*

Heubert, J., & Hauser, R. (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press

Kellaghan, T., Madaus, G., & Airasian, P. (1982). *The effects of standardized testing*.

Boston: Kluwer-Nijhoff.

Koretz, D. (1988). Arriving in Lake Woebegon: Are standardized tests exaggerating achievement and distorting instruction? *American Educator, 12*(2), 8-15, 46-52.

Koretz, D. (1995). Sometimes a cigar is only a cigar, and often a test is only a test. In D. Ravitch (Ed.), *Debating the future of American education: Do we need national standards and assessments?* (pp. 154-166). Washington, DC: Brookings Institution.

LeCompte, M., Preissle, J., & Tesch, R. (1993). *Ethnography and qualitative design in educational research.* (2nd ed.). New York: Academic Press.

LeMahieu, P. (1984). The effects on achievement and instructional content of a program of student monitoring through frequent testing. *Educational Evaluation and Policy Analysis, 6*(2), 175-187.

Lusi, S. F. (1997). *The role of state departments of education in complex school reform.* New York: Teachers College Press.

Madaus, G. (1988). The influence of testing on the curriculum. In L. Tanner (Ed.), *Critical issues in curriculum: 87th yearbook of the NSSE, Part 1* . Chicago: University of Chicago Press.

National Center for History in the Schools. (1994). *National standards for United States history.* Los Angeles: Author.

National Council for Social Studies. (1994). *Expectations of excellence.* Washington, DC: Author.

Natriello, G., & Pallas, A. (1998). The development and impact of high stakes testing. Available on-line: http://www.columbia.edu/~gjn6/histake.html

New York State Education Department. (1996). *Learning standards for social studies.* Albany, NY: Author.

New York State Education Department. (1998). *Social studies resource guide.* Albany, NY: Author.

New York State Education Department. (1999). *Global history and geography Regents examination: Test sampler draft.* Albany, NY: Author.

Noble, A., & Smith, M. L. (1994). Old and new beliefs about measurement-driven reform: "Build it and they will come." *Educational Policy, 8*(2), 111-136.

Popham, J., Cruse, K., Rankin, S., Sandifer, P., & Williams, P. (1985). Measurement-driven instruction. It's on the road. *Phi Delta Kappan, 66*(9), 628-634.

Popham, W. J. (1998). Farewell, curriculum: Confessions of an assessment convert. *Phi Delta Kappan, 79*, 380-384.

Ravitch, D. (Ed.). (1995). *Debating the future of American education: Do we need national standards and assessments?* Washington, DC: Brookings Institution.

Resnick, D., & Resnick, L. (1985). Standards, curriculum, and performance: A historical and comparative perspective. *Educational Researcher, 14*(4), 5-20

Salmon-Cox, L. (1981) Teachers and standardized achievement tests: What's really happening? *Phi Delta Kappan, 62*, 631-634.

Schwille, J., Porter, A., Belli, G., Floden, R., Freeman, D., & Knappen, L. (1983).

Teachers as policy brokers in the content of elementary school mathematics. In L. Shulman & G. Sykes (Eds.), *Handbook of teaching and policy* (pp. 370-391). New York: Longman.

Shanker, A. (19[illegible]5). The case for high stakes and real consequences. In D. Ravitch (Ed.), *Debating the future of American education: Do we need national standards and assessments?* (pp. 145-153). Washington, DC: Brookings Institution.

Smith, M., & O'Day, J. (1991). Systemic school reform. In S. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing* (pp. 233-267). New York: Falmer.

Smith, M. L. (1991). Meanings of test preparation. *American Educational Research Journal, 28*(3), 521-542.

Smith, M. L., Heinecke, W., & Noble, A. (1999). Assessment policy and political spectacle. *Teachers College Record, 101*(2), 157-191.

Smylie, M. (1995). Teacher learning in the workplace; Implications for school reform. In T. Guskey & M. Huberman (Eds.), *Professional development in education: New paradigms and practices* (pp. 92-113). New York: Teachers College Press.

Stiggins, R., & Conklin, N. (1992). *In teachers' hands: Investigating the practices of classroom assessment.* Albany, NY: SUNY Press.

Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia.* Cambridge: Harvard University Press.

Wiles, D. (1996). *Networking high performance in New York's secondary education: The Regents curriculum story.* Lanham, MD: University Press of America.

Wolf, R. (1998). National standards: Do we need them? *Educational Researcher, 27*. 22-5.

About the Author

S. G. Grant
517 Baldy Hall
State University of New York at Buffalo
Buffalo, New York 14260

Email: sggrant@acsu.buffalo.edu
Telephone: 716-645-6493

S. G. Grant is an assistant professor of Social Studies Education in the Department of Learning and Instruction. He has published papers in both social studies and general education journals. His most recent journal publications have been in *Theory and Research in Social Education* and the American Educational Research Journal. In the fall of 1998, he published his first book, *Reforming Reading, Writing, and Mathematics: Teacher's Responses and the Prospects for Systemic Change*. An article on the influence of state-level tests on teachers' classroom practices is forthcoming in *Teachers College Record*.

Appendix

FOCUS GROUP PROTOCOL

Spring, 1998

- Introduction: Why we are here. Guidelines and ground rules

- METAPHORS

  Moderators and participants introduce themselves to group.

  To get started, introduce yourself to someone next to you and describe an image or metaphor that characterizes your thinking and/or feelings about the new state assessments.

  After they have shared in pairs, have them share their metaphors with the group.

  Have participants discuss and elaborate on the metaphors. Lead a discussion of the metaphors. What do they say about our thinking? Common features? Significant differences.

  Direct the discussion toward the next question-what do these assessments mean to you?.

- MEANING OF ASSESSMENTS

  What do/will these assessments mean to you? Your school? Your students?

  Transition to next question-are you prepared to deal with these implications?

- BEING PREPARED

  How prepared to deal with these assessments do you feel? How are you being prepared? What are you being prepared for? What opportunities do you have to talk about the assessments and related issues?

  Build on these expressions to move toward a discussion of needs.

  What help do you need?

  This discussion should lead naturally to talk of challenges.

- CHALLENGES

  What challenges/concerns do you anticipate? How will you deal with these challenges/concerns? Who do you expect will help you?

- CLOSURE

  What has this conversation made you think about concerning teaching and testing (e.g., issue, question, new image) ?

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (480-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallos
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apemberton@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
aacosta@cucea.udg.mx

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Maria Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatistica
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Hugo Casanova Cardiel
Universidad Nacional Autónoma de México

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiernr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

BEST COPY AVAILABLE

This article has been retrieved 1215 times since February 28, 2000

prior vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 8 Number 15 February 28, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Forces for Change in Mathematics Education: The Case of TIMSS

Donald S. Macnab
Northern College
Aberdeen, Scotland

Abstract
The results of the Third International Study in Mathematics and Science Education (TIMSS) were published in 1996/7. Since that time the participating countries have reacted in a variety of ways to the comparative performance of their students. This article investigates the diverse effects these reactions have had on mathematics curricula and teaching methodologies in a selection of these countries, within the context of a wider analysis of the motivations which determine change in education.

Introduction

What causes schools' mathematics curricula and teaching methodologies to change over time? To what extent do they change in a rational response to external objective considerations; to what extent subjectively in accordance with beliefs and social pressures? What does success mean in relation to change? Often enough, the effect of change (planned or otherwise) is to metamorphose antecedent success criteria to validate the change, at least in the short term. In the world of politics this is a commonly recognised practice; in education, less so. Fullan (1993) documents many such instances in education from the 1960s onwards. Reviewing the last 30 years, he concluded that "we have been fighting an uphill battle.... We need a different formulation to get at the heart of the problem, a different hill, so to speak. We need, in short, a new mindset about educational change."(p 3). For an analysis in a Scottish context, see Macnab (1999a).

In Fullan's words, the essence of the difficulty is that "we have an educational system that is fundamentally conservative. The way that teachers are trained, the way

that schools are organised, the way the educational hierarchy operates, and the way that education is treated by political decision-makers results in a system that is more likely to retain the *status quo* than to change. When change is attempted under such circumstances it results in defensiveness, superficiality, or at best short-lived pockets of success." (Fullan, 1993, p. 3).

All those involved in promoting and implementing change do so from a sense of moral purpose to improve education. In a study of educational innovation in science mathematics and technology education in 13 countries (Black & Atkin, 1996), the authors conclude that "things are much more complicated than they seem.... Comparisons [between different countries] illustrate how the historical perspective and the cultural embedding—of educational thinking, of conceptions of change, and of the nature of the particular subjects involved—all have a profound effect on any process of change. [These comparisons] also illustrate the complexity of change. Fashionable opposites, such as top-down v. bottom-up, or teacher-active v. teacher-passive, are not helpful. In the real world action and change take place in more complex ways and at intermediate points along these bi-polar axes. There is another reason why change is complex. When it succeeds, it often does so for unforeseen causes. Those who think they control it sometimes find that unpredictable inner imperatives have passed control to others. Planned hierarchies of people collapse. Students may be better motivated but learn less. Teachers may be enthusiastic but students resistant, or vice-versa." (Black & Atkin, 1996, pp. 1-2).

Black and Atkin devote a chapter of their book to the question "What drives reform?" They comment that "every country that participated in our international study is dissatisfied with that education of its students in science, mathematics, or technology. Every country is trying to make changes.... Every country seems to be more or less unhappy with what it has today.... At any moment, however, each country will be preoccupied about different perceived ills.... Each country is fighting its own demons. But there is a paradox. All the most important pressures and influences that promote change in science, mathematics, and technology education in schools keep re-appearing as we move from one country to another. None appears only in a single country, and in that sense little is unique. Yet the countries are different and distinct, because each attributes a different weight to particular problems and to how they combine and interact. No country is ever exactly in phase with any other because each is a creature of its own unique history and evolution." (Black & Atkin, 1996, pp. 12-13).

In an earlier study, (Adams & Chen, 1981), the authors ask "Why then is the history of innovation such a doleful one? Why, according to the literature, is failure its companion so frequently? Why, given the burning enthusiasm of the advocates of reform, do teachers remain unimpressed, even glum, and administrators shudder?" (p. 1). In the final two paragraphs of their book they conclude a further set of questions commenting that, "the questions, it seems are endless.... [T]o finish the book on such a note of uncertainty is distressingly unimaginative." (p. 282). They do not, however, provide clear-cut answers to the questions with which they began.

The evidence from these studies and others is that the central imperative and dilemma underlying the change process in education is a sense of dissatisfaction with the status quo giving rise to the feeling that change is necessary, combined with confusion about its purpose, and uncertainty about the nature and value of its outcomes, with potential resulting disappointment and frustration for planners and teachers alike.

TIMSS and Change

The Third International Mathematics and Science Study (TIMSS), the largest international survey of attainment in mathematics and science ever attempted, took place in 1994/5 in over 40 countries, (Martin et al., 1996, 1997). Details of the underlying research questions and project design are contained in Robitaille, (1996a). For detailed technical reports see Martin and Kelly (1996, 1997). Two main groups of children were tested: Population 1, 8/9 years old, and Population 2, 13/14 years old. In addition, a third population, students in their "final year" of secondary school, was tested. A summary of the average scores of the various nations is presented in Table 1.

Table 1

TIMSS 1996/97 National Average Scores: Mathematics

	Pop. 1 (8/9 yrs)	Pop. 2 (13/14 yrs)	Pop. 3 "Final Year"
(AUSTRALIA)	546	530	522
(AUSTRIA)	559	539	518
BELGIUM-FLEMISH		565*	
(BELGIUM-FRENCH)	526		
(BULGARIA)	540		
CANADA	532	527	519
(COLOMBIA)	385		
CYPRUS	502	474	446
CZECH REPUBLIC	567	564	466
(DENMARK)		502	547
(FRANCE)		538	523
ENGLAND	513+*	506+*	
(GERMANY)		509+*	495
GREECE	492	464	
HONG KONG	587	588	
(HUNGARY)	548	537	483
ICELAND	474	487	534
IRAN, ISLAMIC REP.	429	428	
IRELAND	550	527	
(ISRAEL)	531	522+	
(ITALY)			476
JAPAN	597	605	
KOREA	611	607	
(KUWAIT)	400	392	
(LATVIA)	525	493*	
(LITHUANIA)		477+	469
(NETHERLANDS)	577	541	560
NEW ZEALAND	499	508	522
NORWAY	502	503	528
PORTUGAL	475	454	
(ROMANIA)	482		
(RUSSIAN FEDERATION)		535	471
SCOTLAND	520*	498	
SINGAPORE	625	643	
SLOVAK REPUBLIC		547	
(SLOVENIA)	552	541	512
(SOUTH AFRICA)		354	356
SPAIN		487	
SWEDEN		519	552
SWITZERLAND		545*	540
(THAILAND)	490	522	
UNITED STATES	545	500*	461

Mathematics International Average = 529 for Pop. 1
Mathematics International Average = 513 for Pop. 2
Mathematics General Knowledge International Average = 500 for Pop 3

Nations not meeting international sampling or other guidelines are shown in parentheses.

Nations in which more than 10% of the population was excluded from testing are shown with a +. (In Latvia, only Latvian speaking students were tested, which represents less than 65% of the population.)

Nations in which a participation rate of 75% of the schools and students combined was achieved only after replacement for refusals were substituted are shown with a *.

Sources:

- Mullis, I.V.S. et al. (1997) Mathematics Achievement in the Primary School Years. Table 1.1. Boston College: Chestnut, MA.
- Beaton, A. et al. (1996) Mathematics achievement in the middle school years. Table 1.1. Boston College. Chesnut Hill, MA.
- Mullis, I.V.S. et al. (1997) Mathematics and Science Achievement in the Final Year of Secondary School. Table 2.1. Boston College: Chestnut, MA.

TIMSS caused or was partly responsible for the initiation of curricular change in mathematics and science education in a number of the participating countries—mostly, but not entirely, the poorer performing countries. What follows is a survey of what happened in 23 of these countries. Information was obtained from a questionnaire sent to TIMSS representatives in participating countries, from TIMSS country reports, and from official documents and related sources.

The 23 countries for which information was available were as follows:

Argentina	Belgium(Flemish)	Belgium(French)
Canada	Cyprus	Czech Republic
Denmark	England	France
Germany	Hong Kong	Iran
Ireland	Israel	Japan
New Zealand	Norway	Scotland
Singapore	Spain	Sweden
Switzerland	USA	

The range of possible effects of TIMSS was structured under the following headings:

- Nature of official response to TIMSS.
- Degree of publicity given to TIMSS.
- Changes to mathematics curricula as a result of TIMSS.
- Changes to teaching methodology in mathematics as a result of TIMSS.
- General comments on the effect of TIMSS.

Nature of Official Response to TIMSS

In 14 of the 23 countries there was a national response to TIMSS, namely:

Belgium(Flemish)	Cyprus	Denmark
England	France	Germany
Iran	Japan	New Zealand
Norway	Scotland	Singapore
Sweden	USA	

The nature of the response varied from country to country as shown below.

Type of Response	Countries
PUBLICATION OF AN OFFICIAL REPORT	Belgium(Flemish) Canada(*) Denmark France HongKong(*) Iran Japan New Zealand Norway(*) Scotland Singapore Spain Sweden USA * Issued by the national TIMSS team.
NATIONAL REGIONAL CONFERENCES	Belgium(Flemish) England Iran Japan Scotland
FORMATION OF NATIONAL REGIONAL POLICY GROUPS TO PROMOTE CHANGE	Cyprus England Germany Iran Norway Scotland USA
PLANNING IMPLEMENTATION OF POLICY INITIATIVES	Cyprus Germany
INITIATION OF DEVELOPMENTAL PROJECTS	Belgium(Flemish) Norway USA

Publicity Given to TIMSS

Type of publicity	Countries
WIDESPREAD THROUGH MEDIA	Belgium(French)(*) Cyprus England Germany Norway Scotland Sweden Singapore Switzerland USA * For Science only.
MINOR ITEM IN NEWS MEDIA	Hong Kong Iran Ireland Israel Czech Republic Japan Spain
WITHIN EDUCATIONAL COMMUNITY	Belgium(Flemish) Canada Denmark New Zealand
LIMITED TO THOSE IN SENIOR EDUCATIONAL POSITIONS	France
NO PUBLICITY OUTSIDE RESEARCH TEAM	Argentina

Changes to Mathematics Curricula and Teaching Methodology as a Result of TIMSS

England, Cyprus, Denmark, France, Japan, Norway, Scotland, and Sweden all indicated a variety of changes in curricular emphasis, while England, Denmark, France, Japan, and Scotland also indicated changes to teaching methodology, mainly in the direction of increasing active pupil participation in the learning process

Individual Country Effects

We now look at the effect of TIMSS, country by country. Essentially direct quotations from questionnaires or official documents are given in quotation marks.

- ARGENTINA
  Results not included in official TIMSS report. Little governmental interest in the outcomes.
- BELGIUM(FLEMISH)
  Only Population 2 (13/14 years old) tested. No curricular action taken due (a) to the relatively high position in the comparative tables, and (b) to a perception that there were variables affecting student achievement which TIMSS had not considered.
- BELGIUM(FRENCH)
  Only Population 2 tested, performing moderately well. Main emphasis on Science results, with little publicity given to mathematics.
- CANADA
  In Canada there is no Federal Ministry of Education. Educational decision-making rests with individual provinces. For details, see Robitaille (1997a). The Canada TIMSS team have published two detailed reports, (Robitaille, 1996b, 1997b). Individual Canadian provinces—for, example British Columbia and Ontario—have revised their mathematics curricula in the wake of the TIMSS survey.
- CYPRUS
  Cypriot students performed relatively poorly in both Populations. Mathematics curriculum is under scrutiny. Some topics to be deleted from the curriculum.
- CZECH REPUBLIC
  In both Populations 1 and 2 Czech performance was good. "The Czech ministry of Education used the results to argue against innovation. Critics of Czech mathematics education based their arguments for change on TIMSS background variables—attitude to the subject , for instance."
- DENMARK
  Only population 2 tested. "Ministry of Education has focused on gender differences. Greater emphasis to be given to participation of girls in mathematics and science. Comparisons are being made between TIMSS results and national tests."
- ENGLAND
  England performed relatively poorly in the TIMSS tests. Detailed results will be found in Keys et al. (1996,1997). The main reaction was the setting up of a Numeracy Task Force which produced two Reports—*Numeracy Matters* and *The Implementation of the National Numeracy Strategy*—(Reynolds, 1998a,b), in which, as the second title indicates, a national numeracy strategy for England is developed. The essence of the strategy is contained in the following set of practices recommended to Primary school teachers (Reynolds, 1998b, p. 16):
  - teaching all pupils a daily 45 to 60 mathematics lesson;
  - teaching mathematics to all pupils within a class at the same time, with a high proportion of lessons concentrating on the development of numeracy skills;
  - teaching mathematics to the whole class or to groups for a high proportion of the time, promoting participation from, and co-operation between, pupils;

- including oral and mental work within each daily mathematics lesson;
- providing regular mathematical activities and exercises that pupils can do at home.

The complementary National Numeracy Project (NNP) with its detailed *Framework for Teaching Mathematics: Reception to Year 6* (Department for Education and Employment, 1999) emphasises the enhanced importance given to numeracy in the primary mathematics curriculum. A first evaluation of NNP is available from The National Foundation for Educational Research in England and Wales, (Minnis et al., 1999))

- FRANCE

  France participated in Population 2 only, performing moderately well somewhat ahead of England and Scotland. A national government report was published but there do not appear to be direct links between the TIMSS results and curricular change in mathematics.
- GERMANY

  Germany participated in Population 2 only, performing similarly overall to England and Scotland. "The Federal State Commission for Education Policy and Promotion of Research installed a group of experts to examine deficits in Science and Mathematics education and make suggestions for change. Their report was in published November 1997. As a consequence of this report an interstate five year program was installed with 15 of the 16 states (Laender) taking part. Under the co-ordination of the Institute for Science education (IPN) in Kiel, an intervention program was instigated in 180 schools to optimize science and mathematics instruction."
- HONG KONG

  Hong Kong students performed well. No government response. Minor item on news media. The Hong Kong TIMSS team have published two reports (TIMSS Hong Kong, 1996,1997).
- IRAN

  Iranian students performed comparatively very poorly in both Populations. "A group of educational experts has been formed to identify the reasons for students' low performance. During the last two years (i.e. 1997/8) many steps have been taken by the group and the national research co-ordinator in order to create positive attitudes to the outcomes of the project (for curricular change)and as a result tangible changes have been observed among educational policy makers as well as senior education experts. More emphasis to given to topics of proportion, data analysis, and measurement."
- IRELAND

  No direct publicity or government interest. Irish students performed somewhat better than those in England and Scotland but not markedly so.
- ISRAEL

  Israeli students overall performance was similar to that of England and Scotland. "Reports analysing national standing relative to other countries were published (in Hebrew) in the maths teachers journal for each of the TIMSS Populations. Very few take the results seriously. Many look for excuses and find ways to ignore TIMSS results."
- JAPAN

  Japanese students performed very well in both populations. "TIMSS revealed that Japanese children didn't like (mathematics). Therefore spontaneous activities were emphasised. In order to find time for this, topics were deleted from the curriculum. Greater emphasis was placed on children's' mathematical activities." A report of the Japan National Curriculum Council (1988) included the following recommendations:
  - "greater emphasis on practical and problem-solving activities, and on real-life contexts, in the process of acquisition of basic knowledge and skills in number, quantity, and geometrical figure;
  - "some reduction in curriculum content, in particular complicated computation and the use of complicated geometrical figures;
  - "use of repetitious learning as a help in mastering computation skills;
  - "establishing a new subject in upper secondary school incorporating

mathematical history and statistical processing of daily events, this subject to be a required elective."

- NEW ZEALAND

The performance of New Zealand students was very similar overall to England and Scotland. A full report is contained in Garden, (1996,1997) The New Zealand Government set up a Mathematics and Science Taskforce which reported in December 1997 (NZ Ministry of Education, 1997). Quoting from the initial Background Section of the report, "The Taskforce was established because of reported difficulties of classroom teachers (especially primary teachers) in implementing the new curricula for mathematics and science and in the light of the reported results of the Third International Mathematics and Science Study." In Section 2 of the report, entitled *Overriding Issues*, five concerns are identified and analysed. These are:

1. "The need to raise expectations;
2. "Under achievement amongst Maori and Pacific island students;
3. "Professional skills and knowledge of teachers;
4. "Material resources for teachers;
5. "Professional development."

In particular, the report places considerable stress on the availability of effective material resources, stating that its recommendations are made in a spirit of pragmatism and "are based on the realities if the current situation in schools, and not on idealistic notions of teachers' ability to invent rich activities by themselves and teach them with the pedagogical knowledge of an experienced researcher in (mathematics)education."

- NORWAY

Norwegian children performed similarly to those in England and Scotland in Population 2, but rather less well in Population 1. The main effect of TIMSS has been an increased emphasis on mathematics in the training of primary teachers. "Statistics to be given lesser emphasis."

- SCOTLAND

Scottish children performed disappointingly in both Populations 1 and 2 (Scottish Office Education and Industry Department, 1996, 1997a). The reasons for this are not fully understood and a variety of explanations have been put forward. For one analysis and overview see Macnab (1999). Scotland has also an internal standards survey—the Assessment of Achievement Project (AAP)—which has reported a continuing decline in standards of mathematics attainment since 1983, (Macnab et al., 1988; Robertson et al., 1993,1996; Scottish Office Education and Industry Department, 1998). The evidence of these reports has been largely ignored by the educational community for reasons explored in Macnab (1999a). However, publication of the TIMSS results has led to an official government report, *Improving Mathematics 5- 14* (Scottish Office Education and Industry Department 1997b), which put forward a series of recommendations for improving the situation, based at least partly on the perceptions of HM Inspectorate of Schools (Scotland) regarding characteristics of teaching in high performing TIMSS countries mainly in the Far East, and including:

  - Moving from mixed ability to some form of setting by ability.
  - Moving from individualised approaches to learning to more teacher-led whole class activity;
  - Reducing dependence on the calculator;
  - Increasing pupils facility in mental arithmetic.

Roughly contemporaneously with the publication of the report three regional conferences were organised to which both teachers and education administrators were invited. The effects of the report and the conferences on the teaching and learning of mathematics in Scottish schools will be the subject of a separate article, (Macnab, 1999b). They are outlined briefly in the section on Discussion of Survey Outcomes.

- SINGAPORE

Singapore students performed well in the TIMSS tests. A national report has been published on the TIMSS website. http://TIMSS.bc.edu. This report listed 7

possible reasons for this success.

1. THE HOMOGENEITY AND COHERENCE OF THE EDUCATION SYSTEM.
2. CHANGES TO THE CURRICULUM - placing greater emphasis on the development of mathematical concepts and the ability to apply them to solve mathematical problems.
3. THE WORKING ETHOS OF TEACHERS.
4. TRAINING AND PROFESSIONAL DEVELOPMENT.
5. HOME ENVIRONMENT - the virtue of hard work and the need to strive for excellence is ingrained in students in Singapore from an early age.
6. PEER INFLUENCE - while students in Singapore feel that doing well in schools is important, what is perhaps more important is that they also perceive their friends to place a similar emphasis on academic achievement.
7. FOSTERING OF INTEREST IN MATHEMATICS AND SCIENCE - the climate of opinion in Singapore is conducive to the learning of mathematics and science.

- SPAIN

  Spain participated in Population 2 only. No official government response. "There is no tradition of evaluation in Spain and up to now there are no channels created by the administration to spread and give relevance and impact on possible consequences to the outcomes of evaluations in which we take part, no matter whether they are national or international evaluations." A report in Spanish has been published by INCE, the Instituto Nacional de Calidad y Evaluacion, in Madrid.
- SWEDEN

  Sweden participated in Population 2 only, performing slightly better than England and Scotland. National government reports have been published in Swedish . Curriculum change is underway but not because of TIMMS as such.
- SWITZERLAND

  Switzerland participated in Population 2 only, performing moderately well. No government report has been published and no program of curricular change initiated.
- USA

  The United States did not come out well from the test results, although at both age levels it was placed above the UK countries. A national curriculum development program, *Attaining Excellence*, has been prepared involving a set of video-taped lessons from classrooms in the US, Germany, and Japan, together with an action strategy for improving achievement in mathematics and science. Two books have been published—*A Splintered Vision* (ASV) (Schmidt et al., 1997b) and *Facing the Consequences*(FC) (Schmidt et al., 1998)—which analyse the US results in their international setting and discuss in detail their consequences for US mathematics education. These publications reveal considerable soul-searching regarding the causes of the poor performance of the US. Three of the main conclusions reached are that US schools mathematics curricula are:
    - Too fragmented and lack coherence;
    - Cover too many topics and lack depth;
    - Concentrate too much on skills and too little on problem-solving.

Discussion

The most obvious outcome of the study is the difference in the degree of attention individual responding countries gave to the TIMSS results and in their reactions to them, varying from the extensive documentation emerging from the USA, and to a lesser extent the UK and New Zealand, to the almost nil. reaction in Argentina. In a number of countries - France and Sweden, for example - curricular change in mathematics education is in progress but not directly because of TIMSS.

The case of Scotland is interesting. The main recommendations for change contained in *Improving Mathematics Education 5-14* concerned matters such as

increased emphasis on whole-class teaching, inter-active teaching, and mental arithmetic, rather on the mathematics curriculum as a whole, its content and coherence. These recommendations were, moreover, agreed and accepted with virtually no dissent at the February 1998 Conferences (McKaig, 1998). There was not felt either by teachers or by the schools inspectorate - who in Scotland have a curriculum development role - to be any need to revise the 1992 curriculum document *National Guidelines: Mathematics 5-14*, which sets out official guidance on the mathematics curriculum and standards of attainment in the Primary and early Secondary years; indeed, the curriculum development emphasis from 1998 has been on Environmental Education.

This being so, it is a valid question to ask why the near unanimity on the way forward occurred. If teachers were indeed so persuaded of the rightness of the recommendations, why did they not implement them sooner? If not, why the sudden apparent enthusiasm to implement them now? It is still too early to judge in what measure implementation will actually take place, but an early survey (Macnab, 1999b) suggests that those at the conferences have moved to put at least some of the recommended changes into place and that school pupils perceive that change has occurred.

In England Wales, on the other hand, a much greater degree of prescription has been applied, with the publication of *The National Numeracy Strategy: Framework for Teaching Mathematics from Reception to Year 6*. This bulky loose-leaf format document, with a Foreword by the Secretary of State for Education and Employment in England and Wales, has been implemented in Session 1999/2000. It sets out not only macro aspects of teaching such as methodology and classroom organisation, but includes also a breakdown of lesson structure with time guides for the various elements. Detailed guidance on Oral Work, on Teaching Input and associated Pupil Activities, and on Lesson Conclusions is given. By far the greater part of the document, however, is devoted to a description of pupil learning outcomes relating to numerical work, of which the following example from Year 1 conveys the general character:

"Pupils should (be able to):

- Respond rapidly to oral questions phrased in a variety of ways such as:
  - 4 take away 2.
  - Take 2 from 7.
  - 7 subtract 3,.
  - Subtract 2 from 11,
  - 8 less than 9,.
  - What number must I take from 14 to leave 10?
  - What is the difference between 14 and 12?
  - How many more than 3 is 9?
  - How many less than 6 is 4?
  - 6 taken from a number leaves 3. What is the number?
  - Find pairs of numbers with a difference of 2.
  - I think of a number. I take away 3. My answer is 7. What is my number?
- Record simple mental subtractions in number sentence using + and - signs."

There are thus quite considerable differences between the two areas of the UK—England and Wales, and Scotland—in the degree of detailed guidance provided, and in the degree of consequential apparent leeway available., reflecting to some extent differing perceptions of the scale of the problem and so of the scale of reform required. Time alone will tell which of the two will be the more effective in implementation and in the effect on pupils' standards of attainment, although official figures (Summer 1999) have been published to show that standards in England and Wales are improving, in advance of the across-the board introduction of the Strategy. In Scotland we may have to wait for the results of the next round of the Assessment of Achievement Survey scheduled for Year 2000.

In the US different states have a freedom to devise their own mathematics curricula. California, for example, has prepared a set of mathematics standards (California, 1999) of which the Introduction says:

> These standards are based on the premise that all students are capable of learning rigorous mathematics and learning it well, and all are capable of learning more than is currently expected. Proficiency in mathematics is not an innate characteristic; it is achieved through persistence, effort and practice in the part of students and rigorous and effective instruction on the part of teachers.....The standards emphasise computational and procedural skills, conceptual understanding, and problem-solving. These three components of mathematical instruction and learning are not separate from each other; instead they are intertwined and mutually reinforcing.

We can see from these examples and from the generality of the survey evidence that a perception of the need for curricular reform in mathematics education is widespread, but that there is no overall consensus on the nature of the change required. I have argued elsewhere (Macnab, 1999c) that what may be missing in at least some of the poorer performing countries is the necessary will to ensure success in mathematics, by administrators, by teachers, by pupils and students, a will admirably expressed in the California Standards document quoted from above.

Surveys such as TIMSS perform a valuable service in that they give participating countries the opportunity in mathematics (and science) education to "see oorselves as ithers see us", to quote from Scotland's national poet Robert Burns. The survey reported here demonstrates that not all the countries made use of this opportunity; of those that did, not all were prepared to accept what was revealed; and that among those who did accept the verdict of TIMSS, there was not agreement as to the nature and depth of the changes required. Mathematics has a long history of being badly taught and worse understood. It would be pleasant that this time TIMSS will indeed make a difference.

References

Adams, R.S. & Chen, C. (1981). *The Process of Educational Innovation: An International Perspective.* London: Kogan Page.

Askew, M. & Wiliam, D. (1995). *Recent Research in Mathematics Education 5-16.* London: HMSO.

Black, P. & Atkin, J., M. (Eds.) (1996). *Changing the Subject: Innovations in Mathematics and Science Education.* London: Routledge.

Brown, M. (1996). FIMS and SIMS: the first two IEA International Mathematics Surveys. *Assessment in Education, 3,* 193-212.

Brown, M. (1998). The Tyranny of the International Horse Race, in Slee, R., Weiner, G., with Tomlinson, S. *School Effectiveness for Whom?* London: Falmer Press.

Brown, M. et al. (1998). Is the National Numeracy Strategy Research-based? *British Journal of Educational Studies, 46,* 362-385.

California State Board of Education (1999). *The Mathematics Content Standards for California Public Schools, Kindergarten Through Grade Twelve.* Sacramento CA: California State Board of Education.

Department for Employment and Education (1999). *The National Numeracy Strategy: Framework for Teaching Mathematics from Reception to Year 6.* London: DfEE.

Fullan, M. (1993). *Change Forces: Probing the Depths of Educational Reform.* London: Falmer Press.

Garden, R.A. (1996). *Mathematics Performance of New Zealand: Form 2 and Form 3 Students.* Wellington NZ: Ministry of Education.

Garden, R.A. (1997). *Mathematics and Science Performance in Middle Primary*

School. Wellington NZ: Ministry of Education.

Keys, W., Harris, S., Fernandes, C. (1996). *Third International Mathematics and Science Study First National Report Part 1.* London: NFER.

Keys, W.; Harris, S. & Fernandes, C. (1997). *Third International Mathematics and Science Study First National Report Part 2.* London: NFER.

Law, N, (Ed.), (1996). *Science and Mathematics Achievements at Junior Secondary Level in Hong Kong.* Hong Kong: University of Hong Kong.

Law, N, (Ed.), (1997). *Science and Mathematics Achievements at the Mid-Primary Level in Hong Kong.* Hong Kong: University of Hong Kong.

McKaig, G. (1998). *Improving Mathematics Education 5- 14: Conference Report.* Glasgow: St. Andrew's College.

Macnab, D.S.; Page, J. & Kennedy, M. (1989). Assessment of Achievement Programme Second Round Mathematics 1988. Aberdeen: Northern College.

Macnab, D.S. (1999a). Mathematics Education in Scottish Schools: An Uncertain Vision? *Scottish Educational Review, 31,* 10-20.

Macnab, D.S. (1999b). *Implementing Change in Mathematics Education: A Paper Presented to the 1999 Annual Conference of the Scottish Educational Research Association.* Aberdeen: Northern College.

Macnab, D.S. (1999c). *Improving Standards in Mathematics Education: Values, Vision, and TIMSS.* Aberdeen: Northern College.

Martin, M.O. & Kelly, D. L. (1996). *TIMSS Technical Report Volume 1:Design and Development.* Boston: Boston College.

Martin, M.O. & Kelly, D. L. (1996). *TIMSS Technical Report Volume 2: Implementation and Analysis.* Boston: Boston College.

Martin, M.O. et al. (1996). *Mathematics Achievement in the Middle School Years.* Boston: Boston College.

Martin, M.O. et al. (1997). *Mathematics Achievement in the Primary School Years.* Boston: Boston College.

Minnis, M. et al. (1998). *National Numeracy project Technical.* Report Slough: NFER

New Zealand Ministry of Education (1997). *Report of the Mathematics and Science Taskforce.* Wellington NZ: Ministry of Education.

Reynolds, D. et al. (1998a). *Numeracy Matters.* London: DfEE.

Reynolds, D. et al. (1998b). *The Implementation of the National Numeracy Strategy: Then Final Report of the Numeracy Task Force.* London: DfEE.

Robertson, I.J. & Meechan, R. C. (1992). *Assessment of Achievement Programme (Scotland) Mathematics Third Round 1991.* Glasgow: Jordanhill College.

Robertson, I.J. et al. (1996). *Assessment of Achievement Programme (Scotland) Fourth Survey of Mathematics 1994.* Glasgow: University of Strathclyde.

Robitaille, D.F. et al. (1993) *TIMSS Monograph No. 1: Curriculum Frameworks for Mathematics and Science.* Vancouver. Pacific Educational Press

BEST COPY AVAILABLE

Robitaille, D.F. et al. (1996a). *TIMSS Monograph No. 2: Research Questions and Study Design.* Vancouver Pacific Educational Press.

Robitaille, D.F. et al. (1996b). *TIMSS-Canada Report Volume 1: Grade 8.* Vancouver: University of British Columbia.

Robitaille, D.F. (Ed.) (1997a). *National Contexts for Mathematics and Science Education: An Encyclopaedia of the Education Systems Participating in TIMSS.* Vancouver: Pacific Educational Press.

Robitaille, D.F. et al. (1997b). *TIMSS-Canada Report Volume 2: Grade 4.* Vancouver: University of British Columbia.

Schmidt, W. H. et al. (1996). *Characterising Pedagogical Flow An Investigation of Mathematics and Science Teaching in Six Countries.* Dordrecht: Kluwer.

Scottish Office Education Department (1991). *Curriculum and Assessment in Scotland, National Guidelines Mathematics 5-14.* Edinburgh: SOED.

Scottish Office Education and Industry Department (1996). *Achievements of Secondary 1 and Secondary 2 Pupils in Mathematics and Science (TIMSS).* Edinburgh: SOEID.

Scottish Office Education and Industry Department (1997a). *Achievements of Primary 4 and Primary 5 Pupils in Mathematics and Science (TIMSS).* Edinburgh: SOEID.

Scottish Office Education and Industry Department (1997b). *Improving Mathematics Education 5-14.* Edinburgh: SOEID.

Scottish Office Education and Industry Department (1997). *Assessment of Achievement Programme: Fifth Round Survey of Mathematics.* Edinburgh: SOEID.

About the Author

Donald S. Macnab

Senior Research Fellow
Northern College
Aberdeen Campus
Hilton Place
Aberdeen AB24 4FA
UK

Phone: +44 1224 283552
Fax: +44 1224 283900
Email: d.s.macnab@norcol.ac.uk

Donald Macnab is Senior Research Fellow and Professor of Mathematics Education at Northern College, a teacher education institution affiliated with the UK Open University. He has published two books and a number of papers, and led a Scottish Education Department study into the mathematical attainments of Scottish Primary and Secondary schoolchildren - the Assessment of Achievement Programme. He is currently undertaking an extensive investigation into factors and processes with underly and influence curriculum development and change, with a particular reference to mathematics.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is

http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb. casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--- Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 1346 times since March 8, 2000

prior vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 8 Number 16 March 8, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES.**
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education.*

The Relationship between the Reliability and Cost of Performance Assessments

Jay Parkes
The University of New Mexico

Abstract
Performance assessments have come upon two major roadblocks: low reliability coefficients and high cost. Recent speculation has posited that the two are directly related such that cost must rise in order to increase reliability. This understanding may be an oversimplification of the relationship. Two empirical demonstrations are offered to show that more than one combination of sources of error may result in a desired generalizability coefficient and that it is possible to increase the number of observations while also decreasing cost.

The movement toward performance assessments for large-scale assessment purposes has encountered two major obstacles: first, such assessments have difficulty demonstrating highly reliable scores, and second, they tend to be very expensive. How these two problems are thought to be related influences the proposed solutions. This in turn will directly affect policies about the use of such assessments.

The problem of poor reliability in performance assessment scores stems from the lack of agreement among tasks, raters and other sources of measurement error. This is exhibited in a variety of types of performance assessments by several concurrent lines of inquiry, including: those by Shavelson and colleagues (e.g. Shavelson and Baxter,1992; and Shavelson, Baxter, and Gao, 1993); those from the Vermont Portfolio Assessment program (e.g. Koretz, Klein, McCaffrey, and Stecher, 1994; Koretz, Stecher, Klein, and McCaffrey, 1994; and Koretz, Stecher, Klein, McCaffrey, and Deibert, 1994); and one by McWilliam and Ware (1994).

Shavelson and colleagues have worked primarily with performance assessments in elementary level general science. By using the framework of generalizability theory, they have demonstrated that the greatest contributing facet to low generalizability coefficients is the task (e.g. Shavelson, Baxter and Gao, 1993). Furthermore, they project that by increasing the number of tasks a higher generalizability coefficient will result. Koretz and colleagues have worked with portfolio assessments of math and writing and identified raters and tasks as sources of error variance (Koretz, Stecher, Klein, McCaffrey, and Deibert, 1994). They, too, explore the possibility of increasing the number of tasks and the number of raters to achieve a more acceptable estimate of reliability. McWilliam and Ware (1994) examined the assessment of young children's engagement, and identified the number of sessions or observations as being a large

source of error variance. They estimated the minimum number of sessions that would be necessary to create an acceptably reliable assessment.

A second major concern with performance assessments is their high cost (Picus, 1994). Performance assessments are widely believed to be more expensive than multiple-choice testing (Catterall & Winters, 1994; Hardy, 1996; Linn, Baker & Dunbar, 1991; U.S. General Accounting Office, 1993), though the costs of performance assessments will vary considerably based on the exact nature of the assessment (Monk, 1996; U.S. General Accounting Office, 1993). Reckase (1995) demonstrated that it is possible to produce a writing portfolio assessment procedure that meets current standards of psychometric quality; but such a procedure, compared to current multiple-choice methods, would be a "very expensive alternative (p. 14)." White (1986), however, holds that, when designed properly, a direct assessment of writing can be conducted with comparable expense to that of multiple-choice assessment. This divergence notwithstanding, White (1986) recognized that the expenses are different for the two forms, the money being used mostly for raters in a direct assessment of writing. Hoover and Bray (1995) to some extent validated this claim by showing that the Iowa Writing Test could be conducted for approximately the same cost as the Iowa test of Basic Skills, albeit the former covered a much smaller domain than the latter.

These two problems of low reliability coefficients and high cost in performance assessment are often directly linked. If the solution to low generalizability is to increase the number of tasks, raters, etc., then the cost must also increase (e.g. Picus, 1994). There are a number of issues, however, that make this more complicated than it first appears.

The first issue is the automatic acceptance of the direct relationship between the number of observations in an assessment and the reliability of scores from that assessment. This acceptance is promulgated by a long history with the Spearman-Brown Prophecy Formula used to address this issue with objective item assessments. In a multiple-choice test, it is possible to estimate the number of items necessary to reach a desired reliability coefficient. For example, if a test contains 50 multiple-choice items and the reliability coefficient for scores from that test is 0.76, the Spearman-Brown Prophecy Formula can be employed to estimate how many items would need to be added to increase the reliability estimate to 0.85. There is direct (though asymptotic) relationship between the number of items used and the magnitude of the reliability coefficient. In a performance assessment, however, the relationship between a reliability estimate and the number of observations is more complicated because there are more sources of error. In a multiple-choice test, the items represent the only source of error. In a performance assessment, tasks, raters, occasions and potentially many other sources of error are possible. The implication is twofold. First, there may be more than one combination of raters, tasks, etc. that will result in a reliability estimate of a given magnitude. Second, it is possible that fewer observations could lead to a larger estimate of the reliability of scores from a performance assessment. Therefore, it is no longer axiomatic that increasing reliability means adding more observations.

The second issue is that cost and reliability are seldom addressed simultaneously. By and large this is due to the methodologies employed for such projections. In an assessment procedure with multiple sources of error, the most common projective technique is a liberalization of the Spearman-Brown Prophecy Formula, the decision study, or d-study from the generalizability theory framework. The d-study approach to addressing the joint issues of cost and reliability is less than desirable in a couple of ways.

D-studies are often done one at a time by considering different combinations of sources of error. That means that when the first combination to reach the desired reliability estimate is reached, the process stops. If there are several combinations of sources of error that would satisfy the desired reliability threshold, they probably would not be uncovered in this manner.

The d-study approach does not take cost information into consideration, which leaves the direct relationship between the number of observations and cost to dictate the best combination of sources of error. Assuming that d-studies are conducted in such a manner that multiple combinations of sources of error are identified, all meeting a minimum reliability estimate, the one with the fewest total observations is likely to be selected for implementation. It might be possible that more total observations could actually be less expensive. Without explicitly examining cost information, there is no way to know for sure.

The goal should be an optimal assessment design where optimal is defined as the most reliable and least expensive. There is a technique that allows all of these issues to be handled simultaneously in one analysis. Sanders, Theunissen, and Baas (1989, 1991, 1992) proposed the use of a branch-and-bound integer programming algorithm which searches for and identifies the optimal number of levels for each facet while taking into account each facet's contribution to the generalizability coefficient and each facet's cost as well as any other practical constraint. This technique appears to be promising. It can

exhaustively search all possible combinations of levels of facets, within given parameters, something that could be a daunting task to perform "by hand" using only psychometric constraints. Thus it gives reasonable assurance that the optimal solution has been located.

A second advantage of this technique is that it can accommodate a wide variety of logistical, economic, or other constraints. So cost data and reliability data, as well as other relevant issues, can be used simultaneously to define an optimal assessment design.

These issues and procedures will now be demonstrated using two different studies. The first study concludes that, depending on the definition of "optimal," there are many possible best combinations of facets to produce a predetermined generalizability coefficient. The second study produces data supporting the Sanders, et al. (1991) statement that it is possible to decrease the number of observations and/ or the total cost while increasing the generalizability coefficient. Both studies are based on the same set of data.

The Optimization Studies

Subjects. Fifty subjects enrolled in an undergraduate educational psychology class participated in the study. Twenty- eight percent of the sample were males and seventy-two percent were females. The sample also contained a mix of White, Asian-American, and Hispanic subjects. By class, the sample consisted of freshmen (20%), sophomores (52%), juniors (21%), seniors (5%), with the remainder unidentified. The sample had taken an average of 1.26 writing courses with a range from 0 to 3.

Procedures. Each subject read three articles—one about instructional approaches, and two articles about performance assessments—prior to attending the first of two 2 1/2 hour sessions. During the first session, subjects filled out a demographic questionnaire and wrote a separate 300 to 500 word essay about each of two prompts. During the second session, subjects wrote the other two prompts. In total, they wrote an expressive piece and a persuasive piece about the instructional approaches and an expressive piece and a persuasive piece about performance assessments. Four different orders of the prompts were counterbalanced to allow investigation of practice effects or other effects that may arise by writing the essays in a particular order.

Scoring the essays. Three graduate students in Educational Psychology served as raters and were trained. These raters were given the scoring rubric and discussed it; then, they scored a sample paper as a group. Using a slightly modified version of the Diederich scale (Diederich, 1974), each rater then read all 200 pieces of writing. The seven items on the scale were summed to achieve each subject's score on each piece of writing.

The Variance Models

The studies are based on a three-facet mixed design: mode of discourse (m), writing prompt (p), and rater (r). The object of measurement is student's overall writing ability (s). In the data collection design, prompts are nested within mode (i.e., p:m) and both cross raters and students. In the generalizability framework, the variance model is:

$$\sigma^2_{(\text{system})} = \sigma^2_s + \sigma^2_r + \sigma^2_m + \sigma^2_{p:m} + \sigma^2_{sr} + \sigma^2_{sm} + \sigma^2_{mr} + \sigma^2_{srm} + \sigma^2_{(p:m)s} + \sigma^2_{(p:m)r} + \sigma^2_{(p:m)sr} \quad (1)$$

The variance components for the sample in this study were estimated using the GENOVA software program (Crick and Brennan, 1983). Based on a review of the literature on modes of discourse (Crusius, 1989), there are at most five modes in existence. Therefore, for the estimation of variance components, the universe of modes was defined as having 5 levels. For all other facets, the universes were defined as infinite. The variance components estimated are shown in Table 1.

Table 1
Estimated Variance Components for Studies One and Two

Source of variation	Variance components
Subject (s)	5.8275728
Mode (m)	0*
Prompt:mode (p:m)	0*
Rater (r)	5.6756912
sm	0*
s(p:m)	2.6025238
sr	0.6714422
smr	0.3008503
sr(p:m)	11.8791415

*Note. Negative variance components were set equal to zero, following Brennan (1992).

For all subsequent optimization analyses, the relative model of measurement was used wherein the relative error variances were estimated through:

$$\sigma^2_{(\delta)} = \frac{\sigma^2_{sr}}{n_r} + \frac{\sigma^2_{sm}}{n_m} + \frac{\sigma^2_{srm}}{n_r n_m} + \frac{\sigma^2_{(p:m)s}}{n_m n_p} + \frac{\sigma^2_{(p:m)sr}}{n_r n_m n_p} \quad (2)$$

where n_r, n_m, and n_p are the number of raters, modes, and prompts respectively.

The G-coefficient of interest was therefore:

$$E\rho^2 = \frac{\sigma^2_s}{\sigma^2_s + \sigma^2_{(\delta)}} \quad (3)$$

Study One

In this study, results of a generalizability study and data describing the number of person-hours necessary to score the assessment have been used. Four different scenarios are presented, each with a different set of constraints, each producing a different optimal solution. The first scenario optimized the problem using only psychometric constraints; the second took a relative human factor constraint into consideration; the third used a specific human factor constraint; and the fourth used specific economic constraints.

The Optimization Scenarios

A branch-and-bound integer programming algorithm, a linear programming technique, was employed to estimate the optimal combination of raters, prompts within modes, and modes themselves. This investigation used the solver function of Microsoft EXCEL, version 5.0, to execute the algorithm. For all four scenarios, the variance components from Table 1 were entered into the worksheet. All four scenarios investigated shared a common objective function and a common set of constraints. In Scenarios 2, 3, and 4, additional constraints were considered. The common problem to be solved across all scenarios is:

Objective Function:

$$\text{Minimize } L = n_m\, n_{p:m}\, n_r ; \quad (4)$$

Subject to:

$$E\rho^2 = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_{(\sigma)}^2} \geq 0.8 \quad (5)$$

$$n_m < \text{or} = 5, \quad (6)$$

$$n_m\, n_{p:m}\, n_r \text{ are integers,} \quad (7)$$

$$\text{and } n_m, n_{p:m}, \text{ and } n_r > \text{or} = 1. \quad (8)$$

The objective function is to minimize the total number of observations needed. Constraint (5) specifies the minimal acceptable level of a generalizability coefficient. Constraint (6) specifies that there are no more than 5 possible modes of discourse. Constraints (7) and (8) ensure that solutions will be positive whole numbers.

In Scenario 1, the objective function defined in (4) subject to constraints (5) through (8) was submitted to the branch-and- bound search algorithm. The results of this search can be found in Table 2, which shows that, to attain a g-coefficient of at least 0.8, the minimum numbers are 4 modes with 2 prompts each while employing two raters to score each prompt in each mode. Based on data obtained from the sample, the average time needed to rate each prompt in each mode in this study was 0.092 hour (approximately 5.5 minutes). The total amount of time needed to rate the writings from ns subjects under any given scenario is then:

$$\text{Total person-hours} = n_m\, n_{p:m}\, n_r\, n_s(.092) \quad (9)$$

Applying Equation (9), the total person-hours needed for Scenario 1 for 50 subjects is 73.6.

Table 2
Results of Study One
Number of Cases Needed to Meet the Constraints

	Actual	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Additional Constraints			$n_r \geq n_m(n_{p:m})$	$n_m(n_{p:m}) \leq 6$	$n_m(n_{p:m})n_r(50)(.092) \leq$ 60 person-hours $n_m(n_{p:m})n_r(50)(.092) \leq$ 70 person-hours
Mode	2	4	1	1	4
Prompt: Mode	2	2	4	6	2
Rater	3	2	5	3	2
Obj. Function	12	16	20	18	16
Manhours	55.2	73.6	92	82.8	73.6
G Coefficient	0.75	0.80	0.80	0.80	0.80

An apparent practical problem with Scenario 1 is the demand on the examinee. A better solution might be one in which the burden of reliability is shifted away from the demand on the examinee to a demand on ratings per piece of writing. In Scenario 2, a new constraint was added to shift this demand to ratings. The additional constraint and the results can be found in table 2. To attain a g-coefficient of at least 0.8 while minimizing the burden on the examinee, the minimal design is one in which each examinee responds to 4 different prompts in a single mode of discourse. Each piece of writing needs to be rated by 5 raters. Under this scenario, the total number of writings from each examinee is only four. However, the total amount of person-hours needed for the rating of 50 subjects increases to 92 person- hours.

In Scenario 3, a compromise between Scenarios 1 and 2 was investigated by constraining the total number of pieces to six or less (see table 2). Under this scenario, each examinee must produce 6 pieces of writing in a single mode. On the other hand,

only 3 raters are needed for each piece to attain a g-coefficient of 0.8 or higher. The total person-hours for 50 subjects in this case is 82.8.

Scenario 4 investigated the cost factor. The lowest number of person-hours so far has been 73.6 in Scenario 1. Scenario 4 attempted to explore the possibility of a person-hour estimate lower than that. Table 2 illustrates the two constraints attempted. neither of which produced a feasible solution. In other words, it is not possible to expend less than 70 person- hours of rating activities to rate the writings used in this study for 50 subjects and still maintain a minimum g-coefficient of 0.8.

Conclusions from Study One

In a single-facet measurement situation, a multiple-choice exam for example, there is only one source of error to draw on to increase a reliability coefficient: items. So a one-to-one relationship exists between the number of the facet and the reliability coefficient: as the number of items increases, so does the reliability coefficient, albeit the relationship is asymptotic at some point. Also, there is a unique minimum number of items that will satisfy the desired reliability coefficient. For example, if a 50-item exam has a reliability coefficient of 0.69, the Spearman-Brown Prophecy Formula may indicate that in order to achieve a coefficient of 0.90, 83 items are needed. In a multi-faceted situation like the one represented here, the relationships are not so clear. With multiple facets, each contributing unequally in proportion to the size of its variance component to the generalizability coefficient, there is no simple one-to-one relationship. Scenario 1 uses psychometric constraints alone (as the Spearman-Brown Prophecy Formula or other projective techniques would) yet mode changes by 2 units, prompt within mode does not change, and raters decreases by one unit. Thus, in multi-faceted situations using only psychometric criteria, the relationship between the facets and the generalizability coefficient is not straightforward or simple.

Neither in a multi-faceted situation is there one combination which will uniquely fulfill the predetermined generalizability coefficient. The first step is to define optimal in some way. The optimization procedure allows a great deal of latitude in doing so. The four scenarios taken together demonstrate that there are many optimal combinations that will fulfill the predetermined generalizability coefficient.

Study Two

The second study is similar to the first except that instead of using person-hours as the economic constraint, it employs dollar figures. Second, instead of minimizing the total number of observations in order to constrain costs, it uses total cost as the objective function. The variance model in Study Two is the same as that in Study One.

The Cost Data

The cost data for this study are taken from Hoover and Bray (1995), who report on cost information for an administration of the Iowa Writing Assessment. The assessment tested the writing skills of 30,000 school students from grades three to twelve, each of whom wrote two pieces of writing. Each sample was scored twice holistically and twice analytically. For this assessment, Hoover and Bray estimate that $138,000 was spent in developing the 40 writing prompts; $174,410 was spent to score the prompts; and $30,000 was spent for materials. This breakdown is consistent with a framework for examining costs explained by Hardy (1996). In order to use this information in the optimization procedure, base units of development, scoring and materials need to be developed. That is, figures need to be obtained that indicate how much adding one rating (for example) to the scenario will change scoring costs, or how much adding one prompt will change development and scoring costs. The cost of development hinges on the total number of prompts developed--in Hoover and Bray (1995), 40—therefore, each prompt costs $3450 to develop ($138,000/40). In that study, each examinee wrote two prompts. Had each written only one prompt, presumably only 20 prompts would have been developed. Therefore, the $3450 is divided by 2, the number of prompts each examinee responded to, producing a cost per prompt required of an examinee of $1725. So that represents the base unit cost for development. Therefore, the development cost function is

1725n_{p:m}n_m$,

where $n_{p:m}$ is the number of prompts each person must write per mode
and n_m is the number of modes.

To obtain the base unit cost for scoring, the total scoring cost ($174, 410) was divided by the number of subjects (30,000), the number of pieces per subject (2), and

the number of raters or readings per piece (2) to produce a unit scoring cost of $1.43 per piece, per rater, per subject. The materials were estimated to cost $1.00 per subject. For the purposes of these analyses, the number of subjects was held constant at 50. Therefore, the total cost function, combining development, scoring and material costs, is:

$$\text{Total Cost} = \$1725n_{p:m}n_m + \$1.43n_{p:m}n_m n_r n_s + \$1.00n_s \quad (10)$$

The Optimization Problem

The variance components from Table 1, the cost function given in equation (10), and the number of prompts within modes, modes, raters, and subjects were entered into the EXCEL worksheet, and the following optimization problem was submitted for analysis.

Objective Function:

$$\text{Minimize } L = \text{Total Cost} = \$1725n_p + \$1.43n_p n_r n_s + \$1.00n_s, \quad (11)$$

subject to constraints (5) through (8) given in Study One.

The results are given in Table 3. Since the procedure was minimizing cost not the number of observation points, the optimal design includes more observation points (27 versus 12) but at less cost and a higher generalizability coefficient.

Table 3
Results of Study Two
Number of Cases Needed to Meet the Constraints

	Actual	Optimal
Mode	2	1
Prompt:Mode	2	3
Rater	3	9
Obj. Function	12	27
Total Cost	$7808	$7156
G Coefficient	0.75	0.80

Conclusions from Study Two

This second study provides empirical support for the claim made by Sanders, Theunissen, and Baas (1989) that it is possible to decrease cost while increasing the generalizability coefficient even when the total number of observation points increases.

Discussion

These studies serve as illustrations of the issues raised in the introduction. The first study demonstrates that it is possible to have many combinations of facets in an assessment design meet some predetermined level of reliability coefficient. The second study demonstrates the advantages of simultaneously considering cost and reliability data in the same analysis, namely, that it is possible to achieve a more reliable but less costly design.

Both of these points need to be taken in consideration during discussions about the cost implications of various solutions to the low reliability problem associated with performance assessment scores. If we assume that the only way to increase the reliability is to increase the number of observations and/ or we assume that increasing reliability will automatically increase cost, these stumbling blocks will not be removed. Policy makers will continue to be very reluctant to choose performance assessments as parts of their assessment plans.

These demonstrations represent a narrow perspective though and were designed to demonstrate only the two issues already mentioned. They are narrow in two ways. First, they may oversimplify the estimation of true costs of performance assessments. Second, they address only reliability and cost and not other concerns.

The costs associated here with performance assessments are expressed in dollars and cents and are rather simple. For example, development costs would change

depending on the number of examinees (Parkes, 1996). More examinees would require that more prompts be developed and the cost would probably change in some exponential fashion. This relationship is held constant by assuming the same number of examinees in each scenario. There are also many other ways to conceptualize cost, some of which would be very difficult to quantify. Monk (1996) and Picus (1994) describe the difficulties in determine the actual "costs" of a performance assessment. There are, of course, the financial expenditures associated with an assessment system. But more nebulously, there will be expenditure of time by students, teachers, and administrators to conduct these assessments. There is also cost in terms of what curriculum changes are made to accommodate the testing. That is, what would students be learning in the time taken for assessment.

The studies reported here are also narrow in that they address only reliability and cost and not other concerns. And there are plenty of other considerations that are equally as important or more important in the design of a performance assessment besides reliability and cost. The content sampling issue is one of these. Deciding how many tasks should constitute an assessment should probably be addressed in terms of content coverage first. Though certain constraints could be added to an optimization problem to account for content coverage issues, it probably not best to handle the issue in that manner. This approach treats each facet of the design equally or weights it based on its contribution to error variance. It therefore works on the implicit assumption that one rater means essentially the same thing as one task, which means essentially the same thing as one occasion, etc. But raters and tasks and occasions all serve different purposes in the assessment and contribute different things to the construct validity of the scores. So to trade three tasks for five ratings is, at best, contrived.

These issues provide a necessary context for the studies reported here but should not distract attention from the two central issues of this paper. First, more than one combination of sources of error may result in a desired generalizability coefficient. Second, it is possible to increase the number of observations while also decreasing cost.

Conclusion

The notion that only one design will generate a g-coefficient of a given value is not accurate. There are many possible combinations of facets, depending on how the optimal solution is defined, that will meet a desired g-coefficient value. The relationship between an assessment design and a corresponding generalizability coefficient needs to be more broadly understood.

The inference that generalizability coefficients and the number of observations are directly related is inappropriate. It is possible that several different designs would achieve acceptable generalizability coefficients. Similarly, a direct relationship between cost and reliability is not exact. Study Two shows that it is possible to increase the generalizability coefficient and the number of observations while decreasing the total cost of the assessment.

The bottom line for policymakers and those involved in performance assessment programs is that it is theoretically possible to have both a reliable and cost-effective performance assessment system. Assuming that low cost is the "line in the sand," those developing performance assessments should not assume that means they must minimize the number of ratings or the number of pieces in an assessment. Indeed, increasing certain aspects, like ratings, might actually end up being cheaper and still produce more reliable scores.

References

Catterall, J. S. & Winters, L. (1994, August). Economic analysis of testing: Competency, certification, and "authentic" assessments (CSE Technical Report 383). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Crick, J. E., and Brennan, R. L. (1982). GENOVA: A generalized analysis of variance system (FORTRAN IV computer program and manual). Iowa City: ACT.

Crusius, T. W. (1989). *Discourse: A critique and synthesis of major theories*. New York: Modern Language Association of America.

Diederich, P. B. (1974). *Measuring Growth in English*. Urbana, IL: National Council of Teachers of English.

Hardy, R. (1996). Performance assessment: Examining the costs. In M. B. Kane & R. Mitchell *Implementing perfomance assessment: Promises, problems, and challenges*

(pp. 107-117). Mahwah, NJ: Lawrence Erlbaum Associates.

Hoover, H. D., and Bray, G. (1995, April). The research and development phase: Can a performance assessment be cost effective? Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Koretz, D., Klein, S., McCaffrey, D., and Stecher, B. (1994). Interim report: The reliability of the Vermont portfolio scores in the 1992-93 school year. (RAND/ RP - 260). Santa Monica, CA: RAND.

Koretz, D., Stecher, B., Klein, S., and McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice, 13* (3), 5-16.

Koretz, D., Stecher, B., Klein, S., McCaffrey, D., and Deibert, E. (1994). Can portfolios assess student performance and influence instruction?. (RAND/ RP-259). Santa Monica, CA: RAND.

Linn, R. L., Baker, E. L., and Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20* (8), 15-21.

Monk, D. H. (1996). Conceptualizing the costs of large-scale pupil performance assessment. In M. B. Kane & R. Mitchell *Implementing perfomance assessment: Promises, problems, and challenges* (pp. 119-137). Mahwah, NJ: Lawrence Erlbaum Associates.

McWilliam, R. A., and Ware, W. B. (1994). The reliability of observations of young children's engagement: An application of generalizability theory. *Journal of Early Intervention, 18* (1), 34-47.

Parkes, J. (1996, April). Optimal designs for performance assessments: The subject factor. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York, NY.

Picus, L. O. (1994, August). A conceptual framework for analyzing the costs of alternative assessment (CSE Technical Report 384). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Reckase, M. D. (1995). Portfolio assessment: A theoretical estimate of score reliability. *Educational Measurement: Issues and Practice, 14*(1), 12-14.

Sanders, P. F., Theunissen, T. J. J. M., and Baas, S. M. (1989). Minimizing the number of observations: A generalization of the Spearman-Brown formula. *Psychometrika, 54*, 587-598.

Sanders, P. F., Theunissen, T. J. J. M., and Baas, S. M. (1991). Maximizing the coefficient of generalizability under the constraint of limited resources. *Psychometrika, 56*(1), 87-96.

Sanders, P. F., Theunissen, T. J. J. M., and Baas, S. M. (1992). The optimization of decision studies. In M. Wilson (Ed.) *Objective Measurement: Theory into Practice (Vol. 1)*. Norwood, NJ: Ablex.

Shavelson, R. J. and Baxter, G. P. (1992). What we've learned about assessing hands-on science. *Educational Leadership, 49*(8), 20-25.

Shavelson, R. J., Baxter, G. P., and Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement, 30*(3), 215-232.

U.S. General Accounting Office. (1993). Student testing: Current extent and expenditures, with cost estimates for a national examination (GAO/PEMD-93-8). Washington DC: U.S. General Accounting Office, Program Evaluation and Methodology Division.

White, E. M. (1986). Pitfalls in the testing of writing. In K. L. Greenberg, H. S. Wiener, and R. A. Donovan (Eds.), *Writing Assessment: Issues and strategies*. New York: Longman.

About the Author

Jay Parkes

Educational Psychology Program
128 Simpson Hall
University of New Mexico
Albuquerque, NM 87131-1246

E-mail: parkes@unm.edu

Jay Parkes is an Assistant Professor of Educational Psychology at the University of New Mexico. He received his Ph.D. in Educational Psychology from the Pennsylvania State University in 1998 specializing in applied measurement and statistics. His research interests include generalizability in performance assessment and cognitive and motivational aspects of performance assessments.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Andrew Coulson
a_coulson@msn.com

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Richard M. Jaeger
University of North Carolina— Greensboro

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs- Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
New York University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
Ann Leavenworth Center for Accelerated Learning

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois--UC

Robert Stonehill
U.S. Department of Education

Robert T. Stout
Arizona State University

U.S. Department of Education Arizona State University
David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doc.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 491 times since March 14, 2000

Education Policy Analysis Archives

Volume 8 Number 17 March 14, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

The Appalachian Rural Systemic Initiative: Improving Science and Mathematics Student Achievement in Economically Disadvantaged Rural Counties in Central Appalachia Through a School-Based, Teacher Partner Approach

Stephen A. Henderson
Appalachian Rural Systemic Initiative

Wimberly C. Royster
Kentucky Science and Technology Corporation

Abstract
This article was written in response to "Top-Down, Routinized Reform in Low-income, Rural Schools: NSF's Appalachian Rural Systemic Initiative, by Robert Bickel, Terry Tomaskek, and Teresa Hardman Eagle which was published in the *Education Policy Analysis Archives* as Number 12 of Volume 8 on February 21, 2000.

Introduction

"Top-Down, Routinized Reform in Low-Income, Rural Schools: NSF's Appalachian Rural Systemic Initiative" is a description of the authors' opinions (apparently primarily one person's opinion) of the Appalachian Rural Systemic Initiative and one of the strategies utilized to provide information regarding program improvement needs. The article does not accurately describe the ARSI project, is void of data, makes reference to unrelated research, fictionalizes the descriptions of personal observations, and includes more than fifty misrepresentations and/or false statements regarding the project. This rebuttal provides a more complete description of the ARSI project, describes the Program Improvement Review process and its role in the overall project, and provides data which supports the program's overall effectiveness.

It is apparent that the authors did not review the available information regarding the ARSI project or chose not to use that information in their article. ARSI has produced a number of publications and reports detailing the project's activities. The Year 4 ARSI Annual Report, published on the ARSI website since November, clearly describes the ARSI project and successes experienced through this model. Other rural, urban, or state systemic initiative reports may be obtained from the National Science Foundation.

This rebuttal will focus on the following ARSI strengths which are inaccurately

This rebuttal will focus on the following ARSI strengths which are inaccurately portrayed in the article "Top-Down, Routinized Reform in Low-Income Rural Schools: NSF's Appalachian Rural Systemic Initiative":

- ARSI as a "bottom-up" reform initiative.
- ARSI as a multi-dimensional process utilizing the Program Improvement Review as one, of many, means of accomplishing ARSI's aims."
- ARSI's potential to improve student achievement in rural counties in Appalachia.
- ARSI's focus of the uniqueness of rural schools.
- ARSI's successes in regard to science and mathematics program improvement and student achievement.
- The Program Improvement Review process and training procedures for potential reviewers.

The Real ARSI Project

The Appalachian Rural Systemic Initiative (ARSI) has made a major contribution in education reform through the implementation of a truly systemic school and district improvement model. Improved student achievement is being realized as ARSI focuses on K-12 students through the development and support of *catalyst schools* designed to serve as models for other schools in their district. The resulting *catalyst districts* serve as leaders for reform efforts throughout the region.

The ARSI model is based on a "bottom-up" team approach to school reform. A key component of the model is the development of *teacher partners*, who are designated by their schools as mathematics and science leaders. The teacher partner's work is supported by a team of professionals at the building and district level including the building principal, ARSI district liaison, and district superintendent. External support for the teacher partners and the development of catalyst schools and districts comes from five *resource collaboratives* located at university sites across Appalachia. These collaboratives are staffed by a director and mathematics/science specialists who, with support from university mathematics and science educators, provide training for teacher partners and direct services to catalyst schools in their region. Each catalyst school, led by the teacher partner, develops its own school improvement plan based on needs assessments, data analysis, and assessment of the instructional program.

Implementation of the ARSI model has proved to be effective in providing both direction for school reform and a mechanism for technical assistance to catalyst schools. ARSI has provided assistance through the development of school leadership, access to national and regional resources that support mathematics, science, and technology reform efforts, and improvement of the community support base. ARSI has made a major contribution through the development of standards-based curricula, science/mathematics content and pedagogy development workshops for teachers, identification of high quality instructional resources, while providing extensive support for the key ingredient of the ARSI model, the *teacher partner*.

One of the tools used for assessing program improvement needs has been the Science and Mathematics Program Improvement Review. This instrument is used to assess the program's effectiveness against a set of standards developed around "best practices" which are consistent with mathematics and science state and national standards. Needs assessment data gathered through this process has been utilized in both school and district strategic planning efforts.

ARSI as a "Bottom-up" Reform Initiative

The ARSI project utilizes a school-based approach to program improvement. The basic premise of the ARSI model is that reform and improvement of science and mathematics programs is best done in rural schools through the teachers and principals in each school. The ARSI team, consisting of the teacher partner, ARSI district liaison, principal, and superintendent, has been the primary planning group in each district and is supported by the resource collaborative housed at an area university. The ARSI emphasis has been on the identification of program needs, assistance in developing both short range and long range improvement plans, and in the provision of technical assistance in the development of curriculum and selection of appropriate resources. Professional development has been primarily "job-embedded." The primary functions of the teacher partner have included such activities as mentoring of other classroom teachers, modeling inquiry teaching strategies, and assisting teachers plan for inquiry based instruction.

A major service provided by the ARSI staff has been to assist schools and

BEST COPY AVAILABLE

districts with strategic planning. The Program Improvement Review has been a welcomed source of needs assessment data from which the teacher partner, principal and other science and/or math teachers have constructed their own improvement plan. Based on the needs assessment data, ARSI has facilitated school and district reform efforts by providing professional development, assisting in the identification of resources, and providing guidance in regard to curriculum development and instructional improvement. In no case, as implied in the article "Top-down, Routinized Reform in Low- income, Rural Schools: NSF's Appalachian Rural Systemic Initiative," has ARSI dictated how a participating school or district proceeds with their science and/or mathematics program reform efforts or constructed a "one-size fits all" approach to school/district assistance efforts.

After a review of the first four years of the ARSI project, Inverness Research Associates, the ARSI project external evaluator, made the following statements concerning the ARSI approach to school reform:

> "The ARSI model is developmental and works from the inside out. That is, ARSI starts by identifying and building leadership within the district through its work with teacher partners. The teacher partner, with the help of the district liaison, then builds a core group of teachers and administrators who are committed to the reform effort. Eventually the reform effort may move to the level of district policy—curriculum, professional development, etc.—and then out to the community and national scene."
>
> "ARSI is a subtle reform effort that is steadily building within each district a grassroots group of teachers and district leaders - people who are knowledgeable about and, increasingly, advocates for inquiry-based, student-centered, hands-on teaching and learning."

The Program Improvement Review: One of Many Means to ARSI's Goals

The statement "The primary means of accomplishing ARSI's aims is a one-day-one-school visit," indicates a lack of knowledge regarding the ARSI project. (Bickel et al., 2000) ARSI incorporates a wide variety of interventions and assistance to schools in their reform efforts. The primary means of accomplishing ARSI's aims is the utilization of "teacher partners" to mentor other teachers, provide professional development, coordinate curriculum development efforts, obtain quality resources, and work with parent and community groups to promote science and mathematics education. The teacher partner is selected on the basis of his/her general leadership ability, skill as a mathematics or science teacher and potential for providing assistance to other teachers. Teacher partners receive monthly training in both content and pedagogy through the ARSI resource collaboratives. In addition to the training and support provided by the teacher partner, professional development is being provided for teachers in participating district schools by both the ARSI curriculum specialists and university math and science educators. Training is being provided in inquiry instructional techniques, authentic assessment strategies, data analysis, and standards-based mathematics and science content. In all cases, the training provided at the school level has been requested by the school on the basis of needs identified at that level.

The Program Improvement Review is but one tool, of many, utilized by ARSI to provide needs assessment data to schools involved in the ARSI project. In fact, the Program Improvement Review is not a requirement for participation in the ARSI program and is utilized only at the request of the individual school. The process has proved so beneficial, however, that most schools have voluntarily participated in the process and in several cases, districts (ARSI and non-ARSI) have requested that the process be completed in all schools to provide data for program planning.

ARSI Project Potential to Improve Student Achievement in Rural Counties in Appalachia

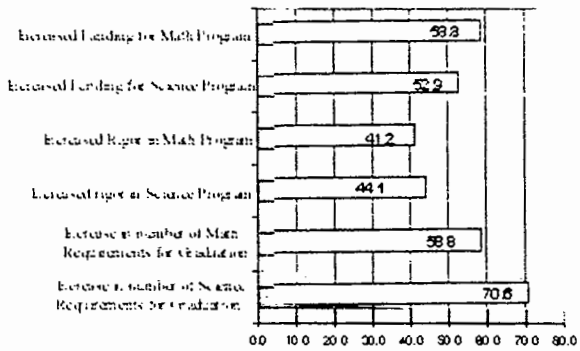
During the four and one-half years of the ARSI project, it has become clear that the school districts in Appalachia differ widely in their "readiness" and ability to participate in significant reform efforts. At the outset of the project none of the participating schools had district-wide curricula in science or mathematics aligned with their state or national standards. School leaders lacked a "vision" of quality mathematics and science programs which would provide direction for reform efforts. Professional development was primarily district based and generally focused on

Professional development was primarily district based and generally focused on generic topics such as improving school discipline or improving student safety in schools. Although these topics are certainly important, teachers also need a consistent, well-planned professional development program focusing on both content and pedagogy.

Professional development, through the ARSI teacher partner has been one of the major foci of the ARSI program. There is clear evidence that the quality of instruction is improving as a result. Improved instruction, use of standards-based materials designed to promote student inquiry, and well defined curricula focusing on state and national standards are now commonplace in ARSI schools and the student achievement data, included in this document, show clearly that use of the ARSI model has resulted in positive results across the region. Another focus area for ARSI has been the development of policies, at both the school and district level, which increase mathematics and science learning opportunities. Policies designed to increase the breadth and rigor of programs and the support for mathematics and science in Appalachian schools, have been implemented in many ARSI districts. See Figure 1.

Figure 1

Percent of ARSI Districts Making Policy Changes



ARSI: Positive Results Across the Appalachian Region and in States Served by ARSI

One of the most positive results of the ARSI project has been the development of skilled and committed leadership for mathematics and science program improvement. "There is no doubt that the greatest contribution of ARSI lies in this area: ARSI is helping districts identify, train and support local leaders who are knowledgeable about math and science reform and empowered to work towards change in schools and classrooms." (Inverness Research Associates, External Review Report, 2000) ARSI's efforts in training teacher partners, ARSI catalyst school principals, and ARSI district liaisons have resulted in a district team that has provided extensive leadership for science and mathematics program reform efforts.

Student achievement data for ARSI catalyst schools validate the impact of the ARSI model. Catalyst schools that started the program during its first year (having had ARSI interventions for two full years), show a dramatic increase in student achievement in both mathematics and science. In science, students scored above the combined states' average and were significantly higher than comparison districts in the Appalachian region. Mathematics scores were slightly below the states' combined average, although the gap was significantly reduced, and students scored well above their Appalachian region counterparts.

As would be expected, the gains for schools involved with the ARSI project for only one year are not as dramatic although ARSI catalyst schools that started the program in its second year demonstrate similar trends. Student achievement in science shows a similar percentage of improvement, as did the students from the inaugural year whereas the mathematics performance increased only slightly. See Figures 2 and 3.

Figure 2

Science Performance

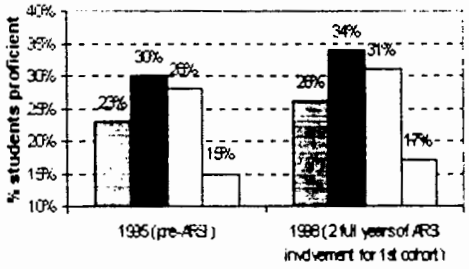
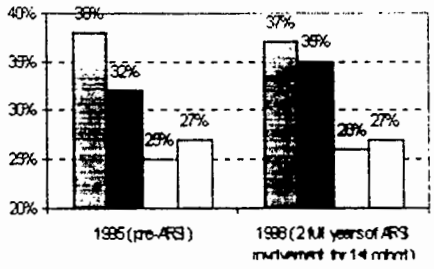


Figure 3

Mathematics Performance



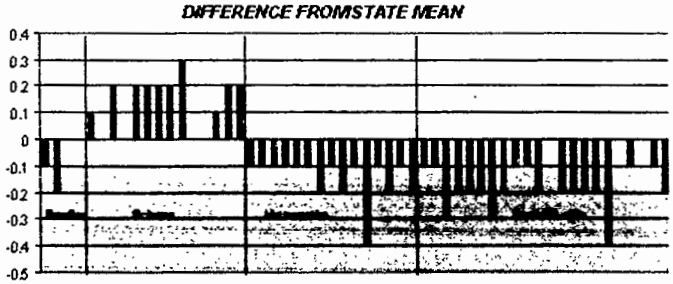
BEST COPY AVAILABLE

In examining individual school data, the results are even more dramatic. See Figure 4. An ARSI school that has had a full range of interventions in science demonstrates the type of results achieved through the project. The school started with the Program Improvement Review which identified several weaknesses including lack of a curriculum in science and little emphasis on inquiry-based instruction.

After the implementation of an aligned, standards-based curriculum and extensive staff development in inquiry based instruction, student achievement in science exceeded the state average in all assessed sub-domain areas whereas student achievement in all other content areas was below the state average.

Figure 4

Example of ARSI Middle School Showing Improvement After ARSI Involvement

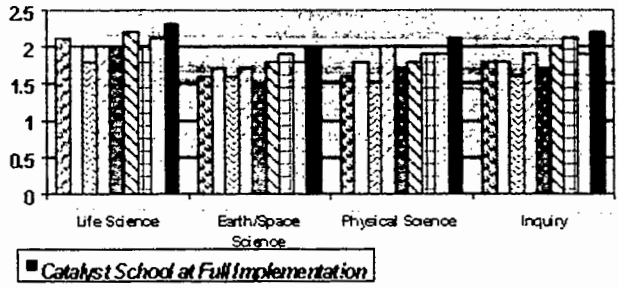


Assessment areas include Reading, Science, Math and Social Studies with only Science showing improvement--the area where Program Improvement Reviews were conducted and extensive work was done by ARSI staff in assisting schools in implementation

The data for another ARSI district with nine (9) elementary schools is equally impressive. As in the previous example, the ARSI catalyst school implemented an aligned, standards-based curriculum and provided inquiry-based instruction professional development for teachers through the ARSI teacher partner. As can be seen in the graph to the right, the ARSI catalyst school scored above all other district elementary schools in every science sub-domain area. See Figure 5 above.

Figure 5

Mean Scores for Elementary Schools in ARSI Catalyst District



These data are not unique. 1999 ARSI schools' state assessment data is currently being analyzed. The preliminary results indicate substantial improvement for nearly all ARSI schools since the inception of the ARSI project in 1996.

ARSI Project Focus on the Uniqueness of Rural Schools

"There is something about "rural-ness" that is important. These are small, closed communities. So, any effort to change the mind set, or to change the value system or the valuing of things, is difficult because it is a closed system. I think what we are seeing is a slow, steady battle to win hearts and minds--and having a local, well respected, well trained, well supported, well chosen teacher partner is the way to go about it. As one district superintendent said, 'Mountain people are just old mules--it is easier to lead them than it is to push them." (Inverness Research Associates, ARSI External Review, 2000)

The ARSI project has been sensitive to the characteristics and needs of *rural communities* since its inception. Characteristics common to rural communities have long been known to researchers and ARSI is cognizant of the necessity of attending to the specific needs of these communities *if* the school reforms initiated are fully implemented and persist beyond the years of ARSI involvement. In addition to being rural, the Appalachian region school districts participating in the ARSI project are similar in that they reside in counties with poverty levels of school age children

greater than 30% (according to the 1990 census) and USDA Beale Numbers 6 or higher.

The principal ARSI goal, "to accelerate performance in science, mathematics, and technology in Central Appalachia," addresses one of the major educational challenges of rural communities. Formal education attainment tends to be lower in these areas. High school completion rates are lower than those in metropolitan areas and fewer rural students complete college (Herzog & Pittman, 1995). Rural students are also less likely to take college preparatory classes (Stern, 1994) often resulting in the need for *remedial classes* in science and/or mathematics upon their entry into a community college or university.

Another goal for the ARSI project, "to develop a sustainable system providing students and teachers with timely, coordinated access to educational resources and services ...." addresses the "isolation" of these communities. Fewer institutions of higher education are located in rural areas and educators feel more professionally isolated than their metropolitan counterparts (Massey & Crosby, 1983; Stern, 1994). "Through ARSI, each of these districts, especially the teacher partners and district liaisons, have become affiliated with at least one university as well as other state resources such as national education laboratories, museums, and other NSF projects." (Inverness Research Associates, ARSI External Review, 2000)

Rural areas often have difficulty attracting and retaining mathematics and science teachers. This results in a large number of teachers teaching "out of field" and generally these teachers are unfamiliar with current resources for standards based mathematics and science instruction. A recent study by the Kentucky Department of Education showed that fully a third of the teachers in Kentucky lack the necessary mathematics background and certification to teach middle school content (Clements, Hartanowicz, and White, 1998). In many of the ARSI districts, the percentage is even higher. The ARSI teacher partner has been a major factor in improving the qualifications of mathematics and science teachers in the participating school districts.

The social norms of rural areas value family, place, and community over other priorities. The school in a rural community is often the "center" for community activities. (deYoung & Lawrence, 1995, Herzog & Outtnmabm 1995, Nachtigal, 1982, Stern, 1994) Recognizing this importance, increasing "community engagement," has also been a major objective of the ARSI project.

The Program Improvement Review: A Tool for Assisting Schools in Identifying Science and Mathematics Program Needs

Since the article, "Top-Down, Routinized Reform in Low-income, Rural Schools: NSF's Appalachian Rural Systemic Initiative," was primarily a critique, be it uninformed, of the Program Improvement Review process, it is important that the procedures utilized and the training program be explained.

The Program Improvement Review is a program assessment process developed to provide schools an "outside" view of their programs as measured against a set of clearly identified standards. The process involves a site visit to the school by a team of trained observers who collect data through interviews with the school principal, teachers, parents, and students, classroom observations, review of the school's curriculum, review of instructional resources, and review of testing procedures and data. A *classroom observation instrument* is used in the Program Improvement Review which guides the reviewer's observations related to student-teacher and student- student interactions. Student engagement and interaction, as well as the teacher's questioning strategies, are critical pieces of the data collected related to inquiry based instruction. Following the site visit, the school is provided a written summary of the site visitors' observations including recommendations for making improvements in the instructional program.

Debriefing with the site visit team occurs immediately following the visit. It takes approximately 10-12 hours to draft a report. After meeting with team members, editing, and publishing the report, the report is delivered to the school in 2-4 weeks. Although reported in the article that "The final report, usually written overnight and presented the next day," there has **never** been a case in which the report was generated overnight and presented the next day.

The Program Improvement Reviews are based on "recognized good practice" and national standards as identified in a set of clearly defined *look-fors*. The *look-fors* are translated into a set of standards which help the reviewer collect data from a variety of sources. The procedures utilized are modeled after the procedures designed by Fenwick English in his *Curriculum Auditing* process as utilized by PDK and site visit procedures developed as part of the U.S. Department of Education's *Blue Ribbon Schools Program*. The approach is not unlike the procedures utilized by the Southern Association of Colleges and Schools (SACS), North Central

Association, or other such accrediting agencies. The primary difference between the Program Improvement Review and these types of programs are the Program Improvement Review's specific emphasis at the *program level*.

No claims regarding "...easy-to-understand, easy to evaluate nature of education achievement in rural Appalachian Schools," have ever been made by ARSI or the developers of the Program Improvement Reviews. Quite the contrary. The reviews are only one piece of assessment data utilized in assisting schools develop both short-range and long-range plans for improvement. The Program Improvement Reviews were developed as a result of a specific need identified by local school districts. The standards and sub-standards are based on the classroom practices of experienced math and science educators and are consistent with standards as specified by NCTM and National Research Council.

The Program Improvement Reviews, as designed and utilized in the ARSI project, have **never** been used to **evaluate** a school or a school program. The ratings, comments, and recommendations are a synopsis of the "one-day snapshot" and designed to provide schools with insight not normally found by "self-evaluations," questionnaires, or other routinely used procedures.

The instrument utilized in West Virginia was developed by West Virginia educators. The procedures described in the article, "Top-down, Routinized Reform in Low-Income, rural Schools: NSF's Appalachian Rural Systemic Initiative," are specific to the West Virginia process which, as initially implemented, is vastly different from the Program Improvement Review process utilized in other ARSI states. The project team at Marshall University developed their own procedures and instrument specific to West Virginia. ARSI gave permission to this team to *adapt the instrument* and, although much different the West Virginia instrument is referred to as a *Program Improvement Review*.

Because of the relatively short time that Program Improvement Reviews have been utilized, approximately 5 years, definitive results are just now being identified. Data are being compiled which shows clearly the impact of the Program Improvement Review Process on individual school reform efforts as part of the ARSI project. In addition to individual school and district data, a database is currently being developed to identify trends among all schools reviewed and the specific needs of schools across Appalachia. As stated, the Program Improvement Review process is an evolving one, based on identified *best practices* and formulated with much input from school clients, both present and future.

Science and Mathematics Program Improvement Review Training Program

The "formal" training session consists a 6-hour session focusing on the various aspects of the process including interviews, classroom observation, and data analysis. The training day begins with an introduction to the process including the assumptions as well as the research and *practice* basis for the procedures utilized. A simulation is utilized to prepare reviewers for conducting the on-site interview sessions. To insure consistency in classroom observation reports, a significant amount of time is spent on the observation and scripting of a classroom setting via videotapes. This is followed by a comprehensive analysis of the participants' observations, a review of student assessment data and how this data is utilized, and a time for reflecting on actual school data for the purpose of preparing a summary report. In regard to the extended description of a "videotape segment" in the training tape in "Top-Down, Routinized Reform in Low-Income, Rural Schools: NSF's Appalachian Rural Systemic Initiative", it is important to note that this part of the scripted observation is approximately 2 minutes long out of a 30 minute training tape.

This formal training session is followed by a "shadowing experience" in which the "trainee" participates in the data collection process and assists with writing various sections of the summary report. In regard to the quality of the report provided the school, it has proven to be very important that potential *reviewers* participate in all phases of the site visit and report writing process *prior to* assuming the role of a program reviewer.

It is also important to note that the West Virginia project (described in "Top-Down, Routinized Reform in Low-Income, Rural Schools: NSF's Appalachian Rural Systemic Initiative") requested that they be allowed to deviate from the normal training program. Although against its better judgement, ARSI complied with this request.

Conclusions

ARSI's model, using a team approach to systemic reform, has produced desired results, namely, standards based instruction in mathematics and science, implementation of supportive policies, convergence of resources for mathematics, science and technology education improvement, a broader base of community support, and increased student achievement.

The four main intervention approaches—Catalyst Schools and Teacher Partners, Program Improvement Reviews, Community Engagement, and Resource Collaboratives/University Partnerships—recognize the importance of "bottom-up" strategies for school reform in rural schools. Among these interventions it has been stated: "The Program Improvement Review and Planning Process may be the most important of all the intervention strategies used by ARSI." (Smith, 1999-2000)

The Program Improvement Review does not operate in a vacuum. ARSI has focused on "school-based" leadership in the form of the ARSI teacher partner supported by the local district team consisting of the school principal, ARSI district liaison, and district superintendent. The ARSI resource collaboratives have served this model through the provision of professional development for the teacher partners, assistance in the identification of quality mathematics and science instructional resources, provision of leadership training for principals, and development of networks with universities and other professionals who can assist in school reform efforts.

The development of a skilled and committed leadership for mathematics and science program improvement has been one of the most significant results of the ARSI project to date. Because of ARSI's training, the district teams now have a "standards- based vision" of mathematics and science instruction which is providing direction for district reform efforts.

It is also apparent that ARSI's focus on K-12 students through the development and support of *catalyst schools* and leadership of the teacher partner has resulted in improved student achievement. Both aggregated state data and individual school data indicate the positive effects of the ARSI project. Because of the success obtained, ARSI catalyst schools are beginning to serve as models for other schools in their district further validating the project's potential for school reform in the Appalachian Region.

The data, obtained after four and one-half years of ARSI activity, clearly indicate that ARSI is a major partner in the school improvement process for low-income rural schools in Appalachia.

References

Appalachian Rural Systemic Initiative. (1999). *ARSI: Empowering Children With Mathematics and Science.* Lexington, KY. 3-9.

English, Fenwick. (1988). *Curriculum Auditing.* Lancaster, PA. Technomic Publishing Co., Inc.

Kannapel, Patricia J. and DeYoung, Alan J. (Fall, 1999). The Rural School Problem in 1999: A Review and Critique of the Literature. *Journal of Research in Rural Education.* Volume 15, No. 2, 67-79.

Smith, Keith. (Winter, 1999-2000). Implementing Systemic Change for Math and Science in Rural Appalachia: The Appalachian Rural Systemic Initiative. *The Rural Educator.* Volume 21, No. 2, 3-8.

St. John, Mark, Brown, Samantha, Carroll, Becky, Hirabayashi, Judy, Horsch, Elizabeth, and Howard, Mike. (2000). Appalachian Rural Systemic Initiative (ARSI): Final Report. External Evaluation Report.

About the Authors

Dr. Stephen A. Henderson
Dr. Henderson is currently Director for the Appalachian Rural Systemic Initiative. Dr. Henderson has extensive experience as a classroom teacher, school administrator, and university science educator in Michigan, Virginia, and Kentucky.

Dr. Wimberly C. Royster
Dr. Royster is currently Principal Investigator for the Appalachian Rural Systemic Initiative. Dr. Royster directs Kentucky's Statewide EPSCoR project and Annenberg Rural Challenge, and is a Vice President and Director of Scientific Research and Policy for the Kentucky Science and Technology Corporation. He has served as

Dean of the Graduate School, Vice Chancellor, and Vice President of Research and Graduate Studies at the University of Kentucky.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umea University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apemberto@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doc.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

Este artículo ha sido consultado 357 veces desde el 20 de marzo de 2000

Education Policy Analysis Archives

Volume 8 Number 18 marzo 20, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES.**
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education.*

Lineamientos de Política Educativa en los Estados Unidos: Debates Actuales; Significados para América Latina

Mónica Eva Pini
Universidad de Buenos Aires

Resumen
El propósito de este trabajo es describir y analizar los lineamientos y debates más relevantes en la política educativa en los Estados Unidos y explorar la universalidad de algunas de las ideas y estrategias que ya se están implementando en América Latina. Las políticas seleccionadas para discutir aquí son las metas de educación de los Estados Unidos para el año 2000 (Goals 2000), y las propuestas de elección de escuela por parte de los padres (parents school choice), especialmente escuelas contratadas (charter schools) y bonos (school vouchers). Estas políticas cuestionan definiciones clave como los objetivos de la educación, el rol del estado y valores democráticos. A través del análisis de diferentes documentos, informes e investigaciones y su contexto de surgimiento, con una perspectiva crítica, se interpretan los significados de los discursos y políticas. El modelo de mercado y las poderosas fuerzas que respaldan muchas de estas propuestas hace necesario que pensemos cómo estas ideas afectan la distribución social de la educación y los ideales democráticos, tanto en los Estados Unidos como en América Latina.

Educational Policy in the United States:
Context of Current Debates; Impact in Latin America

Abstract
The objective of this article is to analyze some of the most relevant debates about current educational policies in the United States, and to

explore the universalility of some of the ideas and strategies which have already been put in place in Latin America. The policies I will discuss are Goals 2000 and school choice proposals, especially those of charter schools and vouchers. These policies question the key definitions of educational goals, the role of the state, and democratic values. Through the analysis of different documents, reports, and research studies, the political context from which they emerged, and a critical perspective, discourses and policies are interpreted. The market model and the strong forces behind many of these proposals makes it necessary for us to think about how these ideas affect the social distribution of education and democratic ideals, both in the United States and in Latin America.

1. Introducción

El propósito de este trabajo es describir y analizar los lineamientos y debates más relevantes en la política educativa en los Estados Unidos. Conocer mejor el marco político institucional en el que se insertan las prácticas educativas norteamericanas nos permite desarrollar una perspectiva más amplia acerca de los procesos de gestación de ideas y estrategias que luego se exportan a (o se importan desde) América Latina desprovistas de contexto y como respuestas válidas universalmente.

Las políticas seleccionadas para describir y analizar en este trabajo son las metas de educación de los Estados Unidos para el año 2000 (*Goals 2000*), que nos brindan una visión general a nivel nacional, y las propuestas de elección de escuela por parte de los padres (parents school choice), bajo sus diferentes formas (Nota 1). Este último tema, que ya está presente en América Latina, es objeto de una de las polémicas más candentes en la educación norteamericana actual, ya que pone en cuestión definiciones fundamentales del sistema tales como para qué y para quiénes educa, cuál es el rol del Estado, la vigencia de valores democráticos como libertad y equidad, la separación entre la Iglesia y el Estado, qué clase de participación se busca, y cómo se distribuye ese bien intangible que es la educación.

En países como Argentina, en los que el debate democrático está dejando lugar al cansancio de un sistema político que no ofrece mejoras sociales para el conjunto de la población, la necesidad de analizar y discutir este tipo de políticas se vuelve fundamental. Para superar, al menos en el texto, la ambigüedad de un término que se utiliza para designar procesos diferentes, el concepto de democracia usado aquí se refiere a "la decisión colectiva y conciente sobre el proceso de producción material de la vida, significa la constitución de la sociedad en sujetos que deciden su destino" (Lechner, 1986, citado por Pini, 1993, p. 7). Por lo tanto, se opone a la subordinación de las relaciones sociales a las leyes del mercado, que pretende abolir la política, y requiere en cambio del fortalecimiento de la responsabilidad social. Decía en otro trabajo (Pini, 1993) que la reproducción de los privilegio y la exclusión de gran parte de la población del reparto de la riqueza cuestiona y restringe la construcción democrática. La misma preocupación orienta ahora este análisis, en vista de que las presiones para introducir el modelo de mercado en educación siguen creciendo.

Una parte importante del trabajo es el desarrollo del contexto histórico político en el que se originan estos objetivos y políticas en los Estados Unidos, y fundamentalmente, las fuerzas que se mueven en la lucha por el control de la escuela pública, que es como decir por el control de la opinión pública. A partir de este marco me propongo reflexionar acerca de la lógica y las conexiones de estas políticas con algunas que se están proponiendo y llevando a cabo en América Latina, particularmente en Argentina.

2. "Goals 2000"

En 1989 se realizó una reunión sobre educación (Charlottesville Education Summit) en la que participaron el Presidente de los Estados Unidos George Bush y todos los gobernadores, liderados por el entonces Gobernador Bill Clinton. En ella se enfatizó

la necesidad de una respuesta nacional para los problemas educativos, y en función de esto se acordaron una serie de compromisos y acciones, entre los que se contaban la creación de Objetivos Nacionales de Educación que proveyeran un marco nacional, dejando a los estados y comunidades flexibilidad para diseñar sus propias estrategias de mejoramiento.

Brevemente, "Goals 2000" (U.S. Department of Education,1996) son ocho objetivos a nivel nacional establecidos por Ley en el año 1994, producto de un acuerdo entre todos los gobernadores de los dos partidos mayoritarios norteamericanos, después de cinco años de discusiones y de más de una década de intentos de dar respuesta a la "crisis educativa". Pero son mucho más que eso, en la medida en que resumen profundas polémicas en el seno de la sociedad, reflejan el nivel de acuerdos alcanzado en un determinado momento, y siguen siendo el centro de continuas revisiones y tensiones.

Los objetivos nacionales en educación son los siguientes:

Para el año 2000:

1. Todos los niños en Norteamérica comenzarán la escuela listos para aprender.
2. La tasa de graduación de la escuela secundaria aumentará al menos al 90%.
3. Todos los estudiantes terminarán los grados 4to., 8vo. y 12vo. habiendo demostrado competencia en contenidos importantes de las materias académicas.
4. Los estudiantes norteamericanos serán los primeros en el mundo en logros en Matemáticas y Ciencias.
5. Todo adulto norteamericano estará alfabetizado y poseerá el conocimiento y las habilidades necesarias para competir en la economía global y ejercitar los derechos y responsabilidades de la ciudadanía.
6. Todas las escuelas de los Estados Unidos estarán libres de drogas, violencia, y de la presencia no autorizada de armas de fuego y alcohol, y ofrecerán un ambiente disciplinado orientado al aprendizaje.
7. La fuerza docente de la Nación tendrá acceso a programas para el mejoramiento continuo de sus habilidades profesionales y la oportunidad de adquirir el conocimiento y las habilidades necesarias para instruir y preparar a todos los estudiantes norteamericanos para el próximo siglo.
8. Todas las escuelas promoverán la asociación (partnership) con personas y entidades de la comunidad, lo que aumentará la implicación de los padres y la participación en el estímulo al crecimiento social, emocional y académico de los niños.

El contenido del Programa refleja la conciencia de que para mejorar la educación hace falta reforzar financieramente algunas áreas, y también que por un lado se respeta la tradición de toma de decisiones a nivel estatal y local, pero por otro se quiere controlar el uso que se haga de los fondos a través de sus resultados (estándares y rendición de cuentas). Sus principales características son:

1. El estado federal da **financiamiento** a los estados y comunidades para apoyar planes y reformas dirigidas a elevar los logros académicos de los estudiantes.
2. **Flexibilidad**, los estados y los distritos escolares pueden utilizar los fondos para una amplia gama de actividades, incluso ya comenzadas, en función del enfoque que adopten para ayudar a los estudiantes a elevar sus estándares académicos. Los objetivos constituyen un marco nacional.
3. Promueve la **participación** ciudadana a través del consenso entre grupos, individuos e instituciones, para una acción concertada y responsable de educadores, empresarios, organizaciones de padres y líderes políticos para su desarrollo.
4. El mejoramiento de los logros académicos se verifica por medio del cumplimiento de **estándares (Nota 2)** (estatales o nacionales).
5. Reconoce los esfuerzos realizados en el mejoramiento, pero declara que no son

suficientes, en especial para achicar la brecha de rendimiento existente entre los estudiantes blancos y los pertenecientes a minorías. Se dirige a **todos** los estudiantes.

6. Las escuelas deben **rendir cuentas** de sus resultados en función de los objetivos definidos por la comunidad, como parte de su compromiso, y también reciben apoyo para mejorar este aspecto.
7. El **perfeccionamiento docente** es una de las claves para el mejoramiento.

Siempre de acuerdo con el documento analizado (U. S. Department, 1996), "en estos primeros dos años el programa ha provisto recursos críticos para una amplia gama de esfuerzos de mejoramiento de las escuelas en función de elevar los logros académicos" (p. 13), entre los que se destacan tres:

- Construcción de relaciones de cooperación entre escuelas, padres, empresarios, universidad y comunidades para mejorar la educación.
- Mejoramiento de las habilidades de los docentes, la evaluación de los alumnos, el curriculum y la instrucción para ayudar a las escuelas a preparar a todos los estudiantes para alcanzar los estándares.
- Incorporación de tecnología educativa en las escuelas para ayudar a los estudiantes a lograr altos estándares.

2.1. Principales supuestos

Como en todo texto, en éste podemos identificar un nivel explícito de discurso, que es el descripto más arriba, y un nivel implícito, que a pesar de no estar escrito constituye el marco ideológico de los objetivos de política. Los principales supuestos que se deducen del texto son los siguientes:

1) Que en la educación está el origen y la solución para gran parte de los males sociales (ver obj.6). Pensar que la educación puede colaborar en la solución de problemas tales como la droga o la violencia es muy común, y es cierto que la educación puede ser una gran ayuda. Pero depositar todo el peso en la educación además de ser un recurso retórico puede ser un falso desafío, ya que es una responsabilidad de toda la sociedad, pero la trampa está en que si esto no se logra (lo que seguramente suceda si no se toman otras medidas) se podrá acusar a la escuela –pública—de no cumplir con esta tarea a pesar del apoyo brindado. Coincido con Liston y Zeigner (1997) en que las intervenciones educativas no pueden, por sí solas, resolver los problemas de desigualdad en las escuelas. Sin embargo, también se menciona que para que el nivel de educación mejore, las reformas deben contemplar iferentes aspectos, social, pedagógico, financiero, administrativo, político, institucional y tecnológico.

2) Que mayor cumplimiento de estándares es sinónimo de mejores niveles de aprendizaje. Esto implica, por una parte, la consideración sólo de aquellos conocimientos que son observables y mensurables, y por otra parte, la convicción de que los tests estandarizados pueden constituir una prueba fehaciente de dichos aprendizajes. Ambos supuestos son discutibles desde otras perspectivas (Apple, 1996; Bertoni et al., 1996; Gimeno Sacristán, 1994). Para avalar la importancia de los estándares, se afirma que "los educadores han aprendido una lección de los negocios y la industria: la clave del éxito es definir claros, altos estándares de rendimiento y un sistema que mida los resultados en relación con esos estándares" (U. S. Department of Education, 1998, p. 8). Creer que los resultados de la educación son representables por estándares medibles por tests constituye, a mi modo de ver, una visión red ccionista y cuantitativista -en parte herencia del positivismo y en parte de la cultura de la eficiencia y la competitividad- que no toma en cuenta que ninguna medida o cuantificación exacta puede reflejar procesos sociales o individuales complejos como el aprendizaje. A lo sumo lo que pueden reflejar las ponderaciones en las que se basan los tests son algunos conocimientos entre los cuales se encuentran las habilidades para responder a esa clase de pruebas.

Iróni camente, se pone el acento en la claridad y rigurosidad de los estándares que los estados se proponen (U. S. Department, 1994, p. 9 y 10), pero luego se afirma que, con pocas excepciones, los tests actualmente programados no reflejan aún los contenidos de los estándares, vale decir, que no hay formas adecuadas de evaluación (p.11). Resulta más extraño todavía si pensamos que, si es difícil 'medir' los estándares, es prácticamente imposible sin pruebas adecuadas. El motivo principal que los estados han dado para justificar esta situación es "el alto costo de desarrollo de mejores formas de evaluación" (p. 11). No es una justificación muy admisible, ya que existe una línea de fondos para desarrollo de la evaluación, que hasta ahora han usado a 8 estados y un consorcio de 22 estados, pero en general se hallan focalizados hacia los grupos co necesidades especiales, y no a los estándares generales.

Al respecto, Apple (1996) se pregunta si los objetivos y estándares nacionales, con instrumentos de evaluación sistematizados, son mejores que los igualmente difundidos pero más ocultos estándares estatales dados por los libros de texto, que ya ejercen un importante control. En todo caso, es mejor tener parámetros explícitos que ocultos, pero la cuestión que permanece es en quién se deposita la autoridad para establecerlos, con todas las implicaciones que esto tiene. Por otra parte, si bajo ciertas condiciones los estándares nacionales legitiman la desigualdad, esto también depende de los objetivos y el uso que se haga de los resultados de la evaluación.

3) Que estudiantes con elevados estándares van a tener trabajos mejor pagados y que van a potenciar la capacidad económica del país frente a la competencia internacional (U. S. Department, 1994, p. 2)**.** No hay evidencia de la relación directa entre buenas calificaciones en la escuela y buenos salarios en el trabajo o mayor competitividad internacional en donde no intervengan otras dimensiones tales como clase social y capital cultural de los estudiantes, así como otras variables derivadas de la situación económica. Según Spring (1997), estas afirmaciones tienen su punto de partida en el análisis de Robert Reich acerca de la situación del mercado de trabajo, cuyos problemas se deberían a la falta de calificación de la mano de obra. Esta teoría tiene sus antecedentes en los años 60 y 70, con el auge de las teorías del capital humano, que, habiendo sido criticadas y cuestionadas, reaparecen sin mbargo con igual fuerza. La esencia sería que la 'nueva' competitividad laboral tiene como centro el conocimiento.

Este análisis no tiene en cuenta la división interna e internacional del trabajo, la creciente desigualdad social, ni el papel que juega el nivel del empleo en la competitividad salarial. No es la educación sino la economía la que determina si crecen o disminuyen los puestos de trabajo y los salarios. Si bien el mayor nivel educativo mejora sin duda las posibilidades individuales de un mejor empleo, paralelamente se verifica el fenómeno de la inflación educacional, por el cual siempre sigue aumentando la brecha entre los más y menos favorecidos en el sistema.

4) Que el rol del gobierno federal en el mejoramiento de la educación consiste en: (a) **apoyo financiero**, en especial para garantizar el acceso de los estudiantes desaventajados o con discapacidades a todos los niveles educativos; (b) **apoyar a las reformas** lideradas por los estados por medio de **investigación y desarrollo, bancos de datos, y ayuda para la difusión de prácticas efectivas**; y (c) **administrar los programas federales en forma flexible**, de manera que apoyen el liderazgo estatal de las reformas educativas. Se trataría de un rol subsidiario del estado central, si no fuera porque al establecer estándares nacionales se coloca en un lugar principal en el control de los objetivos.

En los Estados Unidos hay una larga tradición de comunidades diversas, con organización propia, que se fueron adicionando sin perder sus peculiaridades culturales, y en este sentido la escuela fue primero la encargada de mantenerlas, y luego de difundir los valores democráticos constitutivos y constructores de la nación. Según afirma Popkewitz (1997), el sistema educativo norteamericano carecía hasta hace muy poco de documentos nacionales significativos sobre objetivos o directrices, tampoco existía un ministerio de educación, y sólo se promulgaban leyes

circunstanciales que servían como orientaciones generales.

La educación resulta así ambiguamente responsabilidad del Estado Nacional, efectiva en la medida en que responde a la necesidad de legitimidad política del sistema y a que las políticas sociales del estado de bienestar incluyeron el apoyo a su sostenimiento. Sin embargo, los "Goals 2000" establecen un marco nacional que está ligado a la existencia de estándares nacionales, y éste es un nudo de conflicto, ya que algunos sectores lo consideran un avance centralizador. Con el advenimiento de las ideas neoconservadoras, se ataca al estado intervencionista que coarta la libertad de las personas. A este ataque parecen responder una serie de cambios hacia una mayor flexibilidad introducidos posteriormente (U. S. Department of Education, 1996, p. 20-23), que consisten específicamente en: 1) el mejoramiento de las operaciones, es decir procesos más sencillos de aprobación de fondos; 2) un nue o enfoque de los programas de administración que apunta a mejorar el funcionamiento y la coordinación al interior del Departamento Federal; 3) un aumento de la flexibilidad a través de excepciones a los requerimientos federales; y 4) la búsqueda de la maximización de la flexibilidad a través de Programas de Demostración (Ed-Flex) en seis estados.

Estas modificaciones podrían responder a las demandas de una mayor participación democrática, si no fuera porque la coincidencia con los reclamos de la derecha religiosa y el auge del neoconservadurismo despierta suspicacias al respecto. Al mismo tiempo, esta "retórica" de la participación, como la llama Popkewitz (1997) hace aparecer las reformas como "un reflejo de las prioridades cambiantes de la comunidad". El autor señala que "la idea de 'comunidad' supone la negociación entre los diversos grupos que detentan igual poder" (p. 232). Esto resulta bastante dudoso cuando pensamos en una asociación que incluye a administradores, maestros, padres, empresarios: no todos tienen el mismo peso en las decisiones, y la diferencia de poder hace que algunos puedan condicionar las decisiones de otros.

Entonces, por un lado encontramos una fuerte presión, tanto de la tradición como del funcionamiento y la ideología política vigente, para el ejercicio de las funciones educativas por parte de los estados y los poderes locales. Por el otro, hay una fuerte corriente más ligada a los funcionarios de educación y a los ambientes académicos liberales, orientada a alguna forma de regulación del estado federal, en parte para contrarrestar los valores locales más conservadores, en parte para mantener sus privilegios de 'establishment', en parte para garantizar la igualdad de oportunidades, y por último, pero no lo último, para asegurar la calidad de la educación. Estas fuerzas aparecen en permanente pugna por inclinar la balanza, y cada cual la percibe alarmantemente inclinada para el lado opuesto. Las tensiones descriptas alrededor de los niveles de gobierno y de los estándares fo man parte de una trama de fuerzas y relaciones políticas más amplia que voy a exponer en el apartado siguiente.

2.2. Contexto Político Ideológico

Lo más llamativo de los Objetivos para el 2000 es que se gestaron durante una presidencia republicana en pleno auge del neoconservadurismo, y tuvieron el acuerdo necesario para continuar con el actual presidente demócrata. El neoconservadurismo encarna la unidad de las ideas económicas del libre mercado—originadas en la escuela austríaca liderada por Hayek, cuyos principales seguidores en los Estados Unidos fueron Rothbard, Simon y Friedman—con la derecha religiosa. Ambas tienen fuertes críticas conservadoras a la burocracia gubernamental, en nombre de la 'libertad': de pensamiento, de mercado, de elección. Sin embargo los neoconservadores también sostienen que el estado reducido debería jugar un papel activo en la protección de la moral pública y en la calidad de la enseñanza a través de la censura y los estándares académicos (Spring, 1997).

A pesar de que los neoliberales enfatizan los valores del mercado, y los neoconservadores los valores tradicionales, Apple (1996) afirma que ambos responsabilizan a las escuelas por la mayor parte de los problemas de la sociedad.

Para este autor, dicha alianza "combina los negocios con la Nueva Derecha y con los intelectuales conservadores" (P. 27). Las principales propuestas de este "bloque de poder" incluyen la implementación de programas de elección de escuela como bonos o créditos en los impuestos, el intento de establecer estándares de excelencia; el ataque a las escuelas públicas porque no encarnan los tradicionales valores occidentales; y el reclamo de incluir las necesidades de las empresas en los objetivos de la educación.

Labaree (1997) puntualiza que la presidencia de Reagan representó "un cambio significativo en el discurso de la ingeniería social, ya que los remedios estatales para los problemas sociales perdieron terreno frente a los remedios basados en el mercado" (p. 137). La ideología del gobierno de Reagan afirmaba que la burocracia estatal era inefectiva e ineficiente en el manejo de los servicios sociales, y que el sector privado era más capaz de cumplir ese rol. En 1983, la aparición del informe *Nation at Risk* marcó el comienzo de un poderoso ataque contra la escuela pública e inició la campaña a favor del modelo de mercado, en el que la libertad de elección y la competencia aparecen como los principales valores para la educación.

Consecuentes con ese marco, en 1994 los republicanos redactaron el llamado *Contrato con América*, donde se corporizaban los cinco principios básicos de la 'civilización americana' (Spring,1997, p. 15):

1. Libertad individual.
2. Oportunidad económica.
3. Gobierno limitado.
4. Responsabilidad personal.
5. Seguridad en el hogar y afuera.

Nada relacionado con otros valores democráticos tales como equidad, justicia y responsabilidad social afecta la primacía de los valores individualistas y económicos en estos principios. ¿Cómo se explica entonces, políticamente, la dimensión y continuidad de un programa como *Goals 2000*, dirigido a mejorar las posibilidades de todos los estudiantes? El análisis de Spring (1997) sugiere que lo que sucedió es un "aggiornamiento" de los Nuevos Demócratas para ganar votos en la clase media blanca, que no se sentía representada por un partido que defendía principalmente a los pobres, las minorías raciales, los homosexuales y los pacifistas. Y que también ocurre una división entre los republicanos a causa del fundamentalismo de la derecha religiosa, debido a lo cual muchos se acercaron al Partido Demócrata.

La interpretación de Apple (1996) es que "grupos poderosos del gobierno y la economía, y de los movimientos sociales populistas autoritarios, han sido capaces de redefinir –a menudo de modo regresivo, los términos del debate en educación, estado de bienestar y otras áreas del bien común" (p. 27). Esta lectura parece tener más base, especialmente si pensamos en la similitud que guarda con algunos procesos ideológicos en América Latina, como la pérdida de legitimidad de las instituciones públicas y la instalación del sentido común de que lo privado es mejor.

Parece evidente que la educación es un tema convocante a nivel nacional, o tal vez es el tema elegido por los dos partidos principales para gestar la unidad, frente a una situación considerada "grave". La crisis de la educación norteamericana, o diría mejor de la escuela pública, que viene agitándose con discursos alarmantes especialmente desde la publicación del informe *Nation at Risk* en 1983, es un lugar común, tanto como la búsqueda de soluciones expresadas en diversas propuestas de reforma y tensas polémicas. Popkewitz (1997) expresa elocuentemente el tono catastrófico- patriótico de ése y otros documentos similares, dirigidos sobre todo al ciudadano común: "En lugar de análisis, estos informes ofrecen exhortaciones y profecías. Su lenguaje lamenta la pérdida de gracia de la nación y promueve la rectitud de acción como medio a través del cual es p sible la redención" (p. 166). No hace falta aclarar que la herramienta de esta redención es la escuela. Y tal vez por eso se constituya en el centro de intensas disputas.

Pero podemos considerar otras interpretaciones, como la de Berliner y Biddle (1995), que afirman que la "crisis manufacturada" no es un hecho accidental, sino que "más bien aparece en un contexto histórico específico y fue agitada por críticos identificables, cuyos objetivos políticos podrían ser promovidos usando a los educadores como chivo expiatorio (p. 4)." Por su parte Lind, citado por Spring (1997), se pregunta si esta 'guerra' contra la escuela pública no es una cortina de humo del ala derecha republicana para ocultar la crisis económica y el crecimiento de las desigualdades. En una postura similar, Liston y Zeigner (1997) afirman que "la tan aireada crisis de las escuelas de los Estados Unidos es, en realidad, el reflejo de la crisis general del conjunto de la sociedad" (p. 21).

Al mismo tiempo, encontramos el deseo moralizador de la derecha religiosa, que sostiene que "la solución a los problemas públicos fue la enseñanza de la moralidad y de los valores de la cultura occidental" (Spring, 1997, p. 5). En esta perspectiva, pornografía y comunismo van de la mano con el humanismo secular y la educación sexual, y una de las formas de contrarrestar esta decadencia es, por ejemplo, su propuesta de imponer el rezo en las escuelas públicas, lo cual a su vez cuenta con fuerte soporte económico de algunas empresas y fundaciones, permitiendo promocionar a ciertos candidatos que lideran la difusión de estas ideas.

Por eso esta guerra parece tener distintos frentes y objetivos, ya que se involucran partidos políticos, grupos religiosos y económicos, los Departamentos de Educación Federal y estatales, las fundaciones, los sindicatos, los distritos escolares y gobernadores, etc. en una lucha permanente por el control sobre los contenidos (estándares y curriculum), los textos y la elección de escuelas, principalmente. Sin embargo, Arons (1997) considera que es una pelea mucho más polarizada:

Para fines de 1994 se estaba volviendo menos claro si la estandarización escolar que había sido sostenida por la administración Bush, los gobernadores nacionales, y la administración Clinton, podía ser usada por los conservadores como una herramienta para disputar el control de la escuela pública al 'establishment' o si sería un medio del 'establishment' para resistir la influencia de los conservadores en muchos consejos escolares locales (p. 25).

Es que desde la derecha religiosa, el principal factor disolvente se encuentra en la élite liberal (política y cultural) y la burocracia federal, que promueven políticas nacionales seculares y antisegregacionistas, que no corresponden a iniciativas de las comunidades. Pero en realidad, "la oposición a la burocracia educacional y gubernamental y el deseo de devolver el poder a la gente se basa en la creencia de que esto restaurará los valores tradicionales de la educación." (Spring,1997, p. 15).

Como dije anteriormente, la pelea por el control del sistema educativo se convierte más bien en la lucha por el control de lo que piensa la gente. En palabras de Apple (1996) "la decisión de definir el conocimiento de ciertos grupos como el más legítimo, como el conocimiento oficial, mientras el conocimiento de otros grupos difícilmente llega a ver la luz del día, dice algo sumamente importante acerca de quién tiene poder en la sociedad" (p. 22). Arons (1997) advierte sobre los riesgos de esta situación en la que cada sector busca imponer su ideología, porque los que fracasen en su intento de control pueden llegar a tratar de debilitar la influencia del sistema educativo.

Así, el espectro se complica, mientras "la última cosa que la derecha religiosa quiere es un curriculum controlado por estándares estatales o nacionales" (Spring, 1997, p. 48), desde otro enfoque ideológico, tanto para Arons (1997) como para Apple (1996) este mayor control gubernamental representa una amenaza para la democracia, llegando el segundo a afirmar que el aumento del centralismo conduce directamente a la privatización.

En el acuerdo de los *Goals 2000* parecen haberse plasmado las ideas de la derecha y los demócratas radicales con respecto a la libertad local y a la menor intervención del Estado Nacional, las ideas republicanas de estándares nacionales para un mayor

control de calidad y competitividad, y las ideas demócratas de que todos tienen que tener las mismas oportunidades y de que el gobierno federal es el responsable de garantizarlo, sosteniendo programas compensatorios. Sin embargo, en las enmiendas realizadas al Programa en 1996 se observan cambios en la dirección de una mayor flexibilidad hacia los estados y agencias locales, eliminando regulaciones y estableciendo un uso más permisivo de los fondos. Estas modificaciones pueden leerse dentro del marco de la presión que ejercen los conservadores en su defensa de los poderes locales, bajo el reclamo de mayor libertad.

Este es el contexto en el que también se desarrolla el movimiento por la elección de escuela, que constituye una de las propuestas de reforma más importantes en los Estados Unidos, que dentro de este marco de lucha política por el control de la educación es defendida tanto por sectores conservadores como progresistas, obligándonos, una vez más, a intentar definir contextualmente los términos de esta lucha por los significados de los discursos y prácticas en educación.

3.1. El movimiento por la elección de escuela ("school choice")

El tema de la elección de escuelas por parte de los padres es uno de los más controvertidos en la educación norteamericana actual. Una de las razones por las cuales el movimiento está creciendo fuertemente es que sus fundamentos son defendidos desde grupos muy diversos en el espectro ideológico. Dentro de este movimiento encontramos también diferentes posiciones que van desde la defensa de la elección sólo entre escuelas públicas, como un medio para mejorar la educación pública, pasando por los que quieren ampliarla también a las privadas no confesionales, hasta los que postulan la libre elección de escuela pública o privada (Nota 3) sin distinción, lo cual trae fuertes polémicas acerca de la separación constitucional entre la Iglesia y el Estado. Los que se oponen ven estas posturas como un ataque a la educación pública, y una válvula de escape para evitar una auténtica reforma del sistema. Los conceptos de democracia, libertad, competencia, mercado y desigualdad social están en la base de estas discusiones.

El reclamo básico de la elección de escuela es que las escuelas deben tener más autonomía para trabajar mejor, y los padres mejores opciones que la escuela pública de la zona, a menudo percibida como deficiente. Entre las principales propuestas que impulsan los partidarios de la elección de escuela por parte de los padres encontramos: escuelas charter, programas de bonos, escuelas imán y escuela en el hogar.

Sobre las escuelas imán diré brevemente que fueron las primeras propuestas para optar en la escuela pública dentro del distrito. Fueron introducidas en los años sesenta como programas alternativos en general menos tradicionales que las otras escuelas y en los setenta constituyeron uno de los instrumentos para las políticas de desegregación, especialmente en los barrios marginales. Los resultados fueron dispares en cuanto a ese objetivo, pero aparentemente positivos con respecto al mejoramiento de los aprendizajes (Murphy J., Gilmer, S., Weise, R., y Page, A., 1998). No me voy a extender más aquí, primero, porque la experiencia no es actualmente relevante para la discusión sobre elección y privatización, y segundo, porque merece un tratamiento más profundo, quizás en un próximo estudio.

En este trabajo me voy a referir especialmente a las propuestas de escuelas con contrato (charter) y de bonos escolares (vouchers) porque son las que se están difundiendo con cierta rapidez también en América Latina. Luego de describir y analizar sus características y modos de implementación, retomaré la discusión sobre los diferentes grupos identificables en la lucha política por redefinir la educación pública.

3.1. Las escuelas con contrato (charter school)

Las escuelas charter constituyen uno de los fenómenos más dinámicos de la actual reforma educativa norteamericana. Como dije más arriba, forman parte del creciente

movimiento de elección de escuela, pero con particulares características que las diferencian de las otras propuestas y determinan su enorme difusión, ya que son vistas como una alternativa a la escuela pública tradicional.

Las escuelas charter son escuelas que firman un contrato con el estado o el distrito por el cual reciben excepciones con respecto a ciertas normas generales, y fondos del gobierno para poder cumplir con los objetivos establecidos en dicho contrato (charter). Cada estado determina por ley qué características y duración podrán tener los contratos en su territorio, y los requisitos para poder ser renovados. Debido a que hay una gran variedad de leyes y a que el contrato es específico en cada caso, encontramos una enorme diversidad en las características de estas escuelas entre un estado y otro, y aún dentro del mismo estado. Esta flexibilidad hace que las escuelas charter tengan características tanto de escuelas públicas como privadas. Me propongo analizar el carácter público de estas instituciones y explorar si proveen mejor educación que las escuelas públicas comunes, así como qué grupos y con qué di cursos político ideológicos establecen la agenda de esta polémica.

3.1.1. El carácter público de las escuelas charter

En esta sección trato de establecer si estas escuelas siguen siendo públicas y qué las diferencia de las privadas. ¿Las escuelas charter son consideradas públicas debido al origen de sus fondos, a que deben rendir cuentas al estado, o a que deben ser abiertas a todos los estudiantes sin distinción ni requisitos de admisión? Un punto de partida útil puede ser la definición de escuelas charter por parte de un ente oficial. La Oficina de Investigación y Mejoramiento Educational del Departamento Federal de Educación (U. S. Department of Education, 1998) caracteriza las escuelas charter como escuelas públicas, lo que las hace diferentes es su contrato –un contrato con el estado o con el distrito. Cada contrato establece lo que la escuela planea hacer para alcanzar los objetivos educacionales; en función de éstos la escuela recibe fondos públicos por un período determinado. El contrato libera a los titulares de la escuela de regulaciones que de otra manera se aplican a todas las escuelas (p. 1).

Esta definición enfatiza los elementos financieros y estructurales de las escuelas charter. Sin embargo, no satisface la pregunta acerca de cuáles son las reales diferencias entre éstas y las escuelas privadas en los Estados Unidos. Esta distinción es importante porque la escuela pública es abierta a todos, y la educación común es esencial para que los niños aprendan, junto con los demás contenidos y actividades, el respeto a los derechos, la tolerancia hacia las diferencias y la aceptación de los otros, elementos fundamentales para la vida social en una democracia. Las dimensiones que voy a explorar en función de clarificar las diferencias con las escuelas privadas son: (1) financiamiento; (2) posibilidad de elección; (3) control; (4) status legal; y (5) accesibilidad.

1) Financiamiento. Es una de las dos principales dimensiones que utiliza un documento del Centro Nacional de Estadística Educacional (NCES) para determinar el carácter público de una escuela (U. S. Department of Education, 1997). Todas las escuelas charter reciben fondos públicos, pero el manejo operativo no es comparable con el de las demás escuelas ya que en muchos casos las normas que los rigen difiere básicamente. Lo usual es que reciban el mismo monto por alumno que recibe cualquier escuela del distrito, y esto pone en desventaja a los grupos promotores que no tienen el capital suficiente para la infraestructura y elementos necesarios para comenzar. En síntesis, es más fácil el punto de partida para los que ya administran una escuela pública, o para las corporaciones privadas. Con respecto a los alumnos, si la escuela no está en la zona correspondiente a su domicilio en muchos casos deben asumir el gasto de transporte, qu de otro modo es gratis.

2) Elección. La posibilidad de elección de escuela por parte de los padres es la segunda dimensión fundamental utilizada por el mencionado documento (U. S. Department of Education, 1997). El estudio afirma que ésta "ha estado asociada tradicionalmente con las escuelas privadas" (p. 3). Sin embargo, los investigadores encontraron que en 1993 al menos 11% de los estudiantes que cursaban los grados 3ro. a 12vo. concurrían a escuelas públicas elegidas por sus padres debido a algún

tipo de influencias, y los padres del 39% de los estudiantes podían elegir escuela mediante la elección del barrio donde vivían. Solamente el 41% (menos de la mitad) de los estudiantes concurrían a escuelas públicas asignadas sobre la cual sus padres no habían ejercido ninguna elección directa ni indirecta. Sorprendentemente, esta segunda característica no es suficiente para diferenciar escuelas públicas de pri adas porque más de la mitad de los padres pueden elegir. Además, el estudio muestra que las familias con mayores ingresos tienen más posibilidades de elección, lo cual coincide con las conclusiones de otros estudios realizados en varios países desarrollados, que indica que la elección de escuela puede incrementar la estratificación social (Gewirtz, Ball, y Bowe, 1995; Lauder y Hughes, y Watson, Waslander, Thrupp, Strathdee, Simiyu, Dupuis, McGlinn, y Hamlin, 1999; Patrinos y Ariasingam, 1998; Whitty, Power, y Halpin, 1998). Mientras que los defensores de la elección afirman que ésta busca expandir las oportunidades de las familias pobres, los estudios indican que este sistema beneficia a los grupos privilegiados de la sociedad, que están en mejor posibilidad de elegir.

3) Control. Las escuelas charter son evaluadas por sus resultados, esto significa que están sujetas a una rendición de cuentas (accountability). El proceso educativo es complejo y requiere de un análisis cualitativo y cuantitativo para entender sus resultados. Una evaluación reciente de las escuelas charter en Los Angeles ha desarrollado un enfoque comprensivo que reconoce que "los resultados y éxitos de cada escuela pueden ser atribuidos a diversos factores" (WestEd, 1998, p. 3). El informe del segundo año de la investigación mencionada anteriormente (U. S. Department of Education, 1998) expresa: "Las escuelas tienen autonomía con respecto a las normas a cambio de que rindan cuentas por sus resultados", pero "los estatutos que regulan los contratos difieren de estado a estado tanto como la amplitud y naturaleza de la autonomía que permiten" (p. 9). En todos los casos alguna autoridad educativa (por ejemplo, el consejo de distrito, el consejo del estado, otra institución estatal o más de una) puede ser garante del contrato. Sin embargo, las leyes definen contratos y excepcionalidades muy diferentes a lo largo del pais. Por ejemplo, en New Mexico, la ley impone a las escuelas charter prácticamente las mismas regulaciones que a las escuelas públicas tradicionales, en tanto en la mayoría de los demás estados las leyes autorizan a las escuelas charter excepciones automáticas con respecto a los códigos comunes. Por lo tanto, podríamos preguntarnos cómo cada estado desarrolla el proceso de evaluación de sus escuelas charter –qué parámetros y qué clase de indicadores de éxito utilizan, tales como rendimiento de los alumnos, satisfacción de la comunidad, administración de los fondos, etc., en especial si tenemos en cuenta las dificultades mencionadas más arriba para evaluar los resultados de los programas desarrollados en función de *Goals 2000*.

4) Status legal. En muchos estados las escuelas charter son entidades legales independientes que pueden seleccionar y/o negociar con su personal (U. S. Department of Education, 1998, p. 111). De acuerdo con el citado documento, hay dos tendencias que están emergiendo en lo que respecta a legislación: estados con leyes más antiguas están ampliando los límites en el número de charter que permiten, mientras que otros estados están flexibilizando los procesos de autorización. A pesar de que no sabemos aún qué aspectos incluye exactamente esta flexibilización, un estudio de Morando Rhim (1998) advierte que "hay un fuerte interés en el sector privado en capitalizar el mercado educativo" (p. 51). Este trabajo también alerta sobre los procesos de subcontratación que han comenzado en el sistema público (por ejemplo en Massachusetts), "el crecimiento de las escuelas charter y el crecimiento si ultáneo de la administración privada de charters durante los últimos tres años, representa una aparente 'segunda ola' o renovación del sistema de privatización escolar" (p. 47). La permisividad, ambigüedad o inespecificidad de las leyes ponen a las escuelas en situación más vulnerable frente a los intereses de las corporaciones educativas privadas.

Un ejemplo es el caso del Proyecto Edison (Nota 4), respecto del cual Spring (1997) afirma que como la batalla por los bonos no fue muy exitosa para los conservadores, ellos encontraron una buena opción en las escuelas charter, financiadas por el

gobierno. Los estados de Colorado y Massachusetts "dejaron pasar leyes de escuelas charter que permiten a los estados y al sistema educativo firmar contratos con . . . contratistas privados (por ejemplo, el Proyecto Edison). Con las escuelas charter, la operación de escuelas con fines de lucro es una posibilidad" (p. 63). Como Spring (1997) y Morando Rhim (1998) sugieren, la inclusión de instancias de lucro en las instituciones públicas, las hace en parte privadas. Este es solamente el principio, pero si las corporaciones pueden hacer buenos negocios, seguramente tratarán de aumentar sus beneficios a expensas de los contribuyentes.

5) Accesibilidad. Las escuelas charter son públicas porque son abiertas a todos los estudiantes, son gratuitas, no pueden poner requisitos de admisión y, al menos en teoría, reciben población social y étnicamente diversa. Como las demás escuelas públicas reciben mayor impacto de los problemas sociales relacionados con la pobreza, la violencia y el abuso de alcohol y drogas que las escuelas privadas (U. S. Department of Education, 1997).

En síntesis, las escuelas charter son públicas porque reciben financiamiento estatal, porque están abiertas a todos los estudiantes, y porque su funcionamiento está bajo el control del estado, que puede no renovar su contrato si no demuestran haber cumplido con los objetivos propuestos en él. Sin embargo, en la práctica, la flexibilidad, la ambigüedad y la permisividad de las leyes, el creciente interés de los sectores privados en el negocio de la educación, y la presión política a favor de los modelos de mercado, hace necesario contextualizar y definir mejor los objetivos públicos de la educación, para dar un sentido democrático a las escuelas charter.

Algunas experiencias modelo han sido valiosas al mostrar las posibilidades que una mayor flexibilidad puede brindar a las escuelas que trabajan con población marginal. Meier (1995) afirma que la principal diferencia entre una escuela y otra es el status social y económico de los estudiantes, no su carácter de pública o privada. Sin embargo, los cambios en Central Park East (Nota 5) fueron posibles porque tuvieron el apoyo del sistema educativo público; los cambios no fueron en una sola escuela sino en todo el distrito, aún cuando los resultados en cada escuela fueron diferentes. Meier cree que aunque la elección de escuela ha sido defendida por los enemigos de la escuela pública, ésta "es una herramienta esencial para salvarla" (p. 97), porque "la alternativa a la privatización es buena educación pública" (p. 103).

Este ataque a la escuela pública desde muchos defensores de la elección hace difícil pensar que no es una forma de sacar ventaja de la crisis de legitimidad de las instituciones públicas y la burocracia estatal (Anderson, 1998). Una de las alternativas más probables consiste en convertir a las escuelas charter en el primer paso para crear el sentido común favorable a la privatización a través de los bonos, u otras estrategias. El siguiente artículo periodístico parece significativo en ese sentido, en particular porque Shokraii es una de las analistas de The Heritage Foundation, de la cual hablaré más adelante:

Los que ven la elección de escuela de más cerca ven las escuelas charter como la reforma educativa con más probabilidades de éxito. Esto es en parte porque los planes de charter usan dinero público para sostener escuelas privadas, pero no amenazan a las escuelas públicas de la nación o a los grupos que están detrás, como los sindicatos docentes. Por esas razones, Shokraii-Rees afirmó, "Es mucho más fácil venderles las escuelas charter a los políticos y al público . . . . dan a los padres una experiencia de elección de escuela que no habían tenido antes . . . . ayudan a legitimar programas de elección de escuela más ambiciosos." (Bray Duff, 1999)

Los conservadores quieren "vender" las escuelas charter a los políticos y al público porque parecen ser menos conflictivas para los sindicatos y grupos que defienden la escuela pública. Sin embargo, el objetivo real parecería ser legitimar bonos y otras estrategias que significan financiamiento público para escuelas privadas.

3.1.2. Las escuelas charter ¿son mejores?

El Estudio Nacional de Escuelas Charter (U. S. Department of Education, 1998) es

el programa de investigación más amplio realizado sobre estas escuelas. El Informe del Segundo Año presenta información del año escolar 1996-1997 (Nota 6) sobre escuelas charter. A pesar de los cambios ocurridos desde esa fecha, y de algunas limitaciones metodológicas, ésta es la más completa fuente de información sobre la implementación de charter. De acuerdo con el informe, algunas de las principales características de este fenómeno hasta 1997 eran las siguientes:

- Su número crece rápidamente.
- La mayoría de los contratos fueron renovados.
- La mayor parte de las escuelas charter son más pequeñas que las escuelas públicas comunes.
- Muchas de ellas son escuelas nuevas, creadas como charter.
- En muchos estados hay una proporción más alta de escuelas charter que reciben predominantemente alumnos de color, en tanto en otros estados reciben similar o algo más alta proporción de estudiantes blancos.
- La mayoría de las escuelas charter tienen una composición étnica y socioeconómica similar al resto del distrito, pero alrededor de un tercio tienen una mayor proporción de estudiantes de color y/o pobres.

Los elementos de esta información que quiero destacar son: (1) expansión, (2) tamaño, y (3) distribución étnica y socioeconómica de la matrícula.

1) Expansión. Desde 1993 se han multiplicado las escuelas charter y el número de estados que han sancionado leyes de charter. Algunas de las causas se relacionan con la disminución de la credibilidad del estado y sus instituciones en relación con la democracia representativa, pero hay cierta específica y profunda "crisis" del sistema educativo público que está en el centro del debate. El diagnóstico y la solución dependen de quién define la crisis. Republicanos y conservadores consideran que la escuela pública ha fallado completamente porque los estudiantes no tienen un rendimiento competitivo, y por eso proponen la elección de escuela. Para ellos esto significa el modelo de libre mercado para mejorar la educación (Schokraii, 1998), y tienen importante apoyo político y monetario para desarrollar y difundir sus ideas (Spring, 1997; National Committee for Responsive Philantropy, 1999). Intelectuale críticos y liberales (Berliner y Biddle, 1995; Spring, 1997; Liston y Zeichner, 1996) interpretan la supuesta crisis de la educación como un modo de ocultar la crisis política y económica actual y no enfrentar los reales problemas de las escuelas. Su respuesta es mejorar la equidad social y la educación pública.

2) Tamaño. La mayor parte de las escuelas charter son más chicas que las escuelas públicas tradicionales, más del 60% tienen menos de 200 alumnos, mientras sólo alrededor del 16% de las otras escuelas públicas tienen ese número (U. S. Department of Education, 1998). Dada la amplia variedad de programas y contratos, quién podría asegurar que su "éxito"—de acuerdo con los indicadores de satisfacción de los padres—se debe a que los programas responden a las necesidades de la comunidad, y no a que las escuelas más pequeñas tienen mayores posibilidades—porque tienen más tiempo y relaciones más personalizadas con los alumnos—de enseñar mejor (Waymack y Drury, 1999).

En el caso de las escuela de Meier (1995), todo el distrito había reducido el tamaño de las escuelas. Esto no significa que este puede ser la única y mágica estrategia de cambio, pero si brinda la oportunidad de que los docentes conozcan mejor a sus alumnos y sus padres, y a los administradores la posibilidad de realizar otro tipo de manejo, de mejorar la calidad de las relaciones y realizar más enriquecedoras experiencias, ¿por qué no reducir el tamaño de todas las escuelas? La respuesta podría ser, en primer lugar, porque el criterio que ha prevalecido para las escuelas públicas –con excepción de las charter-- ha sido el de uso eficiente de los recursos; en segundo lugar, porque el resultado sería el aumento del presupuesto educativo estatal y distrital, lo cual resulta contradictorio con la lógica hegemónica del mercado.

3) Distribución étnica y socioeconómica de la matrícula. El mismo informe

mencionado más arriba (U. S. Department of Education, 1998, p. 72) expresa que "la mayoría de las escuelas charter tiene características demográficas similares a otras escuelas públicas, excepto que una de tres está enfocada a minorías o estudiantes económicamente desaventajados". La distribución parece asemejarse a las otras escuelas públicas, pero al menos un tercio de las escuelas charter atiende a estudiantes pobres o pertenecientes a minorías étnicas. Las tablas 1 y 2 muestran la comparación entre la distribución étnica en todas las escuelas públicas y en las escuelas charter, en general y en algunos estados.

Tabla 1
Porcentajes de matrícula estimados en escuelas charter (1996-97) y en todas las escuelas públicas y en los 16 estados con leyes de charter (1994-95), por categoría étnica/racial.

Categorías raciales	Escuelas charter de la muestra	Escuelas públicas en los estados con ley charter
Blanco, no de origen hispano	58,1%	59,9%
Negro, no de origen hispano	16,8	14,6
Hispano	16,3	19,5
Asiático o de las Islas del Pacífico	3,1	3,9
Indio Americano o Nativo de Alaska	5,3	2,1
Otros	0,4	NA

Tabla 2
Promedio estimado de porcentajes de estudiantes blancos matriculados en escuelas charter (1996-97) y escuelas públicas en algunos estados.

Estados	Todas las escuelas públicas %	Escuelas charter %
Wisconsin	87,07%	78,7%
Massachusetts	80,06	69,8
Michigan	79,04	54,8
Minnesota	85,9	52,8
Texas	50,4	17,6
California	46,4	56,5
Colorado	74,3	83,5
Arizona	56,1	58,4

Fuente: elaboración propia de U.S. Department of Education, 1998, p. 48-49.

Aún cuando en la distribución general el promedio no varía demasiado, en la comparación por estado es evidente la mayor diferencia en algunos de ellos. Si bien

la orientación hacia estudiantes menos favorecidos puede significar mejorar sus posibilidades, al mismo tiempo puede constituir una nueva forma de segregación, en muchos casos voluntaria, por parte de familias que no sienten que las escuelas oficiales están respondiendo a las necesidades de sus hijos.

Los datos que muestra un trabajo de evaluación de escuelas charter en Los Angeles (WestEd, 1998) son similares a los del informe del Departamento de Estado. Por ejemplo, observan que "las tres escuelas charter independientes tienen mayoría de población hispana . . . . las dos escuelas dependientes, por el contrario, tienen menos del 20% de hispanos . . . . cuatro de las cinco escuelas con contrato renovado muestran un creciente porcentaje de estudiantes que tienen competencia limitada en el idioma inglés. En dos de las escuelas, el porcentaje de estudiantes económicamente desaventajado ha crecido" (p. 12). La información parece indicar que cada escuela tiende a trabajar con grupos más homogéneos, acrecentando sus grupos mayoritarios (blanco, negro o hispano) y reduciendo la integración. Estos resultados son consistentes con la investigación de Nathan (1996), si bien él lo interpreta como una expansión de las oportuni ades para los alumnos pobres o pertenecientes a minorías étnicas.

Por otra parte, un estudio pormenorizado de las escuelas charter de Arizona (Cobb y Glass, 1999) descubre lo que otras evaluaciones no muestran. Los autores afirman que la pregunta clave no es qué porcentaje de cada grupo étnico hay en las escuelas charter sino cómo estos grupos están distribuidos entre las escuelas charter y las públicas comunes cercanas o de la misma zona. Utilizando una metodología que resulta muy reveladora, el mapeo digital, junto con información censal y datos de matrícula, realizan una serie de comparaciones contextualizadas entre escuelas públicas comunes y charter dentro de diferentes zonas rurales y urbanas, encontrando que más de la mitad de las escuelas charter muestran un grado importante de segregación. No solamente las charter registran un porcentaje mucho más alto de estudiantes "blancos," sino que aquéllas que atienden a mayoría de estudiantes pertenecientes a minorías suelen ser escuelas vocacionales, es decir que no preparan para la universidad, o escuelas que reciben alumnos expulsados por el sistema tradicional.

Basados en este análisis, ¿podríamos afirmar que las charter son mejores escuelas? Quizás lo son en muchos sentidos: relación más personalizada con alumnos y padres y mejor aprendizaje en muchas de ellas, pero existen otros esfuerzos de la misma clase en escuelas ignoradas. Sin embargo, el riesgo—y en muchos casos la realidad—de una mayor segregación es alto, y hasta ahora, las evidencias no aseguran que el éxito de las escuelas charter está más relacionado con mejores resultados de los alumnos en los tests (contra el discurso que afirma esto) que con estrategias políticas e insatisfacción de los padres con las escuelas públicas.

La evaluación de WestEd (1998) sugiere algunos factores que pueden ser significativos para el mejor funcionamiento de muchas escuelas charter: experiencias previas de reforma impulsadas desde el distrito o desde el estado, tiempo y trabajo extra aportado por directores y maestros, políticas para captar alumnos, y una amplia variedad de programas dirigidos a alentar la participación de los padres (por ejemplo, cursos sobre cómo ayudar a los estudiantes en la casa, sobre cómo participar en las reuniones y asambleas de la escuela, o relacionados con el trabajo). Que los padres que "eligieron" escuela charter estén más satisfechos (U. S. Department of Education, 1997) no necesariamente indica que esto se debe a mejores resultados, sino que la mayor atención, las relaciones y la autoafirmación jugan un papel en estos sentimientos. ¿Las escuelas charter son necesarias para cambiar la educación? Tal vez no. ¿Por qué no p omover que todas las escuelas públicas desarrollen sus propias innovaciones, dándoles todo el apoyo y la flexibilidad necesaria para mejorar?

En palabras de Sarason (1996, p. 379), los educadores "son reactores, no proactores". Para él, "bonos, escuelas charter y privatización son algunos de los indicadores de que el público (incluyendo al establishment político) está deseando

tomar nuevas direcciones como nunca antes . . . . Los educadores expresan su desaprobación hacia estas propuestas, pero no nos dicen qué proponen y aprueban ellos". Sin embargo, no es seguro que las autoridades escuchen a los docentes cuando estas propuestas emergen, cuando lo importante parece ser la agenda neoliberal-neoconservadora que incluye a la educación como la próxima conquista del libre mercado.

3.1.3. Intereses en juego

La Asociación de Padres y Maestros (PTA) y la Federación Norteamericana de Docentes sostiene que sólo las escuelas públicas deben usar fondos del estado, y defienden algunos principios que aseguran que las escuelas charter sigan siendo públicas. Con relación a las escuelas charter, PTA (1998) considera que las mismas constituyen sólo una vía para la reforma escolar, pero que deben seguir ciertos principios a fin de mantener intacta la integridad de las escuelas públicas. La Federación Norteamericana de Docentes ha declarado que las escuelas charter pueden representar una gran oportunidad para el mejoramiento, pero sus análisis (1996) indican que hay ciertos riesgos relacionados con la falta de requerimientos de los estados con respecto a estándares, la falta de claridad en cuanto a cuotas y donaciones, certificación de los docentes, y sistemas de información. En función de esto realizan ciertas recomen aciones hacia los políticos para las próximas leyes o modificaciones de las actuales. Ninguna de estas dos importantes asociaciones analiza en sus documentos los intereses políticos que están en juego ni las consecuencias sociales de la elección de escuela en el contexto actual.

Determinadas escuelas charter pueden ser interesantes experiencias que ayuden a estudiantes desaventajados a tener una mejor oportunidad, pero en una perspectiva más general, son parte de la lucha de los conservadores contra o por el control del sistema educativo público, en la cual, bajo el reclamo de mayor libertad para los padres, promueven que la competencia por la educación sea regulada por las "fuerzas del mercado". Por ejemplo, la Heritage Foundation tiene 254 documentos sobre escuelas charter en su página web. Su presidente, Edwin Feulner (The Heritage Foundation, 1998a), pregunta "¿Cuán malas son nuestras escuelas públicas?" en un comentario de una página en el que se declara "fanático de las escuelas charter". Su posición se basa en originales definiciones como por ejemplo que las escuelas charter en Houston usan "instrucción directa, un método tradicional de enseñanza dirigido a las habilidades básicas de lectura, escritura y matemáticas"; sus administradores "pueden incluso ignorar los estúpidos requerimientos de certificación"; y "ellas pueden rescatar al menos algunos niños del decadente sistema educacional norteamericano." Este, claramente, no es un comentario académico sino político, y su tono es altamente provocativo. Feulner denigra las escuelas públicas y la certificación de los docentes, reclamando los métodos tradicionales de instrucción que enseñan "habilidades básicas", y desconociendo el importante desarrollo de la práctica y la investigación pedagógica.

Al mismo tiempo, los conservadores consideran que las escuelas charter sirven mejor a la "sociedad civil que los monopólicos/compulsivos organismos de gobierno" (Spring, 1997, p. 63). Sin embargo, el resultado de la asistencia obligatoria universal que ellos critican "es un sistema que se ha vuelto paulatinamente más inclusivo e igualitario" (Labaree, 1997, p. 6). O quizás esta sea la razón de su ataque, porque la idea de educación que estos grupos tienen está más ligada a mantener los privilegios de las élites y a enseñar a las mayorías lo básico, en relación con el lugar que cada uno más probablemente ocupará en la economía. La creación de más escuelas especiales de cualquier clase continúa erosionando el principio de equidad y universalidad de la educación.

3.2. Los programas de bonos (vouchers)

El financiamiento público de las escuelas privadas es uno de los temas más polémicos en la educación norteamericana actual, ya que incluye argumentos filosóficos y políticos acerca de la escuela pública, la relación entre la iglesia y el estado, el valor y la posibilidad de elección, y el acceso universal a la educación. El

sistema de bonos, o elección de escuelas privadas, permite a los padres usar fondos públicos para enviar a sus hijos a escuelas privadas. En la mayoría de los casos cada estudiante recibe una suma equivalente a lo que el estado hubiera gastado en el sistema público.

El sistema de bonos es una forma de mercantilización de la educación, es decir que la educación puede ser intercambiada como cualquier otro bien en el mercado capitalista, el cual, teóricamente, está regulado por la oferta y la demanda creada por productores y consumidores libres e iguales, pero en la práctica, la libertad de consumir es tan amplia como el límite de una tarjeta de crédito o del efectivo disponible.

Mercantilizar la educación significa abandonar una larga tradición de educación pública, que en los Estados Unidos se ha caracterizado por financiamiento estatal, asistencia obligatoria, acceso universal, gratuidad, provisión de transporte en la zona, edificios, docentes, materiales, libros y servicios compensatorios (por ejemplo almuerzo o programas especiales). El sistema educativo público ha sido uno de los pilares de la democracia en este país y en otros, y ha colaborado en el mejoramiento de la calidad de vida de la sociedad en su conjunto. Sin intentar idealizar la escuela pública, desde una perspectiva sociológica está claro que el acceso universal a la educación ha generado mayor inclusión e igualdad en el sistema (Labaree, 1997). Esto constituye un gran logro, basado en una concepción de la educación como bien público. Si la educación se considerara un bien privado, sus prioridades serían el beneficio individual y la competencia, y la exclusión se convertiría en una característica básica en lugar de un problema.

En lo que sigue haré primero una revisión de los programas de bonos que se desarrollan en los Estados Unidos y en otros países, analizaré luego los términos actuales del debate y el contexto político ideológico a fin de explorar las consecuencias políticas y sociales de la implementación de estos programas. A pesar de que la elección de escuela privada se promueve desde diferentes posiciones del espectro ideológico, el actual contexto y tendencias en los Estados Unidos indicarían que la mayoría de las propuestas se orientan a la mercantilización de la educación.

3.2.1. Historia de los programas de bonos en los Estados Unidos

Cuando los educadores en los Estados Unidos oyen hablar de programas de bonos, seguramente lo primero que piensan es en los de Milwaukee y Cleveland (Nota 7). Milwaukee comenzó su primer programa piloto en 1990 cuando aproximadamente 500 estudiantes de familias pobres recibieron bonos para asistir a escuelas privadas. Por primera vez en 1999 los bonos pudieron usarse para concurrir a escuelas religiosas. En Cleveland, la mayoría de los alumnos de escuela primaria incluidos en el plan de bonos, concurren a escuelas religiosas. La constitucionalidad de ambos programas ha sido cuestionada en las cortes porque el financiamiento estatal a escuelas privadas se considera una violación de la separación entre la iglesia y el estado (Roberts, 1999). En estos dos casos, las Cortes Supremas de Wisconsin y de Ohio decidieron que los programas de bonos no violaban las constituciones federal ni estatales. Posteriormente, la Corte Suprema de los Estados Unidos declinó la apelació en el caso de Wisconsin (Walsh, 1999).

Sin embargo, la historia de los bonos escolares no empieza con estos programas recientes. De acuerdo con Spring (1997), Milton Friedman fue el primer norteamericano que propuso el uso de bonos como medio para permitir la elección de escuela, y que el gobierno financie la educación pero no maneje las escuelas. Spring (1997) afirma que desde los años cincuenta a los noventa, los conservadores han proclamado que el mayor problema que afecta a las escuelas es el control burocrático. En coincidencia con Fridman, muchos conservadores abrazan el concepto del libre mercado, pero en cambio rechazan la idea de abandonar totalmente el control, especialmente en el plano social y moral.

Guiados por otro tipo de ideas, Christopher Jenks en Arkansas, y John Coons y Stephen Sugarman en California, fueron investigadores preocupados por la

inequidad en educación a fines de los años sesenta. Jenks elaboró un modelo de elección de escuela que se aplicó en Alum Rock, Arkansas, entre 1969 y 1973. El objetivo del proyecto era mejorar las oportunidades educacionales para los niños desfavorecidos, para quienes la zonificación tradicional de las escuelas constituía segregación (Lauder et al., 1999). En California, Coons y Sugarman defendieron los bonos como una solución potencial para la inequidad escolar en un exitoso caso en la Corte en 1971. El estado podía dar a las familias de zonas pobres bonos para llevar el gasto de educación en esos distritos al mismo nivel del que tenían los distritos ricos (Miller, 1999).

No tan conocidos como los programas de Cleveland y Milwaukee son los de Vermont y Maine, donde las leyes de educación permiten que los distritos escolares pequeños que no tienen escuela secundaria paguen la escuela de los estudiantes en otros distritos. La ley de Vermont, de 1869, permitía que los alumnos se transfirieran a cualquier escuela pública o privada de su preferencia, hasta que una decisión de la Corte Suprema de Vermont en 1961 restringió el programa a escuelas no confesionales (Mathis y Pearl, 1999). A diferencia de Vermont, la ley del Estado de Maine siempre prohibió específicamente el uso del financiamiento público para concurrir a escuelas religiosas.

Otros programas operan en una línea similar a los mencionados. En 1997, la Legislatura del Estado de Arizona sancionó una reglamentación de crédito en las tasas que otorga crédito de hasta $500 en los impuestos a los contribuyentes a programas de becas para concurrir a escuelas privadas. Este crédito no incluye bonos en el mismo sentido que los anteriores, pero sí beneficia a los estudiantes que asisten a escuelas religiosas. En un caso presentado, la Corte Suprema de Arizona decidió que el crédito en las tasas no viola ni la Constitución de los Estados Unidos ni la del Estado. Esta decisión fue apelada ante la Corte Suprema de los Estados Unidos y el proceso continúa (Walsh, 1999).

La tendencia a usar dinero público para la educación privada está creciendo a lo largo de la nación. La Legislatura del Estado de Florida acaba de aprobar el primer programa de bonos del país que comprende a todo un estado; Texas, New Mexico y Pennsylvania podrían ser los siguientes. También el Intendente de la Ciudad de New York, Rudolph Giuliani ha propuesto la introducción de bonos escolares para ayudar a los estudiantes pobres. Si los bonos escolares violan la separación constitucional entre la iglesia y el estado, o si pueden ayudar a los estudiantes pobres son sólo dos preguntas en un debate más amplio, que tiene muchas más implicaciones. Mientras los defensores del modelo de mercado presentan el sistema de bonos escolares como un fenómeno creciente e inevitable (Walsh, 1999), no está claro cuál es su aporte para solucionar los graves problemas que enfrentan las escuelas públicas.

3.2.2. Algunas conclusiones de estudios de investigación

La literatura reciente no muestra precisamente que el sistema de bonos favorezca una educación mejor y más equitativa para todos los niños sino más bien que sirve a los intereses de los defensores del mercado. Lo que sigue es un resumen de las conclusiones de uno de los estudios más importantes realizados sobre programas de bonos en los Estados Unidos.

John Witte (1998) analizó los cinco primeros años de la experiencia de Milwaukee. Como ya dije, inicialmente el programa no autorizaba su uso para concurrir a escuelas confesionales, estaba orientado sólo a alumnos de escuelas públicas y limitaba el número de estudiantes con bonos por escuela. En 1995 la ley fue modificada permitiendo entrar a las escuelas religiosas, incluyendo a estudiantes de escuelas privadas, y anulando la restricción de cantidad de alumnos en una escuela.

Contrariamente a lo que los defensores de la elección de escuela hubieran pronosticado, la inscripción en el programa de bonos creció en forma lenta, si bien constante, a lo largo de los cinco años, sin alcanzar nunca el número máximo de estudiantes que la ley permitía. El perfil demográfico de la matrícula era de familias muy pobres, la mayoría negros e hispanos, con un promedio de dos hijos por

familia. Los padres mostraban un mayor nivel de educación, más compromiso con la escuela y un mayor grado de insatisfacción con la escuela anterior que los padres que no formaban parte del programa. La interpretación que Witte hace de esto es la siguiente:

Por un lado, el programa demostró claramente que un programa puede focalizarse exitosamente en las familias pobres que han tenido malas experiencias en sus escuelas públicas previas. De este modo, el programa creó el tipo de oportunidad igualadora que pretendía. Por el otro, uno podría también afirmar que el programa está privando a las escuelas públicas de familias que tienen padres más educados y que están comprometidos activamente en la educación de sus hijos—es decir, el tipo de padres que potencialmente podrían colaborar con los esfuerzos de mejoramiento (p. 236- 237).

Los alumnos que usaban bonos tuvieron calificaciones y tasas (altas) de abandono similares a los de las escuelas públicas de la ciudad.

3.2.2. Elección de escuela en otros países

H ay otras investigaciones sobre la implementación de programas de elección de escuela en diferentes países. Si bien los contextos son diferentes, algunas de las conclusiones son suficientemente consistentes como para intentar resumirlas aquí. Un estudio realizado por Gewirtz et al. (1995) en Inglaterra sugiere que los resultados materiales de los programas de elección de escuela incluyen el aumento de la diferenciación y segregación, una redistribución de recursos negativa para los que más necesitan y la redefinición y restricción de los fines de la educación.

Un estudio internacional realizado por Whitty et al. (1998) sugiere que en Inglaterra y Gales la elección de escuelas privadas ha beneficiado a los niños provenientes de hogares de clase media, muchos de los cuales hubieran asistido a escuelas privadas de todos modos, más que a los niños pertenecientes a hogares de trabajadores de las zonas pobres de la ciudad. En Suecia, el mismo estudio indica que, de acuerdo con Miron, "el financiamiento público de escuelas privadas está favoreciendo el crecimiento de la segregación social en las áreas urbanas porque esas escuelas pueden controlar la admisión mucho más que las escuelas municipales" (p. 122). Los datos de Australia muestran que un sector privado orientado a minorías privilegiadas está creciendo a expensas de las escuelas públicas.

Un proyecto de investigación realizado en Nueva Zelandia (Lauder et al., 1999) confirma la evidencia previa de que los programas de elección de escuela mejoran las oportunidades de aquéllos que ya están en mejores condiciones para elegir, exacerbando la polarización existente en base a la segregación por zona de residencia.

3.2.4. El debate actual en los Estados Unidos

A pesar de que el número real de estudiantes incluidos en programas de bonos en los Estados Unidos es por ahora bastante reducido, es considerable el grado de agitación alrededor del tema, reflejado en los discursos públicos, la cobertura de los medios, las propuestas de ley, los casos en las cortes, y los informes de diferentes fundaciones y asociaciones vinculadas a la educación. El amplio y acalorado debate parece favorecer a los que atacan a la escuela pública y querrían que el sistema educativo se ajustara al modelo de mercado. Como he dicho anteriormente, la definición de educación como bien público o privado es un punto central de este debate. Mientras los defensores de la elección de escuela insisten en los beneficios para las familias pobres, los opositores afirman que la escuela pública asegura igualdad de oportunidades para todos los niños, mientras que la privatiz ción deja que la "libre" competencia seleccione ganadores y perdedores.

Argumento s a favor de los bonos. El principal argumento que utilizan sus defensores es que los bonos escolares ayudarán a los niños pobres a mejorar sus logros ya que no deberán permanecer en las 'mediocres' escuelas públicas. Si bien

éste parece ser un objetivo progresista, no hay suficientes ejemplos que avalen esta postura. El ataque y la desvalorización de las escuelas públicas es un punto fuerte en la lucha de los neoliberales en contra de la burocracia estatal.

Otro argumento importante es el de devolver a los padres el poder de tomar decisiones sobre la educación de sus hijos, que los neoliberales relacionan con la reducción del poder del estado sobre la vida de la gente, criticando su monopolio sobre la educación. En este sentido, también afirman que dar a los padres la posibilidad de elegir mediante los bonos, obligará a las escuelas públicas a competir en un sistema de mercado para atraer o retener alumnos. De este modo, las escuelas tendrán que mejorar su calidad o perderán estudiantes llegando incluso a cerrar. Otro argumento económico es que la gente que paga impuestos y manda a sus hijos a escuelas privadas, como parte de los impuestos financia la educación pública, paga por la educación dos veces.

Los neoconservadores que defienden la elección privada de escuela también exigen que se respeten valores liberales como la libertad—libertad religiosa y libertad de expresión—permitiendo a los padres educar a sus hijos de acuerdo con sus creencias religiosas. Detrás de muchos de estos argumentos están los intereses económicos de las escuelas privadas y las corporaciones que quieren aprovechar el negocio de la educación (Spring, 1997).

Los argumentos en contra. Los que se oponen a los bonos escolares enfatizan la idea de que los mismos amenazan los fundamentos de la escuela pública, negando a los niños la posibilidad de compartir experiencias de aprendizaje con otros cuyas características sociales, raciales o culturales sean diferentes. Para algunos académicos liberales y críticos, cambiar el foco de la educación para todos a la elección individual en el mercado educativo beneficiaría solamente a los que ya tienen privilegios. Al mismo tiempo, los programas de bonos, usados como una solución rápida para el bajo rendimiento de los alumnos, pueden desplazar la atención de los críticos problemas sociales, económicos y políticos que se encuentran en la base de los problemas educacionales, como la pobreza y la discriminación. Con respecto al reclamo de libertad religiosa y de expresión, los opositores responden que usar fondos públicos para pagar escuelas confesionales viola la separación constitucional entre la iglesia y el estado. Otro argumento usado contra los bonos escolares es que reducirían los recursos imprescindibles de las escuelas públicas, aumentando las ganancias de las escuelas privadas a expensas del estado. En especial las escuelas públicas de las zonas marginales verían reducidos sus ya exiguos presupuestos mientras el dinero disponible para las escuelas privadas aumenta, agrandando a su vez las diferencias entre unas y otras.

Mientras los programas de bonos escolares son a menudo vistos como una solución en zonas donde las escuelas públicas no satisfacen las necesidades de los alumnos, particularmente las que atienden a chicos pobres o pertenecientes a minorías, los opositores afirman que el monto de dinero que los estudiantes reciben –aproximadamente entre $2100 y $2500 anuales, según Miner (1992)-- no les permite el acceso a una escuela privada de alto nivel. A menudo, las familias pueden afrontar sólo una elección limitada a una escuela religiosa o de minoría. Adicionalmente, las escuelas confesionales tienen requisitos de ingreso, lo cual significa que la elección no es de los padres, sino de las escuelas, reforzando la discriminación. O algunas escuelas pueden volverse étnicamente homogéneas a través de la segregación voluntaria de los padres. Otro argumento en contra es la falta de control ya que las escuelas privadas no deben rendir cuentas. Por otra parte, no está demostrado que las escuelas privadas son mejores que las públicas, en cambio algunas investigaciones muestran que los logros de alumnos incluidos en programas de bonos escolares fueron similares a los de los no incluidos (Witte, 1998).

3.3. Grupos y fuerzas que disputan por el control de la educación

La posición de la Asociación Nacional de Padres y Maestros (PTA, 1997) sobre la

elección de escuela consiste en defender los derechos de todos los niños y el mejoramiento de la educación pública. Apoyan "las elecciones educacionales dentro de las escuelas públicas" y algunos principios tales como que la escuela brinde información apropiada a los padres, transporte gratis para los estudiantes, procesos de admisión justos y abiertos, conducta no discriminatoria, y que los fondos públicos sean sólo para las escuelas públicas.

Sin embargo, de acuerdo con algunos autores (Murphy, 1999; Murphy et al., 1998), la elección de escuela dentro del sector público es una de las iniciativas que se orienta a introducir las fuerzas del mercado dentro del sistema, ya que los contratos constituyen un grado de privatización de la escuela pública. Los padres no participan como ciudadanos o miembros de comunidades sino como consumidores, por lo tanto, las soluciones se ubican en la esfera económica en lugar de en la esfera política. Spring (1997) analiza en detalle los grupos e intereses que están detrás de esta agenda. Por ejemplo, la Heritage Foundation "es una parte importante de la red de trabajo de la derecha" (p. 33). La Heritage Foundation, el Hudson Institute, el Manhattan Institute, y el American Enterprise Institute son importantes bases para los analistas neoconservadores: Denis Doyle, Chester Finn Jr., Diane Ravitch y Bruno Manno, entre otros, difunden estas ideas e influe cian a la opinión pública. La siguiente cita es de un documento de la Heritage Foundation:

La elección de escuela hizo sólidos avances en 1997. Los principios de competencia en el libre mercado y de libertad para los padres de elegir la mejor educación para sus hijos ganaron el apoyo de muchas legislaturas estatales, gobernadores, educadores y padres –especialmente padres de zonas pobres urbanas. . . . Una razón de este apoyo ha sido que muchas investigaciones han demostrado la continua declinación de las escuelas públicas en los resultados de los tests, en el nivel de seguridad, en los recursos de que disponen los docentes, y sobre todo en la falta de rendición de cuentas, especialmente en ciudades importantes como el Distrito de Columbia. (Shokraii, 1998)

Los informes como éste utilizan conceptos significativamente progresistas para su propaganda, tales como libertad y ayuda a los estudiantes pobres, agitando al mismo tiempo la crisis de la escuela pública. Otros aspectos seductores que esta fundación proclama son la participación de los padres en las decisiones y el mejoramiento de la calidad educativa:

La elección de escuela . . . es la reforma educativa más promisoria en los Estados Unidos hoy . . . ella sola transfiere poder de los burócratas a los padres en decisiones educacionales básicas y brinda a los niños pobres que asisten a las peores escuelas la opción inmediata de una educación de mejor calidad. (Bolick, 1997)

Lo nuevo en esta lucha es que el discurso político de los conservadores se apropia, modificándolas, de las preocupaciones de los liberales. Para Popkewitz (1997), la demanda de "elección", una metáfora política y económica con profunda atracción simbólica. es un ejemplo. El afirma que el modelo educacional de contrato se relaciona con la privatización en otros sectores de la economía, la cultura y la política, como por ejemplo, la salud, la jubilación, el transporte. Sus defensores expresan que si los padres eligen la escuela de sus hijos, las fuerzas del mercado producirán motivación y mejores resultados para los que anteriormente carecían de opciones. Siguiendo al mismo autor, y tal como vimos en los ejemplos precedentes, la corriente privatista se presenta como una expresión de toda la sociedad.

Los "think tanks" conservadores trabajan hacia dos frentes: la opinión pública y el sector empresarial. Hacia la opinión pública, enfatizan la crítica hacia la burocracia y el bajo rendimiento de los estudiantes, y levantan valores progresistas como la libertad de elección, la decisión de los padres y las mejores alternativas para los pobres, buscando convertir su retórica en sentido común. Hacia el sector empresarial se dirigen con un lenguaje comercial (por ejemplo, "vender" las escuelas charter), y

le brindan el soporte ideológico para reclamar la eliminación de regulaciones y lograr buenas inversiones. Según un informe sobre la distribución de fondos entre las fundaciones, estos grupos

continúan promoviendo una visión altamente ideológica a través de múltiples frentes políticos, reclamando la privatización de la esfera pública y la elevación del mercado como uno de los principales mecanismos de mediación social y distribución de recursos. Los "think tanks" conservadores atraen crecientes contribuciones de las corporaciones interesadas en afectar el proceso político (National Committee for Responsive Philantropy, 1999).

Para los neoconservadores la equidad no constituye un requisito para la democracia, en tanto ésta se reduce a un marco legal para hacer negocios. En los textos citados más arriba podemos leer expresiones en las que se destaca la preocupación por los estudiantes pobres y sus familias y por aumentar su poder de decisión y participación. En este sentido, Anderson (1998) analiza el problema de la elección de escuela como "estrechamente relacionado con la discusión sobre el papel de los ciudadanos en una sociedad democrática" (p. 584). Los defensores de la "libre elección" ven a los padres como consumidores y consideran su participación fuera del contexto político y social, y al margen de "una teoría de la ciudadanía en una sociedad democrática" (p. 585). Las teorías político filosóficas de la ciudadanía y la democracia han sido reemplazadas en el discurso educ cional por los conceptos económicos de eficiencia, competencia, libertad de consumo y subcontratación. Uno de los libros más invocados y polémicos en este sentido ha sido el de Chub y Moe (1990) cuya tesis principal es que el problema de la educación es el funcionamiento democrático, porque esto acarrea burocracia e ineficiencia, y por lo tanto la solución es la elección en el mercado educativo privado, que asegura a los padres mayor libertad. Este uso del término libertad resulta engañoso, porque "ofrece la promesa de poder y obstruye la relación entre el beneficio deseado y los recursos necesarios para obtenerlo" (Munín, 1999, p. 24). Habiendo pruebas suficientes de que el acceso al consumo no es libre sino limitado a los recursos disponibles, la adopción del modelo de mercado en educación está en contradicción con los principios democráticos de equidad y universalidad que implica la noción de educación como un bien público.

Labaree (1997) realiza una importante distinción entre la consideración de la educación como un bien público o privado, su relación con los objetivos educacionales prevalecientes, y las consecuencias que esto tiene sobre el credencialismo y la estratificación social: "la creciente hegemonía del objetivo de movilidad social y su estrecho enfoque de la educación basado en el consumidor ha llevado a la reconceptualización de la misma como un bien puramente privado" (p. 51). Para este autor la educación como bien público tiene un significado incluyente y proporciona beneficios sociales compartidos, mientras que como bien privado se vuelve excluyente y brinda beneficios individuales selectivos. Una de las consecuencias de ésta última concepción es el triunfo del credencialismo por sobre el aprendizaje, lo cual, a su vez, aumenta la estratificación social.

Sus defensores presentan la mercantilización de la educación como un hecho inevitable, y tienen potente apoyo político y económico para lograrla. Como no existe un mercado con competencia perfecta, esta clase de estrategias continúan mejorando las oportunidades de los privilegiados, y aumentando la exclusión. Nada más lejos de los ideales democráticos representados por los sistemas educativos públicos, que deben ser sostenidos y mejorados para poder ofrecer una mejor educación para todos.

El análisis de esta controversia contribuye a echar luz sobre los significados políticos de estos procesos. La información evidencia que intereses políticos y económicos muy fuertes se están embanderando en problemas educacionales como parte de su ataque al estado en general, y al sistema educativo público en particular como una de los últimos restos del estado de bienestar que todavía subsiste.

4. Implicaciones para América Latina

Las políticas de apertura económica, ajuste presupuestario y achicamiento del estado que se desarrollaron en América Latina como parte de la agenda conservadora de los ochenta (si bien en Chile y Argentina comenzaron antes) tuvieron como principales estrategias el recorte del gasto público, la desregulación y las privatizaciones (Gamarra, 1994). Es interesante hacer un repaso rápido de los efectos más visibles que tuvieron estas medidas en Argentina, por ejemplo, ya que se tomaron en nombre del mejoramiento de los servicios, la eficiencia en la administración y contra el monopolio del estado burocrático, que son argumentos similares a los que se utilizan para defender la mercantilización de la escuela. Si bien algunos servicios mejoraron, como la telefonía, ésta no es la característica general; baste mencionar los graves problemas con la electricidad y con el transporte aéreo, en los que la falta de inversión se puede calificar eufemísticamente de irresponsable. Si hablamos de los ferrocarriles, cuyo servicio es de calidad dispar según el recorrido, es claro que no se pudieron privatizar los ramales no lucrativos, ni tampoco mantener, lo cual aisló a centenares de comunidades en el interior del país, privándolas del medio de transporte más tradicional y económico. Tampoco se eliminaron los monopolios, sino que las compañías pasaron a manos privadas dividiéndose en algunos casos por zonas, pero manteniendo la clientela cautiva en cada área geográfica.

El corolario del repaso de estas políticas y algunos de sus efectos es la reiterada observación de que estas soluciones generadas desde los países centrales para todo el "tercer mundo" o "los países en desarrollo" sin tener en cuenta las historias, culturas y peculiaridades, terminan siendo en el mejor de los casos un mal negocio para la gente, y en el peor, un retroceso en el que las pérdidas sociales y humanas son irrecuperables. Según Coraggio, la lógica de las actuales políticas sociales del Banco Mundial (Coraggio y Torres, 1997), que dan marco a las políticas educativas, puede interpretarse en tres sentidos principales: 1) para "*continuar* el proceso de desarrollo humano"; 2) para "*compensar* coyunturalmente los efectos de la revolución tecnológica y económica que caracteriza a la globalización"; y 3) para "*instrumentar* la política econó ica" (p. 14-15). A pesar de que este autor tiene una visión optimista acerca de las posibilidades de operar individual o sectorialmente dentro de las contradicciones y limitaciones de esta realidad, no cabe duda de que el conservadurismo tiene una "avasalladora iniciativa" en el discurso dominante, de la que las fuerzas que podrían oponerse carecen, actuando en forma débil y defensiva.

A pesar de que el impacto de las políticas neoconservadoras ha sido negativo en múltiples ámbitos de la vida social y económica, similares propuestas para educación mantienen –o renuevan-- su vigor. Junto con las recomendaciones de las agencias internacionales y los bancos para la economía, consecuentes recetas con similares fundamentos se difundieron para los sistemas educativos: descentralización, sistemas de evaluación, énfasis en la educación básica y las habilidades para el trabajo, eficiencia en el manejo de los fondos. La descentralización fue una de las más difundidas en la mayor parte de los países, ya sea a través de la municipalización o de la federalización. Entre la mayor parte de los investigadores hay acuerdo en que una consecuencia de la descentralización en la mayor parte de los países de la región fue el aumento de la desigualdad, ya que se transfi ieron servicios educativos con insuficiente o nula transferencia de fondos, a distritos y estados o provincias con muy diferenciada capacidad de absorberlos y sostenerlos (Arnove, 1997; Munín, 1998; Tiramonti, 1998). Los efectos más relevantes fueron: 1) Desinversión en infraestructura y equipamiento, 2) Disminución de los salarios docentes, 3) Aumento de las tareas asistenciales en la escuela, 4) Dificultades de los gobiernos provinciales para afrontar las necesidades de las escuelas, y 5) Menor exigencia sobre la calidad de la enseñanza (Pini y Cigliutti, 1999). De hecho las condiciones y posibilidades de enseñanza y aprendizaje en las escuelas de los distritos pobres empeoraron. Este efecto negativo no se compensó con un aumento de la democracia o la participación en las comunidades, ya que en general cada sistema local mantuvo las características jerárquicas de sus orígenes.

La propuesta básica actual del Banco Mundial para el sistema educativo es minimizar la gratuidad y adoptar lo más posible el modelo de mercado (Coraggio y Torres, 1997). El correlato de las tendencias o corrientes pro- mercado en educación en los Estados Unidos lo encontramos en las recomendaciones para los países "en desarrollo" que se enmarcan dentro de la subsidiariedad del estado o el financiamiento basado en la demanda. Esta idea no es nueva, como casi ninguna de estas propuestas lo es, pero lo nuevo es que entran en las agendas de los bancos e instituciones de desarrollo internacionales y tarde o temprano se convierten en "lo que hay que hacer" en América Latina. En estrecha relación con los movimientos que describimos por la elección de escuela y la introducción del mercado en educación, encontramos que el Banco Mundial propone el financiamiento de la demanda como una tendencia del desarrollo (Patrinos y Ariasingam, 1998 n español, 1997 en inglés), y en Argentina y otros países de América Latina ya se está implementando y escribiendo sobre eso (Filmus, 1998; Llach, 1997).

Patrinos y Ariasingam (1998) definen el financiamiento basado en la demanda como "la canalización directa de fondos públicos a personas, instituciones y comunidades en función de la demanda expresada". El mismo "constituye una opción pragmática para la introducción de reformas necesarias teniendo presentes las necesidades locales y los recursos disponibles" (p. v). Entre los múltiples mecanismos descriptos como parte de esta estrategia se encuentran los bonos escolares, dentro de un conjunto ecléctico que va desde las becas o las donaciones hasta la organización de la comunidad y el trabajo voluntario. Algunos de estos mecanismos se basan en la movilización de recursos de las comunidades y todos en la información de que disponen los interesados para construir y expresar su demanda.

No dudamos de que las comunidades pueden potenciar en gran medida sus recursos, pero esto difícilmente pueda llegar a satisfacer las necesidades de los más desfavorecidos si no aumentan el nivel de empleo y el gasto social. La información y su posibilidad de utilización tampoco están equitativamente distribuidos en la sociedad, con lo cual la capacidad de demanda es prácticamente inexistente para los sectores más pobres, en tanto para los sectores medios y altos es parte de su capital cultural. Más que contribuir al desarrollo, y especialmente al mejoramiento de oportunidades, este tipo de mecanismos tiende a mantener o empeorar las situaciones de inequidad en educación, porque bajo la apariencia de que los recursos se utilizan de manera más eficiente y de que las comunidades tienen más protagonismo, se siguen distribuyendo los mismos de manera regresiva. Muchos de los argumentos y las conclusiones de investigaciones que fundamentan a oposición a la elección y los bonos escolares en los Estados Unidos, que fueron suficientemente expuestos, resultan aún más válidos para América Latina.

Este tipo de propuestas economicistas, además de considerarse universalmente útiles e ignorar todo contexto, desconocen la complejidad de los aspectos y relaciones implicados en el proceso educativo en sus distintas dimensiones. Pero lo que es peor, desprecian los graves riesgos que encierran para nuestro crecimiento democrático las políticas que siguen girando alrededor de las necesidades de la economía y no de las necesidades sociales, tendiendo a segmentar más que a fortalecer y a segregar más que a integrar el ya castigado tejido social.

5. Reflexiones finales

El problema principal, desde mi punto de vista, de la libertad que defienden los neoliberales, es que la libertad, al igual que todos los demás bienes, no tiene una distribución equitativa en la sociedad, a menos que se complemente con el concepto de justicia, que no figura en el glosario neoconservador. La consecuencia concreta de estas prédicas y prácticas en la mayoría de los países en los que penetraron, es el aumento de la desigualdad. La profundización de estos procesos fragiliza y deslegitima gradualmente la democracia, reduciendo cada vez más las opciones y esperanzas de muchos, y aumentando el número de los que ven la violencia individual como su única alternativa.

En los Estados Unidos la polémica entre los que están a favor y en contra de la privatización es fuerte, pero sin embargo hay acuerdo en que para mejorar la educación el estado tiene que reforzar el financiamiento, y en que los docentes son la "fuerza" del mejoramiento. Con respecto a las escuelas charter, algunas brindan mejores condiciones de aprendizaje y oportunidades para aprender a los alumnos desaventajados, pero la mayoría de las escuelas públicas podrían trabajar mejor en condiciones similares. Los resultados hasta ahora muestran que la tendencia es hacia una mayor estratificación y privatización. Desde la perspectiva de la democracia y la equidad podríamos preguntarnos si es correcto que se use más dinero del estado para promover mejores escuelas para algunos en lugar de desarrollar seriamente mejores condiciones de enseñanza y aprendizaje para todos, y si es adecuado que fondos públicos financien lu rativos negocios en educación. En la práctica la constitución de la escuela como una institución del mercado es una forma de controlarla mediante las corporaciones privadas.

Con respecto a los bonos escolares, diferentes estudios sobre programas de elección de escuela que permiten a los alumnos asistir a escuelas privadas usando fondos públicos muestran conclusiones similares. Los resultados de estos programas son la segregación y el aumento de la estratificación social, drenaje de recursos desde escuelas públicas ya empobrecidas hacia escuelas privadas, con frecuencia religiosas, y la conversión de la educación en un negocio. No hay evidencias suficientes de mejoramiento en el rendimiento de los alumnos desaventajados que justifiquen la insistencia en los beneficios para la gente por parte de los defensores de los programas de bonos. No sólo estos programas no son el camino para mejorar la educación sino que tienen consecuencias sociales negativas.

Los estudios consideran el modelo de mercado en educación como algo dado en los países en donde las reformas fueron más amplias. En los Estados Unidos, la mercantilización es todavía una tendencia pero hay fuerzas potentes que luchan por aumentar su influencia. El argumento de que los bonos escolares puede ayudar a los estudiantes pobres y promover la libertad de los padres descansa en el modelo de mercado para corregir problemas sociales, en lugar de buscar soluciones políticas. Esta clase de respuestas han tenido similares consecuencias en educación a las que tuvieron las políticas de desregulación promovidas por los neoliberales en otras áreas sociales desde los ochenta —por ejemplo en salud y previsión: el aumento de la polarización social.

Siguiendo a algunos de los autores citados (Anderson, 1998; Labaree, 1997; Popkewitz, 1997) es necesario considerar los fundamentos de este debate en sus más profundas diferencias filosóficas e ideológicas relacionadas con los conceptos de democracia y equidad. En el discurso de la elección de escuela, los principios políticos de la ciudadanía y la democracia han sido reemplazados por conceptos económicos como eficiencia, competencia, libertad de consumo y contrato. Estas características pueden ser apropiadas para el mundo de los negocios, pero no como valores para educar ciudadanos. En estas definiciones, el ciudadano se ha convertido en consumidor, aislado de su comunidad. La educación como un bien público es inclusiva y proporciona beneficios sociales compartidos, en tanto que como un bien privado, es exclusiva y brinda beneficios individuales selectivos. Si la educación se vuelve un bien privado para satisfacer a consumidores individuales y empresarios, el individualismo y la competencia serán los únicos valores posibles, con el consiguiente retroceso del principio de equidad. En los Estados Unidos aumentará el número de estudiantes segregados y excluidos, y esto erosionará la democracia. En Argentina, al igual que en otros países de América Latina, las mayorías serán segregadas y excluidas, y la democracia, ya frágil, será difícil de construir y mantener.

Notas

1. En el sistema educativo norteamericano cada distrito está dividido en zonas y teóricamente los padres no pueden elegir la escuela para sus hijos porque

tienen que inscribirlos en la que les corresponde según el domicilio, ya que el estado provee el servicio de transporte escolar gratuito dentro de la zona que comprende cada escuela.

2. "Los estándares académicos describen lo que todo estudiante debe saber y ser capaz de hacer en relación con los contenidos académicos de las áreas (por ejemplo Matemáticas, Ciencias, Geografía). También definen cómo los estudiantes demuestran sus habilidades y conocimiento" (U. S. Department of Education, 1996, p. 8).
3. En los Estados Unidos las escuelas privadas, que constituyen alrededor del 15% del total de escuelas, en general no reciben subvención por parte del estado ni tampoco están sujetas a mecanismos de control por parte de las autoridades educativas.
4. Edison Project es una enorme experiencia piloto de escuelas privadas en los Estados Unidos. El objetivo de su creador, Christopher Whittle, es obtener capital para fundar un gran número de escuelas que den beneficio económico. Su estrategia para ganar más es ahorrar dinero reduciendo burocracia y docentes, y aumentando el trabajo voluntario de padres y alumnos (Weiss, 1999).
5. Central Park East es una escuela charter situada en Central Harlem, en New York. Su experiencia como charter mostró la posibilidad de éxito escolar para todos los niños, aún los desaventajados socialmente, basada en una concepción y práctica educativas apropiadas.
6. El año escolar comienza en agosto y termina en mayo.
7. Ciudades del noreste de los Estados Unidos, ubicadas en los estados de Wisconsin y Ohio respectivamente.

Referencias

American Federation of Teachers (1996). Charter school laws: Do they measure up? (Online) Available http://www.aft.org/edissues/chartrep.htm [1999, May 2]

Anderson, G. L. (1998). Toward authentic participation: Deconstructing the discourses of participatory reforms in education. *American Educational Research Journal, 35* (4), 571-603.

Apple, Michael (1996). *Cultural politics and education*. NY: Teachers College Press.

Arnove R. (1997). Neoliberal education policies in Latin America: arguments in favor and against, en C. A. Torres y A. Puiggrós (Eds.) *Latin American education. Comparative perspectives* (79-100). Colorado (U.S.A.) y Oxford (U.K.): Westview Press.

Arons, Stephen (1997). *Short route to chaos. Conscience, community, and the re-constitution of American schooling*. Amherst: University of Massachusetts Press.

Berliner, D. & Biddle, B. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools*. NY: Longman.

Bertoni, A., Poggi, M. y Teobaldo, M. (1996). *Evaluación. Nuevos significados para una práctica compleja*. Buenos Aires: Kapelusz, Colección Triángulos pedagógicos.

Bolick, C. (1997, September 19). School choice, the law, and the Constitution: A primer for parents and reformers, Roe Backgrounder No. 1139. The Heritage Foundation (Online) Available http://www.heritage.org/library/categories/ [1999, May 2].

Bray Duff, A. (1999, January 26). *The states of school reform: Vouchers, tax credits, charters all being tried. Business Daily*, Los Angeles, CA.

Brouillette, L. (1999, April). *What is behind all that rhetoric? A case study of a charter school.* Paper presented at the meeting of the American Educational Research Association, Montreal, Canada.

Chubb, J., y Moe, T. (1990). *Politics, markets, and America's schools*. Washington, D.C.: The Brookings Institution.

Coraggio, J. L. y Torres, R. M. (1997). La educación según el Banco Mundial. Un análisis de sus propuestas y métodos. Buenos Aires: Miño y Dávila.

Cobb, C. D. y Glass, G. V. (1999). Ethnic segregation in Arizona charter schools. Education Policy Analysis Archives, 7 (1) [On-line]. Available: http://epaa.asu.edu/epaa/v7n1/

Coraggio, J. L. (1997). Education policy and human development in the Latin American city, en C. A. Torres y A. Puiggrós (Eds.) *Latin American education. Comparative perspectives* (65-77). Colorado (U.S.A.) y Oxford (U.K.): Westview Press.

Feulner, E. (1998a, August 14). Charter schools are smarter schools. Heritage Commentary. The Heritage Foundation (Online). Available http://www.heritage.org/library/categories/ [1999, May 2]

Filmus, D. (1998). La descentralización educativa en el centro del debate. En A. Isuani y D. Filmus (Comps.) *La Argentina que viene. Análisis y propuestas para una sociedad en transición* (53-88). Buenos Aires: Norma.

Gamarra, E. (1994) Market-oriented reforms and democratization in Latin America: Challenges of the 1990s, en W. Smith, C. Acuña, y E. Gamarra (Eds.) *Latin America political economy in the age of neoliberal reform: theoretical and comparative perspectives for the 1990s.* (1-15) New Brunswick (U.S.A.) and London (U.K.): Transaction Publishers.

Gewirtz, S., Ball, S. J., & Bowe, R. (1995). *Markets, choice, and equity in education*. Buckingham & Philadelphia: Open University Press.

Labaree, D. (1997). *How to succeed in school without really learning: The credentials race in American education*. London & New Haven: Yale University Press.

Lauder, H. & Hughes, D., & Watson, S., Waslander, S., Thrupp, M., Strathdee, R., Simiyu, I., Dupuis, A., McGlinn, J., & Hamlin, J. (1999). *Trading in futures: Why markets in education don't work*. Buckingham & Philadelphia: Open University Press.

Liston, D. y Zeichner, K. (1997). *Formación del profesorado y condiciones sociales de la escolarización* (segunda edición). Madrid: Ediciones Morata, Paideia.

Llach, J. (1997). *Otro siglo, otra Argentina. Una estrategia para el desarrollo económico y social nacida de la convertibilidad y de su historia*. Buenos Aires: Espasa Calpe/Ariel.

Mathis, W. J. & Pearl, M. L. (1999, April). *Public funding of Religious Schools: Legal and Political Implications of Current Court Cases*. Paper presented at the meeting of the American Educational Research Association, Montreal, Canada.

Meier, D. (1995). *The power of their ideas: Lessons for America from a small school in Harlem.* Boston: Beacon Press.

Miller, M. (1999, July). A bold experimcnt to fix city schools. *The Atlantic Monthly*, 15-31.

Miner, B. (1992). From Colorado to California, Milwakee to Baltimore voucher plans proliferate. En R. Lowe y B. Miner (Eds.) *False choices, why school vouchers threaten our children's future, Special Edition of Rethinking schools*, Milwaukee, Wisc.: Rethinking Schools Ltd. P.6- 7.

Morando Rhim, L. (1998, April). *Franchising public schools: An analysis of the linkage of charter schools and private education management companies*. Paper presented at the meeting of the American Educational Research Association, San Diego, CA.

Munín, H. (1999). Introducción a la discusión alemana sobre la "autonomía" de la escuela. En H. Munín (Comp) *La autonomía de la escuela: ¿libertad y equidad? Un recorrido por la discusión alemana de los años noventa* (9- 40). Buenos Aires: Aique.

Murphy, J. (1999, April). *Governing America's schools: The shifting playing field*. Paper presented at the meeting of the American Educational Research Association, Montreal, Canada.

Murphy, J.; Gilmer, S.; Weise, R.; & Page, A. (1998). *Pathways to privatization in education*. Greenwich, Connecticut & London, England: Ablex.

Nathan, Joe (1996). *Charter schools. Creating hope and opportunity for American education*. San Francisco: Jossey-Bass Publishers.

National Committee for Responsive Philanthropy (1999). $1 Billion for ideas: Conservative think tanks in the 1990s. (Online) Available http://www.ncrp.org. [1999, May 2].

Parents and Teachers Association (1997) Public school choice column. (Online) Available http:// pta.org/programs/pubsch2/htm [1999, May 2].

Parents and Teachers Association (1998) National PTA Background Brief: Charter Schools. (Online) Available http://www.pta.org/search97cgi/s97_cgi.e 3; [1999, May 2].

Patrinos, H.A. & Ariasingam, D. L. (1998). *Descentralización de la educación. Financiamiento basado en la demanda*. Washington: Banco Internacional de Reconstrucción y Fomento/BANCO MUNDIAL.

Pini, M. E. (1993). *La democratización del gobierno de la educación pública. La experiencia de los Organos Colegiados en Italia y de los Consejos de Escuela en la Provincia de Buenos Aires*. Tesis de Maestría no publicada. Universidad de Buenos Aires.

Pini y Cigliutti (In press), Participatory reforms in the global context of neoliberalism: The case of Argentina. *Theory into Practice*.

Popkewitz, T. (1997). *Sociología política de las reformas educativas* (segunda edición). Madrid: Ediciones Morata, Paideia.

Roberts, C. (1999, February 28). Only two districts in U. S. use vouchers. *Albuquerque Journal*, B1-5.

Spring, Joel (1997). *Political agendas for education. From the Christian Coalition to the Green Party*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Shokraii, N. (1998, January 30). School choice 1998: A progress report, FYI #172, The Heritage Foundation (Online). Available http://www.heritage.org/library/categories/ [1999, May 2].

Tiramonti, G. (1998). Reestructuración del sistema educativo argentino en la tensión del cambio y la permanencia de las tradiciones políticas. En *Sociedad 12/13* 109-126, Buenos Aires, Argentina: Facultad de Ciencias Sociales (UBA), EUDEBA

U.S. Department of Education (1996, April 30). *Goals 2000: Increasing student achievement through state and local initiatives*. Office of Educational research and Improvement. Educational Resources Information Center (ERIC)

U.S. Department of Education, Office of Educational Research and Improvement, National Center of Educational Statistics (1997). *Public and private schools: How do they differ?* No. 12 (NCES 97-983). Washington, DC: U.S. Government Printing Office.

U.S. Department of Education, Office of Educational Research and Improvement (1998). *A national study of charter schools*. Second Year Report 1998. Washington, DC: U.S. Government Printing Office.

Walsh, M. (1999, June 23). Spate of legal rulings on vouchers, choice yields little consensus. *Education Week, XVIII* (41), 14-15.

Waymack, N. y Drury, D. (1999). Sizing it right: Class-size reduction and student achievement. *Policy Research*, Volume 1, Issue 1. Alexandria, VA: National School Boards Association.

Weiss, M. (1999). La microeconomía de la educación. En H. Munín (comp.), *La autonomía de la escuela: ¿libertad y equidad? Un recorrido por la discusión alemana de los años noventa.* (97-122). Buenos Aires: Aique.

WestEd & University of Southern California (1998, June 30). *The findings and implications of increased flexibility and accountability: An evaluation of charter schools in Los Angeles Unified School District*. Paper presented at the meeting of the American Educational Research Association, Montreal, Canada.

Whitty, G., Power, S. y Halpin, D. (1998). *Devolution and choice in education: the school, the state, and the market*. Buckingham & Philadelphia: Open University Press.

Witte, J. F. (1998). The Milwaukee voucher experiment. *Educational evaluation and policy analysis, 20* (4) 229-251.

Acerca del Autor

Mónica E. Pini

University of New Mexico
Albuquerque, NM 87106
Telefono 505.268.5169
Fax 505.277.5282
Email: mopini@unm.edu

Es Profesora en Ciencias de la Educación y Magister Scientiarum en Administración Pública de la Universidad de Buenos Aires, Argentina. Actualmente es candidata al doctorado en el College of Education en la Universidad de New Mexico, USA. Sus áreas de interés son la política y administración de la educación. Ha trabajado como docente en el nivel primario, secundario, universitario y en Institutos de Formación Docente, ha participado en varios proyectos de investigación educativa en la Ciudad y en la Provincia de Buenos Aires, y en el Programa Nacional de Gestión de la

Capacitación Docente en el Ministerio de Cultura y Educación.

About the Author

Mónica E. Pini

University of New Mexico
Albuquerque, NM 87106
Voice 505.268.5169
Fax 505.277.5282
Email: mopini@unm.edu

Mónica E. Pini is Professor in Sciences of Education and Magister Scientiarum in Public Administration at the University of Buenos Aires, Argentina. She is currently a Ph.D. Candidate in the College of Education at the University of New Mexico, U.S.A. Her areas of interest are politics and administration of education. She has worked as a teacher in elementary, middle, and high school, and as a university professor and in institutes for Teacher Preparation in Argentina. She has participated in several education research projects in Buenos Aires, and in the National Program of Management of Teacher Training at the Ministry of Culture and Education.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
aacosta@cucea.udg.mx

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William Mclnerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

This article has been retrieved 2479 times since March 28, 2000

Education Policy Analysis Archives

Volume 8 Number 19 March 28, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Bridging the Gap between Testing and Technology in Schools

Michael Russell
Boston College

Walt Haney
Boston College

Abstract
The widening gap between the increased use of technology in schools and the absence of computers in state-level testing programs raises important implications for policies related to the use of both technology and testing in schools. In this article, we summarize recent developments in the use of technology in schools and in state level testing programs. We then describe two studies indicating that written tests administered on paper underestimate the achievement of students accustomed to working on computers. We conclude by discussing four approaches to bridging the gap between technology and testing in U.S. schools.

Introduction

The need to improve education in the U.S. has received unprecedented attention recently in the media and in national and state elections. Prescriptions for improving schools have been many, but two of the most common are what might be called the technology and testing remedies.

The technology nostrum holds that the infusion of modern technology into schools will bolster teaching and learning and will prepare students for an increasingly technological workplace. The second prescription, which is often called high stakes testing, holds that standards- based accountability for students, teachers and schools will provide tangible incentives for improvements in teaching and learning. What is little recognized, however, is that these two strategies are working against each other in a sort of educational time warp. Recent research shows that

written tests taken on paper severely underestimate the performance of students accustomed to working on computer (Russell, 1999; Russell & Haney, 1997). The situation is analogous to testing the accounting skills of modern accountants, but restricting them to the use of an abacus for calculations.

The Computer Revolution Goes to School

Although the personal-computer revolution began only twenty years ago and widespread use of the world wide web (WWW) is even more recent, computer technology has already had a dramatic impact on society and schooling. Between 1984 and 1993, the percentage of people using computers in the workplace nearly doubled from 24.6 percent to 45.8 percent. Similarly, the percentage of people owning one or more computers in their home increased rapidly from 8.2 percent in 1984 to 22.8 percent in 1993 to 33.6 percent in 1997 (Newburger, 1997). Although schools have been slower to acquire these technologies, computer use in schools has recently increased rapidly (Zandvliet & Farragher, 1997). While schools had one computer for every 125 students in 1982, they had one for every 9 students in 1995, and 1 for every 6 students in 1998 (Market Data Retrieval, 1999). Not only are more computers in classrooms, but schools are also increasing students' use of computers and access to the Internet. A recent national survey of teachers showed that in 1998, 50 percent of K-12 teachers had students use word processors, 36 percent had them use CD ROMS, and 29 percent had them use the WWW (Becker, 1999). Although it is unclear how computers are affecting student achievement in schools (see, for example, Fabos & Young, 1999, questioning the efficacy of Internet based telecommunications exchange programs in schools), there is little doubt that the computer revolution has gone to school. As a result, more and more students are writing and performing school assignments on computers.

Performance Testing in Schools

Meanwhile, many states are increasingly seeking to hold students, teachers and schools accountable for student learning as measured by state-sponsored tests. According to annual surveys by the Council for Chief State School Officers (CCSSO, 1998), 48 states use statewide tests to assess student performance in different subject areas. Many of these tests are tied to challenging standards for what students should know and be able to do. Scores on these tests are being used to determine whether to: (1) promote students to higher grades; (2) grant high school diplomas; and (3) identify and sanction or reward low- and high-performing schools (Sacks, 1999). Currently, 32 states control, or plan to control, graduation and/or grade promotion via student performance on state-level tests. Because of the limitations of multiple-choice tests, many statewide tests include sections in which students must write extended answers or written explanations of their work. As the recent CCSSO report commented, "Possibly the greatest changes in the nature of state student assessment programs have taken place in the 1990s as more states have incorporated open-ended and performance exercises into their tests, and moved away from reliance on only multiple- choice tests" (CCSSO, 1998, p. 17). In 1996-97, an estimated ten to twelve million students nationwide participated in a state-sponsored testing program that required them to write responses long hand (given a total national K-12 enrollment of about 50 million and open-ended assessments in almost all the states in 3 out of 12 grade levels).

In Ohio, for example, students must pass the written portion of the Ohio Proficiency Test in order to graduate from high school (Fisher & Elliott, 2000). Although many observers have criticized state-sponsored high-stakes tests on a variety of grounds (e.g., Heubert & Hauser, 1999; Sacks, 1999), here we direct attention to a widely unrecognized but critical limitation of depending on these tests to drive educational reform: paper-and-pencil written tests yield misleading information on the capabilities of students accustomed to using computers.

Testing Via Computer

Research on testing via computer goes back several decades and suggests that

for multiple-choice tests, administration via computer yields about the same results, at least on average, as administering tests via paper-and-pencil (Bunderson, Inouye, & Olsen, 1989, Mead & Drasgow, 1993). However, more recent research shows that for young people who have gone to school with computers, open-ended (that is, not multiple choice) questions administered via paper-and-pencil yield severe underestimates of students' skills as compared with the same questions administered via computer (Russell, 1999; Russell & Haney, 1997). In both studies, the effect sizes for students accustomed to working computer ranged from .57 to 1.25. Effect sizes of this magnitude are unusually large and of sufficient size to be of not just statistical, but also practical significance (Cohen, 1988; Wolf, 1986). Effect sizes of this magnitude, for example, imply that the score for the average student in the experimental group tested on computer exceeds that of 72 to 89 percent of the students in the control group tested via paper and pencil.

Our research on this topic began with a puzzle. While evaluating the progress of student learning in the Accelerated Learning Laboratory (ALL), a high-tech school in Worcester, MA, teachers were surprised by the results from the second year of assessments. Although students wrote more often after computers were widely used in the school, student scores on writing tests declined in the second year of the new program. To help solve the puzzle, the school asked us to assist in comparing paper and computer administration of the tests.

In 1995, a randomized experiment was conducted, with one group of sixty-eight students taking math, science and language arts tests, including both multiple-choice and open-ended items, on paper, and another group of forty-six students taking the same tests on computer (but without access to word processing tools, such as spell-checking or grammar-checking). Before scoring, answers written by hand were transcribed so that raters could not distinguish them from those done on computer. There were two major findings. First, the multiple-choice test results did not differ much by mode of administration. Second, the results for the open-ended tests differed significantly by mode of administration. For the ALL School students who were accustomed to writing on the computer, responses written on computer were much better than those written by hand. This finding occurred across all three subjects tested and on both short answer and extended answer items. The effects were so large that when students wrote on paper, only 30 percent performed at a "passing" level; when they wrote on computer, 67 percent "passed" (Russell & Haney, 1997).

Two years later, a more sophisticated study was conducted, this time using open-ended items from the new Massachusetts state test (the Massachusetts Comprehensive Assessment System or MCAS) and the National Assessment of Educational Progress (NAEP) in the areas of language arts, science and math. Again, eighth grade students from two middle schools in Worcester, MA, were randomly assigned to groups. Within each subject area, each group was given the same test items, with one group answering on paper and the other on computer. In addition, data were collected on students' keyboarding speed and prior computer use. As in the first study, all answers written by hand were transcribed to computer text before scoring.

In the second study, which included about two hundred students, large differences between computer and paper-and- pencil administration were again evident on the language arts tests. For students who could keyboard moderately well (20 words per minute or more), performance on computer was much better than on paper. For these students, the difference between performance on computer and on paper was roughly a half standard deviation. According to test norms, this difference is larger than the amount students' scores typically change between grade 7 and grade 8 on standardized tests (Haney, Madaus, & Lyons, 1993, p. 234). For the MCAS, this difference in performance could easily raise students' scores from the "failing" to the "passing" level (Russell, 1999).

Recalling that nearly ten million students took some type of state-sponsored written test last year and that nearly half of the students nationwide use word processors in school, these results suggest that state paper-and- pencil tests may be underestimating the abilities of millions of students annually.

In the second study, however, findings were not consistent across all levels of keyboarding proficiency. As keyboarding speed decreased, the benefit of computer

administration became smaller. And at very low levels of keyboarding speed, taking the test on computer diminished students' performance (effect size of about 0.40 standard deviations). Similarly, taking the math test on computer had a negative effect on students' scores. This effect, however, became less pronounced as keyboarding speed increased.

Bridging the Gap

These studies highlight the importance of the gap between the technology and testing strategies for school improvement. Increasingly, schools are using computers to improve student learning. To measure increases in student learning, states are depending upon tests administered on paper. The open-ended questions on these tests, however, underestimate the achievement of students who regularly use computers. As a result, this mis-match between the mode of learning and the mode of assessment may be underestimating improvements in achievement. This problem is likely to increase as more students become accustomed to writing on computers. There are at least four possible ways to bridge this gap.

First, schools could decrease the amount of time students spend working on computers so that they do not become accustomed to writing on computers. Some schools have already adopted this practice. After reviewing the first study described above and following the introduction of the new paper-and-pencil MCAS test in Massachusetts, the ALL school required students to write more on paper and less on computer (Russell, 1999). In another Massachusetts school system, the principal feared that students who write regularly on computer lose penmanship skills, which might lead to lower scores on the new state test. This school increased penmanship instruction across all grades while also decreasing students' time on computers (Holmes, 1999). Such strategies, in effect reducing computer use in schools to better prepare students for low-tech tests, may be pragmatic given the high stakes attached to many state tests. But they are also short-sighted in light of students' entry after graduation into an increasingly high- tech world and workplace.

A second way to bridge the test-technology gap would be to eliminate paper-and-pencil testing and have students perform open-ended tests on computer. This might seem a sensible solution, but it will not be feasible until all schools obtain an adequate technology infrastructure. Moreover, as shown by problems in recent moves to administer some large-scale tests for adults on computers, computerized testing is not the panacea some had hoped. Among other problems, it adds considerably to the cost of testing and creates new test security concerns. But more importantly, as the second study summarized above indicates, administering open-ended tests only on computer would penalize students with poor keyboarding skills.

A third approach would be to offer students the option of performing open-ended tests on paper or on computer. On the surface, this seems like a sensible solution. However, it would add considerable complexity and cost to test administration and scoring procedures. Although there has not been a large amount of research on the extent to which computer printing versus hand-writing affects ratings of written work, Powers et al. (1994) report that significant effects can occur. Surprisingly, Powers et al. found that computer printed responses produced by adults tended to receive lower scores than the same responses produced by hand. To control for such effects, in offering tests on paper and computer, handwritten responses would need to be converted to computer text. Surely it will be some years before text recognition software is sophisticated enough to convert handwritten responses into computer text. Thus, for the foreseeable future, the cost of transcription would be prohibitive.

But beyond the need to convert responses to the same medium for scoring, the second study summarized above provides evidence that, when given the choice of using computer or paper to write their tests, many students make poor decisions as to which medium they should use. This was evidenced in two ways. First, the correlations between both students' preference for taking tests on computer or on paper and keyboarding speed and between preference and prior computer experience were near zero (less than .18). Second, preference was not found to be a significant factor in predicting students performance. Together, the added complexity of scoring open-ended responses produced in both mediums and students' apparent inaccuracy

in selecting the medium that optimizes their performance suggest that simply giving students the option of performing open-ended tests on computer or on paper would do little to reduce the gap between testing and technology.

A fourth approach, and perhaps the most reasonable solution in the short term, is to recognize the limitations of current testing programs. Without question, both computer technology and performance testing can help improve the quality of education. However, until students' can take tests in the same medium in which they generally work and learn, we must recognize that the scores from high-stakes state tests do not accurately reflect some students' capabilities. Reliance on paper and pencil written test scores to measure or judge student and/or school performance will mischaracterize the achievement of students' accustomed to working on computers. Thus, the gap between the use of technology in schools and testing programs serves as yet another reminder of the dangers of judging students and schools based solely on written test scores.

Note

We would like to acknowledge the help of Jeff Nellhaus and Kit Viator of the Massachusetts Department of Education which allowed inclusion of MCAS items in the second study summarized in this article. Also, we wish to note that this article is an expansion of an opinion essay appearing originally in the *Christian Science Monitor* in July 1999 (Haney & Russell, 1999). Additionally, we thank the National Board on Educational Testing and Public Policy (NBETPP) for its support and suggestions on earlier versions of this article. We also thank two anonymous EPAA reviewers for their suggestions for improving this article. Finally we thank Carol Shilinsky and the staff of the ALL School, and James Caradonio, the Superintendent of the Worcester, MA, Public Schools, for their generous support of the research recounted here.

References

Becker, H. J. (1999). *Internet Use by Teachers: Conditions of Professional Use and Teacher-Directed Student Use*. Irvine, CA: Center for Research on Information Technology and Organizations.

Bunderson, C. V., Inouye, D. K. & Olsen, J. B. (1989). The four generations of computerized educational measurement. In Linn, R. L., *Educational Measurement* (3rd ed.), Washington, D.C.: American Council on Education, pp. 367-407.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd ed.) Hillsadale, NJ: Erlbaum.

Council of Chief State School Officers (1998). *Key State Education Policies on K-12 Education: Standards, Graduation, Assessment, Teacher Licensure, Time and Attendance*. Washington, DC.

Fabos, B. & Young, M. (1999). Telecommunications in the classroom: Rhetoric versus reality. *Review of Educational Research, 69*(3) 217-259.

Fisher, M. & Elliott, S. (2000) Proficiency: The test questioned. *Dayton Daily News* (March 12). http://www.activedayton.com/news/2000/03/12/0313main001.html

Glennan, T. K., & Melmed, A. (1996). *Fostering the use of educational technology: Elements of a national strategy*. Santa Monica, CA: RAND.

Haney, W., Madaus, G., & Lyons, R. (1993). *The Fractured Marketplace for Standardized Testing*. Boston, MA: Kluwer Academic Publishers.

Haney, W. & Russel, M. (1999). Low-tech tests shortchange high-tech students.*Christian Science Monitor*. July 1, 1999.

Holmes, R. (1999). A gender bias in the MCAS? *MetroWest Town Online*. http://www.townonline.com/metrowest/archive/022499/.

Heubert, J. & Hauser, R. (1999). *High stakes: Testing for tracking, promotion and graduation*. (Report of the Committee on Appropriate Test Use). Washington, D.C.: Natioinal Academy Press. (Available on-line at http://www.nap.edu).

Market Data Retrieval. (1999). *Technology in Education 1999*. (A report issued by Market Data Retrieval). Shelton, CN: Market Data Retrieval.

Mead, A. D. & Drasgow, (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin, 114* (3), 449-58.

Newburger, E. C. (1997). *Computer Use In the United States*. Washington, DC: U.S. Census Bureau.

Powers, D., Fowles, M, Farnum, M, & Ramsey, P. (1994). Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays. *Journal of Educational Measurement, 31*(3), 220-233.

Russell, M. & Haney, W. (1997). Testing Writing on Computers: An Experiment Comparing Student Performance on Tests Conducted via Computer and via Paper-and-Pencil. *Educational Policy Analysis Archives, 5*(1). Available online: http://epaa.asu.edu/epaa/v5n3.html.

Russell, M. (1999). Testing Writing on Computers: A Follow-up Study Comparing Performance on Computer and on Paper. *Educational Policy Analysis Archives, 7*(20). Available online: http://epaa.asu.edu/epaa/v7n20/.

Sacks, P. (1999). *Standardized minds*. Reading MA: Perseus Books.

Snyder, T. D. & Hoffman, C. (1990). *Digest of Education Statistics*. Washington, DC: U. S. Department of Education.

Snyder, T. D. & Hoffman, C. (1994). *Digest of Education Statistics*. Washington, DC: U. S. Department of Education.

Wolf, F. (1986). *Meta-analysis: Quantitative methods for research synthesis*. SAGE University series on quatitative applications in the social sciences, series no. 07-059. Newbury Park, CA: SAGE.

Zandvliet, D. & Farragher, P. (1997). A comparison of computer-administered and written tests. *Journal of Research on Computing in Education, 29*(4), 423-438.

About the Authors

Michael Russell

Senior Research Associate
Center for the Study of Testing, Evaluation and Educational Policy
323 Campion Hall, Boston College
Chestnut Hill, MA 02467

Phone: 617.552.0889
Fax: 617.552.8419 Email: russellmh@bc.edu

Michael Russell is a research fellow for the National Board on Educational Testing and Public Policy and a research associate in the Center for the Study of Testing, Evaluation and Educational Policy at Boston College. His research interests include standards based reform, assessment, and educational technology.

Walt Haney

Center for the Study of Testing, Evaluation and Educational Policy
323 Campion Hall, Boston College
Chestnut Hill, MA 02467

Email: haney@bc.edu

Walt Haney, Ed.D., Professor of Education at Boston College and Senior Research Associate in the Center for the Study of Testing Evaluation and Educational Policy (CSTEEP), specializes in educational evaluation and assessment and educational technology. He has published widely on testing and assessment issues in scholarly journals such as the *Harvard Educational Review, Review of Educational Research*, and *Review of Research in Education* and in wide-audience periodicals such as *Educational Leadership, Phi Delta Kappan, the Chronicle of Higher Education* and the "Washington Post." He has served on the editorial boards of *Educational Measurement: Issues and Practice* and the *American Journal of Education* and on the National Advisory Committee of the ERIC Clearinghouse on Assessment and Evaluation.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina-- Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografía e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

Este artículo ha sido consultado 298 veces desde el 29 de avril de 2000

Education Policy Analysis Archives

Volume 8 Number 20 **avril 29, 2000** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES.**
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

La igualdad de oportunidades en el acceso a la educación superior: Una perspectiva socio-familiar para Galicia

Mª Jesús Freire
Universidad de La Coruña
Spain

Resumen
Este trabajo analiza la igualdad de oportunidades de acceso a los estudios universitarios en Galicia (España). El examen se realiza teniendo en cuenta la decisión adoptada políticamente de financiar la educación superior entre un 80% y un 85% del costo real. Al tenor del enorme esfuerzo de financiación pública se examina la composición del alumnado en función del nivel de estudios de los padres. El análisis confirma que con la política de financiación pública aplicada no se podrá alcanzar el objetivo de igualdad de acceso a la universidad.

Abstract
This work analyzes equality of access to the university in Galicia (Spain) as it was influenced by the political decision to finance higher education at between 80% and 85% of its real cost. The composition of the student body with respect to the level of their parents' education is examined. The analysis confirms that in spite of the significant effort at public financing, the objective of equal access will not be reached.

1. Introducción

El crecimiento del sistema universitario en España en las últimas décadas es un hecho ampliamente divulgado y conocido. Los datos publicados muestran que se ha

pasado de 384.424 alumnos matriculados en 1971 a más de un millón y medio en 1995. Esta espectacular crecimiento se ha visto favorecida por el aumento de la oferta, por la creación de nuevas universidades y también por los efectos de la llegada a la universidad de los alumnos que han disfrutado de una educación secundaria obligatoria y gratuita (Ley General de Educación 1970)

La expansión del sistema educativo ha originado aumentos en el gasto público en educación, que ha pasado de representar el 1,7% del Producto Interior Bruto español en 1971, a 4,7% en 1995 (OCDE, 1997). El incremento del gasto público en educación en el último cuarto de siglo ha estado motivado por: a) La expansión en las tasas de escolarización, sobretodo en los niveles secundario y universitario; b) La mejoría en la calidad de la enseñanza, que se refleja en una reduccción del número de estudiantes por profesor; c) Las transferencias a las familias en forma de becas y ayudas, como objetivo de política educativa en favor de asegurar igualdad de oportunidades en materia educativa para todos los ciudadanos.

Un objetivo social unánimemente aceptado es asegurar las mismas oportunidades de acceso a los estudios universitarios. No obstante, los resultados obtenidos en los últimos años, según diferentes trabajos de investigación, han sido contradictorios y la tan anhelada igualdad de oportunidades educativas presenta una evidencia poco concluyente. Por ello, es importante analizar si en los últimos años, teniendo en cuenta la importante cantidad de recursos públicos invertidos en educación, se han amortiguado las desigualdades generadas por la procedencia familiar.

La educación es una actividad de producción mayoritariamente pública; en la que existe una significativa intervención del Estado, característica común en casi todos los países desarrollados. Se trata de una actividad altamente regulada en cuanto a sus contenidos, organización y en los recursos físicos y humanos necesarios para realizarla. La fuente principal de financiación son los fondos públicos. El hecho de que la educación universitaria en España esté subvencionada (en promedio al 80-85 % del costo) por el sector público lleva a pensar que existen importantes razones económicas que justifican esta actuación.

En el campo de la economía, la razón que generalmente se esgrime para justificar la intervención pública hace referencia al concepto de bienes preferentes o de mérito para designar a aquellos que la sociedad considera esenciales para todos los individuos (Musgrave, 1959). La educación y la salud se consideran, según Baumol y Baumol, bienes de mérito, es decir, derivados de un juicio de valor. Otros autores, como Stiglitz, argumentan que los estudiantes obtienen importantes beneficios privados de su educación que les estimularían a llevar a cabo importantes inversiones educativas. A pesar de todo, Stiglitz sostiene que la intervención pública en la educación seguiría estando justificada por objetivos de equidad distributiva.

El objetivo de este trabajo es analizar la igualdad de oportunidades de acceso a los estudios universitarios en Galicia (España). El sistema educativo formal sólo puede cumplir una función igualadora o dar igualdad de oportunidades a las personas que participan de él. Es necesario señalar que la educación universitaria no está abierta a todos, sino que exige un nivel de conocimientos previos que implica haber cursado estudios de nivel medio. Una parte muy importante de las desigualdades que se observan en la enseñanza superior han cristalizado en los estudios anteriores la universidad. Sin embargo, nosotros deseamos reflejar los desequilibrios existentes a nivel universitario con los datos disponibles. En este trabajo, elaboramos unos indicadores socio-familiares que nos permitan cuantificar el carácter equitativo del sistema educativo al nivel de educación superior.

La organización del trabajo es la siguiente: en el apartado 1 hemos hecho una introducción sobre el estado de la cuestión. En el apartado 2, discutimos algunos aspectos del marco teórico de la economía de la educación. El apartado 3, presenta el análisis del acceso a la universidad y el origen socioeconómico del alumno considerando el nivel de estudios de los padres. Finalmente, en el apartado 4 establecemos las conclusiones.

Los datos utilizados son de variada procedencia. Los que representan el núcleo central del trabajo proceden del Sistema Universitario de Galicia, Consellería de Educación y Ordenación Universitaria (varios años). Los datos estadísticos del Sistema Universitario son una encuesta autocumplimentada por los alumnos al realizar la matrícula y que inevitablemente puede originar ciertos sesgos. La otra fuente de datos ha sido la Encuesta de Población Activa (EPA) 2° trimestre elaborada por el Instituto Nacional de Estadística (INE) con una periodicidad trimestral (varios años).

2. Algunas consideraciones del marco teórico

La teoría del capital humano (Schultz, 1961, 1963 y Becker, 1964) explica el aumento de la demanda de educación como consecuencia del incremento de utilidad que el individuo deriva de ella. Las razones que explican los beneficios individuales de la inversión en educación pueden ser muy diversas. Pero, en promedio, se observa que cuánto más elevado es el nivel de educación de una persona mayores ingresos percibirá a lo largo de su vida.

La economía de la educación constituye un campo de estudio con un desarrollo tan rápido como diversificado y llega a ejercer una influencia notable, tanto en el avance del quehacer científico del economista como en el ejercicio de la Política Económica. Pocas cuestiones procedentes de la Ciencia Económica han sido tan rápidamente asumidas por los responsables de la política como la importancia de la educación y su tratamiento de inversión en capital humano.

Consecuencia de todo ello, la industria de la educación superior creció en términos absolutos y relativos en las ultimas décadas. Este crecimiento no es sorprendente cuando se considera la alta prioridad que tradicionalmente la gente ha dado al papel de la educación superior en la sociedad. Se le ha considerado un vehículo poderoso de movilidad social y un determinante importante de estabilidad.

Esta expansión educativa fue, en gran medida, un proceso propugnado y dirigido desde el poder. Por una parte, como medio de asegurar el crecimiento económico y producir una fuerza de trabajo con los conocimientos necesarios para soportar el progreso tecnológico de la sociedad actual, y por otra, asegurar la cohesión política y social.

Es entonces lógico que la política educativa predominante haya estado financiando mayoritariamente la enseñanza universitaria. Con transferencias directas a las instituciones para cubrir los costes de la educación y con subvenciones a los estudiantes para paliar los gastos de matricula y manutención.

El apoyo que los electores han dado a esta política educativa se debe a la creencia de que un mayor gasto público en educación superior contribuye a la supresión de las barreras económicas en el acceso al sistema universitario, y a la idea de que esta actuación facilitará la consecución de la igualdad de oportunidades educativas.

Seamos radicales en el sentido literal del término, es decir, descendamos a la raíz del valor económico de la educación, más concretamente a la identificación de los costes y beneficios a nivel individual y social.

La partida de costes recogería así un componente privado, soportado por el individuo o su familia al prolongar la educación. En este grupo estarían, por un lado, los denominados costes directos que, expresados en forma explícita, serían el pago de la matrícula, los gastos adicionales de alojamiento y aquellos costes que son imputables a la propia actividad escolar. Y por otro lado, los costes de oportunidad de la elección, es decir, la renta que deja de percibir el individuo por estudiar en lugar de dedicarse a una actividad retribuida.

Ciertamente, a medida que se eleva la tasa de desempleo de un país el coste de oportunidad de continuar estudiando tiende a ser cada vez más bajo. En estas

circunstancias el estudiante asigna mayores expectativas de rendimiento al futuro de los que puede asignar al presente. La suma de los costes directos e indirectos representan los costes totales privados.

Los costes sociales son aquellos en los que incurre la sociedad para brindar educación a sus miembros. Sus principales componentes son los gastos de personal y los costes de funcionamiento necesarios para que se desarrolle la actividad educativa.

Como puede apreciarse por la simple enumeración, las partidas de los costes son de difícil cualificación. El paso de la enumeración conceptual a la medición concreta impone tener que establecer numerosos supuestos que siempre estarán sujetos a discusión.

Los beneficios sociales tampoco son fáciles de cuantificar. Normalmente se les denomina externalidades. Existen si más allá de los beneficios propios del sujeto educado se manifiestan ventajas de cualquier clase o naturaleza en otros miembros de la sociedad o en la sociedad de forma indiscriminada. El análisis de la tasa de rendimiento ha inducido a considerar la educación no sólo como un bien de consumo en cuanto a los beneficios que genera a corto plazo, sino también como un bien de inversión que produce unos rendimientos a lo largo de la vida de la persona educada.

De la adecuada comparación entre costes y beneficios individuales y sociales, teniendo en cuenta los perfiles temporales de ambas magnitudes, se obtienen las tasas de rendimiento privadas y sociales de la educación superior. En resumen, el mensaje de la escuela del capital humano es que al invertir en educación se incremente la productividad del individuo y como consecuencia se genera crecimiento económico.

La teoría de la selección o del filtro (Arrow, 1973 y Stiglitz, 1975) aparece como una hipótesis alternativa a la justificación de que la educación aumenta la productividad de los individuos. La teoría defendida por Arrow postula, que el nivel educativo alcanzado por un individuo desempeña la función de filtro para los empleadores que buscan trabajadores con una elevada capacidad laboral. Ante la falta de información de los empresarios sobre la cualificación de los individuos, los títulos académicos actúan como un filtro inicial en el momento de la contratación. La teoría de la selección o del filtro admite que la educación puede ir asociada a mayores ingresos e incluso a mayor productividad pero no constituye su causa.

La contrastación de una y otra teoría está llena de dificultades y de resultados contradictorios. Se pueden encontrar referencias básicas respecto a la economía de la educación (Blaug, 1970, 1976 y 1987), así como, estudios comparativos de la estructura de costes y rentabilidad educativas para un amplio abanico de países (Psacharopoulos y Woodhall (1985).

3. El acceso a la Universidad y el origen socioeconómico del alumno

En la teoría del capital humano (Becker, 1964), las diferencias de retribucionales, en equilibrio, son la resultante de la influencia conjunta de las características innatas, las cualidades naturales, la educación y generalmente las características productivas adquiridas por medio de la inversión en capital humano. Puesto que el nivel educativo alcanzado actúa como uno de los condicionantes de las oportunidades futuras de renta y de la probabilidad de acceder a un puesto de trabajo de un determinado individuo, resulta interesante considerar que factores pueden incidir en la decisión de un estudiante de acceder a la universidad.

En el Sistema Universitario de Galicia la tasa de escolarización han aumentado drásticamente en los últimos años. En el período analizado que comprende los cursos escolares 1990/91, 1993/94 y 1996/97, esta ha tenido un incremento medio anual del 10%.

Esta variación en el número de alumnos que acceden a la universidad gallega se ha visto influida por varios factores que han contribuido a la explosión de la demanda. En primer lugar, cabe señalar la magnitud del cambios que se ha producido en las últimas décadas en la enseñanza secundaria. Donde se ha pasado de una tasa de escolarización del 53% de la población, en el grupo de edad de 16 a 18 años en 1980, al 77% en 1995. En segundo lugar, este incremento, también, se ha visto favorecido por el apoyo económico que a nivel individual se podría concretar en el aumento de las subvenciones a los estudiantes en forma de becas. El porcentaje de estudiantes becarios ha pasado del 10% en la década de los 80 al 20% en los años 90. Por último, ha tenido un efecto muy importante el acercamiento geográfico de los centros universitarios, con la consabida reducción en los costes individuales por los conceptos de alojamiento, manutención y transporte.

A título ilustrativo, es necesario recordar que en el período analizado el distrito universitario gallego pasa de tener una universidad en Santiago de Compostela con Colegios Universitarios, Escuelas Técnicas Superiores, y Escuelas Universitarias (ciclo corto) en las cuatro provincias gallegas (la Coruña, Lugo, Orense, Pontevedra) y Vigo a tres universidades: Universidad de Santiago con campus también en Lugo, Universidad de La Coruña con campus asimismo en Ferrol y Universidad de Vigo con campus en Pontevedra y Orense. (Ley 11/1989, de 20 de julio, Ley de Ordenación del Sistema Universitario de Galicia).

En vista del enorme esfuerzo de financiación pública realizado, la pregunta que nos hacemos es si la composición de los alumnos universitarios gallegos se sigue viendo influida por las circunstancias del entorno familiar más próximo.

Para realizar este análisis se ha considerado el nivel de estudios de los padres(San Segundo y Valiente, 1995) como una aproximación para medir la influencia del medio familiar en los estudiantes cuando toman la decisión de matricularse en la universidad o de abandonar el proceso educativo.

La utilización de este indicador para medir la pertenencia a una determinada clase social, nos permite analizar si el gasto público en él que está incurriendo el gobierno en materia educativa es un medio adecuado para mejorar la igualdad de oportunidades, o tan sólo sirve para reducir el coste de la educación de algunos jóvenes perpetuando las desigualdades existentes.

En el cuadro 1 se describe la evolución del número de alumnos matriculados en el distrito universitario gallego clasificados por el nivel de estudios del padre. En el curso 1990/91 ascendían a 59.767 alumnos, en 1993/94 eran 78.921 y en 1996/97 alcanzan la cifra de 95.304. Los datos presentados confirman que la participación de los estudiantes cuyos padres tienen el nivel de estudios más elevados se reduce respecto al total. En el curso 90/91 los hijos de universitarios representaban el 24,56% y en 1996/97 se ha reducido su participación al 21,81%. En el grupo perteneciente a padres con el menor nivel de enseñanza también se ha producido un cambio significativo. Los hijos de padres analfabetos y sin estudios han pasado de representar el 4,15% del total de alumnos matriculados en el curso 90/91 al 6,8% en el último año analizado.

CUADRO 1
DISTRIBUCIÓN DE LOS ESTUDIANTES CLASIFICADOS SEGÚN LOS ESTUDIOS DEL PADRE. DISTRITO UNIVERSITARIO DE GALICIA (ESPAÑA).

EDUCACIÓN DEL PADRE	1990/91		1993/94		1996/97	
	ALUMNOS	*%*	*ALUMNOS*	*%*	*ALUMNOS*	*%*
ANALFABETO Y SIN ESTUDIOS	2.480	4,15	5.257	6,66	6.486	6,80
PRIMARIOS	24.182	40,46	31.349	39,72	37.623	39,48
BACHILLERATO ELEMENTAL	10.401	17,40	13.027	16,50	15.688	16,46
BACHILLERATO SUPERIOR	8.024	13,43	11.482	14,55	14.722	15,45
DIPLOMADO	7.878	13,18	8.932	11,31	10.602	11,12
LICENCIADO	6.802	11,38	8.874	11,26	10.183	10,69
TOTAL	59.767	100,00	78.921	100,00	95.304	100,00

FUENTE: CONSEJERÍA DE EDUCACIÓN Y ORDENACIÓN UNIVERSITARIA (VARIOS AÑOS)- ELABORACIÓN PROPIA.

A partir del análisis del cuadro 1, es posible afirmar que para el período estudiado la distribución de alumnos cuyos padres tienen un nivel educativo elevado (bachiller superior, diplomado o licenciado) se mantiene estable a lo largo del periodo, con el mismo peso relativo (37% respecto al total). Los datos obtenidos también confirman que no se ha modificado la situación para los alumnos cuyos padres tienen el nivel de enseñanza más bajo. Sin embargo, como ya hemos comentado anteriormente, ha existido una cierta redistribución interna a favor de los hijos de padres analfabetos y sin estudios.

Además, los datos obtenidos reflejan que el 62,74 por ciento de los estudiantes matriculados en el sistema universitario gallego tienen padres con niveles educativos bajos. Esta situación es indicativa de que muchos universitarios están cursando estudios sin contar con un entorno familiar de apoyo. Y de que está teniendo lugar una mejora educativa en la sociedad.

En el cuadro 2, se compara la distribución de la población masculina de 45 a 64 años por nivel educativo con la distribución de estudiantes clasificados según los estudios de los padres, en el último año de referencia 19996/97. El objetivo es investigar si en la población estudiantil están representados los jóvenes en proporción a la importancia de cada colectivo en la estructura poblacional, cualquiera que sea el nivel de estudios de sus progenitores.

CUADRO 2
DISTRIBUCIÓN DE LOS ESTUDIANTES CLASIFICADOS SEGÚN LOS ESTUDIOS DEL PADRE, COMPARADO CON LA DISTRIBUCIÓN DE LA POBLACIÓN MASCULINA ENTRE 45 Y 64 AÑOS. DISTRITO UNIVERSITARIO DE GALICIA (ESPAÑA) 1996.

EDUCACIÓN DEL PADRE	*% ESTUDIANTES*	*% POBLACIÓN*	*DIFERENCIA*
ANALFABETO Y SIN ESTUDIOS	6,80	26,6	-19,8
PRIMARIOS	39,48	50,2	-10,72
BACHILLERATO ELEMENTAL	16,46	10,4	-6,06
BACHILLERATO SUPERIOR	15,45	5,2	+10,25
DIPLOMADO	11,12	3,7	+7,42
LICENCIADO	10,69	3,9	+6,79

FUENTE: CONSEJERÍA DE EDUCACIÓN Y ORDENACIÓN UNIVERSITARIA (VARIOS AÑOS)- ELABORACIÓN PROPIA.

El cuadro 2, analiza la distribución de estudiantes según el nivel de enseñanza de los padres y se compara con los datos de la Encuesta de Población Activa (EPA, INE) del año 1996. Es necesario clarificar que, debido a la información estadística disponible, se está considerando toda la población masculina entre los grupos de edad señalados anteriormente, sin tener en cuenta si tienen hijos universitarios o no.

A la vista de la evidencia contenida en el cuadro 2, se observa que para el grupo de sin estudios y nivel de estudios primarios la diferencia entre la distribución de estudiantes universitarios y la de población es negativa (19,8 y 10,72). Esta dato indica una infrarrepresentación de los padres con bajos niveles de estudios respecto a los jóvenes matriculados en la universidad.

La decisión de no acceder a la universidad por parte de estos colectivos de estudiantes, puede estar motivada por diferentes causas. La más importante es que un alto porcentaje de jóvenes procedentes de familias con bajo nivel cultural no consiguen concluir los estudios de nivel medio. En esta situación, la política educativa de concesión de becas y préstamos no es la más adecuada para cambiar el estado actual de desigualdad. También existen problemas de falta de información a las familias sobre las ayudas al estudio. Y, por último, la función de bienestar de la unidad familiar puede no ser compatible con el objetivo de más educación para los hijos. Ciertamente, el bajo nivel de estudios de los padres incide negativamente sobre la valoración atribuida por los hijos a continuar el proceso de educación formal.

Por lo que respecta a los demás grupos la diferencia entre la distribución de estudiantes y la de población es positiva. Donde se evidencia una mayor sobre-representación es en el nivel más elevados de educación. El grupo de universitarios diplomados y licenciados son el 7,6% de la población masculina de 45 a 64 años y la participación de los alumnos alcanza el 21,81 % del total (diferencia positiva de 14,2).

El análisis realizado permite afirmar que las mayores posibilidades de acceder a la universidad por parte de los jóvenes son tener un padre con titulación universitaria. Una idea generalmente aceptada en el entorno familiar de padre universitario es el hecho de que el nivel educativo influye en el nivel futuro de ingresos.

En el cuadro 3 se analiza la evolución del número de alumnos matriculados clasificados por el nivel de estudios de la madre. En los datos presentados, se observa qué la participación de los jóvenes cuya madre tiene estudios universitarios se mantiene casi estable en el período estudiado (16,1% en 1990/91 y 15,7% en el año 96/97). Donde, sin embargo, se produce un cambio significativo es en el nivel de estudios más bajo, al igual qué ocurría al analizar el nivel educativo del padre.

CUADRO 3
DISTRIBUCIÓN DE LOS ESTUDIANTES CLASIFICADOS SEGÚN LOS

ESTUDIOS DE LA MADRE. DISTRITO UNIVERSITARIO DE GALICIA (ESPAÑA)

EDUCACIÓN DE LA MADRE	1990/91		1993/94		1996/97	
	ALUMNOS	*%*	*ALUMNOS*	*%*	*ALUMNOS*	*%*
ANALFABETO Y SIN ESTUDIOS	3.313	5,54	6.745	8,55	7.963	8,36
PRIMARIOS	29.426	49,23	37.805	47,90	44.834	47,04
BACHILLERATO ELEMENTAL	11.302	18,91	13.432	17,02	16.009	16,80
BACHILLERATO SUPERIOR	6.083	10,18	8.943	11,33	11.541	12,10
DIPLOMADO	7.295	12,21	8.622	10,93	10.822	11,36
LICENCIADO	2.348	3,93	3.374	4,27	4.135	4,34
TOTAL	59.767	100,00	78.921	100,00	95.304	100,00

FUENTE: CONSEJERÍA DE EDUCACIÓN Y ORDENACIÓN UNIVERSITARIA (VARIOS AÑOS)- ELABORACIÓN PROPIA

Los estudiantes con madre analfabeta o sin estudios han pasado de representar el 5,54 por ciento en 1990/91 al 8,36 por ciento en el último año analizado. Pero el grupo con bajo nivel de estudio ha reducido globalmente el ritmo de crecimiento, pasando de representar el 73,68 por ciento en el año 90/91 al 72,2 por ciento en 1996/97. Existe un empeoramiento en la distribución de los estudiantes que proceden de familias donde la madre tiene menor nivel educativo. Sin embargo, la información facilitada por el cuadro 3 muestra una reducción de las desigualdades en las estructuras socio-familiares más bajas y una importante mejora educativa conseguida por estas familias en una generación.

En el cuadro 4, se analizan los datos del curso académico 96/97 con la distribución poblaciónal de mujeres de 45 a 64 años de edad clasificadas por nivel de estudios (EPA, INE, 1996). De nuevo es necesario aclarar que debido a la información estadística disponible, se está considerando toda la población femenina entre los grupos de edad señalados anteriormente, sin tener en cuenta si tienen hijos universitarios o no.

CUADRO 4
DISTRIBUCIÓN DE LOS ESTUDIANTES CLASIFICADOS SEGÚN LOS ESTUDIOS DE LA MADRE, COMPARADO CON LA DISTRIBUCIÓN DE LA POBLACIÓN FEMENINA ENTRE 45 Y 64 AÑOS. DISTRITO UNIVERSITARIO DE GALICIA (ESPAÑA) 1996

EDUCACIÓN DE LA MADRE	*% ESTUDIANTES*	*% POBLACIÓN*	*DIFERENCIA*
ANALFABETO Y SIN ESTUDIOS	8,36	34,4	-26,04
PRIMARIOS	47,04	51,1	-4,06
BACHILLERATO ELEMENTAL	16,80	7,6	+9,2
BACHILLERATO SUPERIOR	12,10	2,4	+9,7
DIPLOMADO	11,36	3,4	+7,96
LICENCIADO	4,34	1,1	+3,24

FUENTE: INSTITUTO NACIONAL DE ESTADÍSTICA (INE) ELABORACIÓN PROPIA

A la vista de los datos estadísticos obtenidos, se puede observar que la participación de los jóvenes universitarios correspondientes al grupo poblacional de mujeres sin estudios es muy inferior respecto de lo que les correspondería con relación a la proporción de población femenina entre 45 y 64 años. La diferencia es negativa, por tanto, están infra-representados en la universidad (-26,04). Lo mismo ocurre con los estudiantes cuyas madres tienen estudios primarios, que representan el 51,1 por ciento de la población, mientras que, los universitarios son el 47,04 por ciento del total de estudiantes, lo que da una diferencia negativa de 4,06.

En el caso de nivel de estudio desde bachiller elemental hasta enseñanza universitaria las diferencias son positivas. La proporción de mujeres diplomadas y licenciadas son el 4,5 por ciento de la población femenina de 45 a 64 años, y la participación de los jóvenes matriculados en la universidad es del 15,7 por ciento (la diferencia es 11,2). El análisis realizado permite confirmar que el acceso a la universidad, por parte de los jóvenes, tiene un determinante importante, y es tener una madre universitaria.

Conclusiones

En el análisis realizado hemos evidenciado el importante papel que el gobierno ha desempeñado en la expansión de la educación superior. Los efectos de externalidades y crecimiento económico que produce la educación han favorecido tanto la financiación pública como la regulación de las actividades educativas.

La extensión de la enseñanza universitaria en España en el ultimo cuarto de siglo ha sido de gran magnitud, estando matriculados en la universidad mas de millón y medio de alumnos en 1995.La tasa de escolarización universitaria se aproxima al 25% en el grupo de edad de 18 a 24 años. España ocupa uno de los primeros lugares entre los países de la OCDE en la escolarización a este nivel. Al mismo tiempo, los programas de becas y prestamos a los estudiantes se han incrementado en las ultimas décadas. El porcentaje de población estudiantil becaria se duplicó (de 10% en 1980 a 20% en 1990).

En el análisis particularizado de la educación superior en Galicia cabe destacar: En primer lugar, la tasa de acceso a la universidad de los jóvenes cuyos padres son analfabetos o sin estudios ha mejorado sensiblemente (del 4,15% en 1990/91 al 6,8% en 1996/97). No obstante, también se ha podido constatar que las desigualdades sociales siguen persistiendo, ya que los hijos de titulados universitarios tienen una tasa de participación del 21,81% en el curso académico 1996/97.

En segundo lugar, el análisis realizado permite concluir que en el ultimo año investigado el 62,74% de los universitarios gallegos, procedían de familias donde el padre tiene un nivel educativo bajo, esta situación es indicativa de que muchos jóvenes no tienen un entorno familiar de apoyo al estudio.

En tercer lugar, si comparamos la distribución de jóvenes según el nivel de estudios de los padres con la estructura poblacional de los varones de 45 a 64 años por nivel de estudios en 1996, los rasgos mas destacados son la infrarrepresentacion de los jóvenes cuyos padres tienen bajos niveles educativos (sin estudios - 19,8 primarios -10,72 y bachillerato elemental -6,06).

Sin embargo, donde se observa una elevada representación de alumnos con diferencias positivas es en los niveles altos de educación de los padres (10,25, 7,42 y 6,79). Esta situación hace evidente que la educación familiar afecta de forma apreciable a la valoración que los jóvenes dan a la enseñanza universitaria.

En cuarto lugar, por lo que respecta a la tasa de acceso de los estudiantes según el nivel de estudios de la madre, se mantiene la misma distribución que ocurría respecto a los padres, con algunas matizaciones que es necesario reseñar. Se produce un cambio significativo al analizar el nivel de cualificación de las mujeres. Se

observa que el nivel de stock de capital humano de la población adulta en Galicia es muy bajo, es decir, casi las tres cuartas partes de los alumnos que acceden a la universidad, proceden de hogares donde la madre tiene como máximo el nivel de estudios medios (el 73,68% en 1990/91 y el 72,2% en 1996/97).

En quinto lugar, los datos obtenidos permiten afirmar que el porcentaje de alumnos cuya madre tiene nivel de estudios de bachillerato superior o universitario está sobrerrepresentado con diferencias positivas respecto a la distribución de la población femenina (9,7, 7,96 y 3,24) en 1996.

Finalmente, el análisis realizado permite afirmar que en la tasa de acceso a la educación superior en Galicia se ha producido un cierto movimiento social, pero la igualdad de oportunidades de entrar en la universidad está muy lejos de alcanzarse.

Seguramente las verdaderas barreras de entrada son anteriores, es decir en la educación secundaria. Por ello, la política de financiación universitaria actual no podrá alcanzar el objetivo de igualdad de acceso, si los medios que se emplean son los de fijar precios de matrícula inferiores al coste real, ampliar indiscriminadamente las becas a los estudiantes, y conceder generosas transferencias a las instituciones educativas. Bajo este contexto, es necesario realizar una revisión del sistema de financiación pública de la educación superior, donde los recursos se asignen de acuerdo con los principios económicos de equidad y eficiencia.

Referencias

Adams, *W.(1977).* Economic Problems Confronting Higher Education. Financing Public Higher Education. *American Economic Review. 67: 86-89.*

Aguiar, I. et al.(1995). *Financiación de la educación superior Especial referencia a Canarias.* Dirección General de Universidades e Investigación Gobierno Canario.

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of Economic Literature,* 19, 4: 483- 536.

Anderson, A. and Bowman, M.J. (eds.) (1965). *Edúcation and economic development.* Chicago. Aldine.

Arrow, K. J. (1973). Higher education as a filter. *Journal of Public Economy.* 2, 3: 193-216.

Baumol, W. J. and Baumol, H. (198l). *Book review: The economics of the performing arts.* New York, St. Martins's Press.

Becker, G.(1960). Underinvestment in college education? *American Economic Review,* 50: 346-354.

Becker, G. (1962). Investment in human capital: A theorical analysis. *Journal of Political Economy,* 70, 5: 9- 49.

Becker, G. (1964). *Human capital. A theoretical and empirical analysis.* New York. National Bureau of Economic Research. Columbia University Press.

Becker, G. (1967). *Human capital and the personal distribution of income: An analytical approach.* Ann Arbor: University of Michigan.

Becker, G. (1975). *Human capital- A theoretical and empirical analysis, with special reference to education.* New York and London. National Bureau of Research.

Benhabib, J. and Spiegel, M. (1994). The role of human capital in economic development. *Journal of Monetary Economics,* 34: 143-173.

Bishop, J. (1977). The effects of public policies on the demand for higher education. *Journal of Human Resources, 12*: 285-307.

Blaug, M. (1966). An economic interpretation of the private demand for education. *Economica,* 33,130:166-182.

Blaug, M. (1966). *Economics of education: A selected annotated bibliography.* London. Pergamon Press.

Blaug, M. (1972). *Economía de la Education.* Textos Escogidos. Madrid: Tecnos.

Blaug, M. (1976). The empirical status of human capital theory: A slightly jaundiced survey. *Journal of Economic Literature, 14,* 3: 827-855.

Blaug, M. (1981). *Educación y empleo.* Madrid: Instituto de Estudios Económicos.

Blaug, M. (1982). *Introducción a la Economía de la Educación.* Mexico: Aguilar.

Blaug, M. y Moreno, J.L.(1984). *Financiación de la educación superior en Europa y España.* Madrid. Siglo Veintiuno.

Bosch, E y Diaz J.(1988). *La educación en España. Una perspectiva económica.* Barcelona: Ariel.

Bowen, G. W.(1977). Economic Problems Confronting Higher Education: An Institutional Perspective. *American Economic Review;* 67: 96- 100.

Calero, J. (1993). *Efectos del gasto público educativo. El sistema de becas universitarias.* Barcelona: Universidad de Barcelona.

Cheit, F. E. (1977). The Benefits and Burdens of Federal Financial Assistance to Higher Education. *American Economic Review,* 67: 90-95.

Cohn, E.(1975). *The economics of education.* Lexington Massachussetts. Lexington Books

Cohn, E. y Geske, T. G.(1990). *The Economics of Education.* Nueva York: Pergamon Press.

Consejo de Universidades (1989). *La financiación de la enseñanza superior.* Madrid: Consejo de Universidades.

Consejo de Universidades (1995). *Informe sobre financiación de la Universidad.* Madrid. Consejo de Universidades.

Figuerola, J. (198 l). *La formación del capital humano.* Madrid: CECA.

Frank, R. H.(1978). Why women learn less: The theory and estimation of differential overqualification. *American Economic Review,* 68: 371-373.

Freire Seoane, M. J.(1991). El acceso a la educación superior en Galicia. *Orientación Económica y Financiera,* 196, 29-35.

Freire Seoane, M. J. (1991). La demanda de educación superior en la Comunidad Autónoma Gallega: Un intento de aproximación cuantitativa. *Información Comercial Española,* 695; 123-133.

Freire Seoane, M. J. (1996). Un análisis de la demanda de educación superior en Galicia a partir de datos cross- section. *Cuadernos de Economía, 66:197-219.*

Freire Seoane, M. J. y Salcines Cristal, J. V.(1997). La equidad en la educación: Un análisis de la educación postobligatoria en la Comunidad Autónoma Gallega *Cuadernos de Economía, 67: 463-490.*

Freire Seoane, M. J. y Salcines Cristal, J.V. (1999). Elección educacional entre escuela pública y escuela privada en Galicia: Un análisis comparativo del gasto de los hogares en las escuelas públicas y privadas.

Friedman, M.(1962). *Capitalism and freedom.* Chicago: University of Chicago Press.

Friedman, M. (1968). The higher schooling in America. *Public Interest,* 11: 108-112.

Fuller, W. et al. (1982). New evidence on the economic determinants of postsecondary schooling choices. *Journal of Human Resources, 17,4: 478-*495.

Grao, J. (comp.) (1988). *Planificación de la educación y mercado de trabajo.* Madrid. Narcea,

Hallak, L.(1991). *Invertir en el futuro. Definir las prioridades educacionales en el mundo en desarrollo.* Madrid: Tecnos-Unesco.

Hansen, W.L. and Weisbrod, B.A. (1969). *Benefits costs and finance of public Higher Education.* Chicago: Markham Publishing Company.

Hope, C. (1983). Postsecundary education enrollment responses by recent high school graduates and older adults. *Journal of Human Resources,* 18, 2: 247-267.

Kodde, D.(1986). Uncertainty and the demand for education. *Review of Economic and Statistc,* 68: 460-467.

Levin, M. R. (1970). Educational investment in an urban society costs, benefits and public policy. *Teachers College Press* 425.

Miller, H. P. and Glick, P.C. (1956). Educational level and potential income. *American Sociological Review,* 21: 307-312.

Mincer, L.(1958). Investment human capital and personal income distribution. *Journal of Political Economy,* 66: 281-302.

Mincer, L. (1962). On the job training: Costs, returns and some implications. *Journal of Political Economy,* 70: 50-79.

Molto, T. y Oroval, E. (eds.)(1982). *Financiación de la enseñanza superior* Barcelona: ICE. Universidad de Barcelona.

Molto, T. y Oroval, E. (1984). *Costes y rendimientos de la enseñanza superior.* Barcelona: ICE. Universidad de Barcelona.

Mora, J.G. (1989). *La demanda de educación superior.* Madrid: Consejo de Universidades.

Mora, J.G. (1991). *Calidad y rendimiento en las instituciones universitarias.* Madrid: Consejo de Universidades.

Mora, J. et al.(1993). *La financiación de las universidades valencianas.* Valencia: Alfons el Magnánim.

Musgrave, R. A.(1959). *The theory of public finance,* New York, Mc. Graw-Hill.

OCDE (I *990). Financing higher education.* Paris. OCDE.

OCDE (1997). *Education at a* glance. Paris. OECD.

Oroval, E.(edit.) (1995). *Planificación, evaluación y financiación de sistemas educativos.* Barcelona. Civitas.

Pissarides, C. H.(1982). From school to university: the demand for post-compulsory education in Britain. *Economic Journal,* 92: 654-667.

Psacharopoulos, G. (1973). *Returns to education: An international comparison.* San Francisco. Elsevier - Jossey Bass,

Psacharopoulos, G. (1975). *Earnings and education in OECD countries.* París: OCDE.

Psacharopoulos, G.(edit.) (1987). *Economics of Education: Research and studies.* Oxford: Pergamon Press.

Psacharopoulos, G. (1993). *Returns to investment in education: a global up date.* Washington D. C. Office of the Director, Latin América and the Caribbean, World Bank.

Psacharopoulos, G. and Ying, Ch. (1992). *Earnings* and *education in Latin America: assessing priorities for schooling investments.* Washington D.C. Technical Department, Latin America and the Caribbean, World Bank.

Psacharopoulos, G. and Woodhall, M.(1987). *Educación para el desarrollo. Un análisis de opciones de inversión.* Madrid: Tecnos. Banco Mundial.

Quintas, J.R. (1983). *Economía y educación.* Madrid: Pirámide.

Quintas, J.R. y Sanmartin, J. (1978). Aspectos económicos de la educación: comparaciones internacionales. *Información Comercial Española,* 537; 37-47.

Revista de Estudios Regionales (1993). *Jornadas sobre enseñanza superior: financiación y empleo.* Un enfoque regional ICE, 36. Universidades de Andalucía.

San Segundo, M.J. (1995). La *intervención publica en* la *educación la financiación de la enseñanza obligatoria,* en Barbera, S.(editor). Estado y economía. Fundación BBV

San Segundo, M. J. y Valiente, A.(1995). *La demanda* de *educación superior y la financiación universitaria.* Economía de la educación, Colegio de Economistas, Las Palmas.España,

Schultz, T.(1959). Investment in man: A economist view. *Social Service Review,* 33:109-117.

Schultz, T. (1960). Capital formation by education. *Journal of Political Economy,* 68: 571-584.

Schultz, T. (1968). El v*alor económico de la educación.* México, Uteha.

Schultz, T. (1980). *Invirtiendo en la gente.* Barcelona: Ariel.

Stiglitz, J. E.(1 988). *La economía del sector publico.* Barcelona. Bosch.

Acerca del Autor

Mª Jesús Freire

Facultad de Ciencias Economicas.
Universidad de La Coruña **(España)**

Spain
E-mail: maje@udc.es

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
aacosta@cucea.udg.mx

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

Este artículo ha sido consultado 172 veces desde el 1 de mayo de 2000

Education Policy Analysis Archives

Volume 8 Number 21 mayo 1, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

El Sistema de Selección de Alumnos de las Universidades Chilenas: Discusión de sus Fundamentos, Resultados y Perspectivas

Sebastián Donoso
Gustavo Hawes
Instituto de Investigación y Desarrollo Educacional
Universidad de Talca
Chile

Resumen

Chile posee la particularidad de disponer de un sistema centralizado de selección de alumnos de pregrado de las universidades de mayor tradición, representadas en el Consejo de Rectores. Este proceso tiene más de treinta años de vigencia continua en sus principales instrumentos y procedimientos así como de sus fundamentos. En la actualidad se encuentran en revisión los contenidos de las distintas pruebas que se aplican, aunque todo parece indicar que el sistema continuará siendo el mismo. La propuesta de cambios proviene de las nuevas condiciones creadas en la educación superior chilena a partir de la Reforma del año 1981, que introdujo modificaciones significativas en la constitución de los planteles, su organización y financiamiento. Se abrió de manera súbita el sistema universitario, hasta entonces basado exclusivamente en ocho universidades, a un conjunto que en la actualidad suma mas de sesenta y cinco. Paralelamente se ha generaron nuevas condiciones y demandas sociales, profesionales y técnicas que han repercutido en el sistema de educación superior en su conjunto y en el proceso de selección de alumnos, las que presionan por cambios más profundos en el sistema de selección que se analiza en este artículo. El texto incluye una descripción del proceso de selección académica vigente para el

ingreso al sistema universitario chileno. A continuación se presenta y analiza la Prueba de Aptitud Académica (PAA), principal instrumento de dicha selección. Finalmente se analiza y critica la PAA desde dos puntos de vista: su pertinencia psicométrica y el modelo de inteligencia implícito.

Abstract
Chile has a peculiar centralised system for the selection of students to undergraduate programs; this is the case for the more traditional universities associated in the Council of Rectors. This process has been in operation for over thirty years, and its major instruments and procedures, as well as its foundations, are still in force. The contents of the different tests are currently under review; however, the system will continue to be the same in the future. Changes come from the new conditions created by the 1981 reform of the Chilean higher education system. Important modifications were introduced in the constitution, organization and financing of universities. The university system was suddenly opened; from the original eight universities, the system expanded to sixty-five. New conditions and social, professional and technical demands are having an impact on the higher education system as a whole and, specifically, on the student selection processes. This article includes a description of the process of academic selection for entrance into the Chilean university system. Next we analyze the Test of Academic Performance (PAA), the main instrument of this selection. Finally the PAA is analyzed and criticized from two points of view: its psychometric relevance and its implicit model of intelligence.

1. Cambios e inconsistencias en el Sistema Universitario Chileno

La reforma de la educación superior chilena del año 1981 es un detonante de cambios que comprenden tanto su estructura como, en otro nivel, la concepción misma de lo que se entiende y concibe por educación superior. Al respecto, se crean nuevos planteles universitarios a partir de las universidades más antiguas (Nota 2) que configuraron un escenario diferente. De un sistema que hasta 1980 estaba basado en ocho universidades con Sedes a lo largo del país, se pasa en 1999 a uno que reúne sesenta y seis universidades, dentro de las cuales se encuentran las ocho iniciales, más otras creadas a partir de las sedes que las universidades antiguas tenían en regiones, las que en conjunto con las anteriores conforman el Consejo de Rectores de las Universidades Chilenas con un total de veinticinco instituciones, y un tercer grupo compuesto por las universidades generadas directamente a partir de la legislación de la Reforma señalada.

A su vez la Reforma en comento implicó definir un conjunto de carreras reservadas exclusivamente para las universidades. Esto se tradujo inmediatamente en la generación de un sistema estratificado que inicialmente consideró doce carreras y en la actualidad incluye diecisiete (Nota 3).

Por otra parte, la Reforma también dio lugar a la aplicación de criterios diferentes de financiamiento para la educación superior en Chile. Se pasó de un sistema de financiamiento de la oferta a uno basado en la demanda. Entre las implicaciones más importantes que trajo consigo este cambio, está que relacionó directamente los puntajes de los alumnos en la Prueba de Aptitud Académica, que es la prueba principal del sistema de selección, con el financiamiento de las instituciones, lo que finalmente derivó, tal como se analiza en el trabajo, en distorsiones de los criterios de admisión.

De esta forma, el sistema universitario chileno transitó en una década desde una estructura cerrada a un mercado abierto, liberalizado en su sentido más lato y sujeto

a los controles de los mismos agentes. Sin embargo este proceso mantiene pendiente algunas discusiones derivadas del enfrentamiento de proyectos distintos en un escenario que soporta altas simultaneidades, algunas de estas disonancias tienen directa relación con el tema en análisis.

Una de las principales desavenencias conceptuales hasta hoy en discusión proviene de la concepción ideológica del Estado Docente, que permea los patrones culturales de la representación del Estado como el proveedor de todos los servicios para la población y hace cargar sobre el aparato estatal la obligación de mantenerlos funcionando en un buen nivel de eficiencia y eficacia. En Chile, el elevado y a veces sobredimensionado valor del impacto de la educación como determinante del empleo y de los ingresos económicos, genera mayor demanda de la población juvenil por ingresar al sistema Universitario. Al detectarse que los establecimientos municipales (que perciben financiamiento estatal), de la cual egresan aproximadamente el 65% de la población del país, tienen rendimientos notoriamente inferiores comparados con los colegios particulares subvencionados por el Estado o pagados directamente, la mirada nuevamente se vuelve hacia el Estado al que se hace garante, aval y responsable de la equidad o derecho para todos los jóvenes chilenos a recibir una educación media de la mejor calidad, que les permita competir en pie de igualdad.

Sin embargo las orientaciones políticas que han inspirado y conducido los procesos de transformación del sistema educacional tienden a una dirección inversa. El Estado se ha desposeído de la educación básica y media a través del proceso de "municipalización" y de privatización de la enseñanza; el cual en la educación superior ha sido mucho más fuerte y radical, privilegiando la dinámica del libre mercado como eje articulador y auto-regulador del sistema de educación superior (Nota 4). Una prueba de ello es que la oferta de matrículas para primer año pasó en un plazo de 18 años, de una asimetría de más de tres postulantes por vacantes a una relación prácticamente igual respecto de la demanda, correspondiendo a las universidades privadas el mayor crecimiento. De esta forma puede señalarse que el sistema de selección que se analiza tiene una relación de identidad directa entre aquellos que finalmente están en condiciones de postular efectivamente a la Universidad con las vacantes que este segmento de universidades dispone. Otra disonancia proviene del rol de articulación que se espera que debiera existir entre la educación media y la universitaria, y de la consistencia del sistema de selección como predictor de los rendimientos futuros. Al hablarse del sistema de selección a las universidades (cuyo sinónimo es también "Prueba de Aptitud Académica" o PAA) se insiste sobre la necesaria y obligada articulación (inmediata) entre la enseñanza media y la superior, asignando al sistema de selección el papel de nexo evaluador de aquélla, lo que es asumido por la construcción de las pruebas). Esta situación deriva en una confusión aceptada socialmente, que se observa en los medios de comunicación, en los padres y estudiantes y, particularmente, en los mismos establecimientos educacionales, tiene importancia, toda vez que éstos son calificados como buenos, regulares o malos según los resultados de sus egresados sobre un conjunto de pruebas que miden sólo un aspecto de los logros educacionales, con las consiguientes consecuencias positivas o negativas que tiene sobre la demanda por matrícula y los niveles de holgura financiera del establecimiento.

Otra importante fuente de ambigüedad está representada por el mercado de las vacantes. El actual sistema de selección de estudiantes surgió como iniciativa de la Universidad de Chile, a la que se sumaron las restantes universidades del país, pasando a ser un proceso nacional (Nota 5). Sin embargo, las condiciones de aquella época eran claramente diferentes: junto a una oferta de vacantes reducida se encontraba una enseñanza media también reducida. La reforma educacional de la enseñanza básica y media del año 1965 significó una explosión demográfica en la enseñanza secundaria y, por consiguiente, en el número de egresados que demandaban educación terciaria, a la que el sistema universitario respondió con un pequeño aumento en las vacantes ofrecidas.

En la actualidad, sin embargo, la desregulación del sistema de educación superior chileno, generado a partir de la legislación de 1981, no ha sido capaz de establecer una auto regulación en cuanto a la explosión carreras y de títulos profesionales ni al aumento de la oferta de vacantes. Este fenómeno es complejo, pues hoy en día se estima superada la etapa de asumir el mercado como una fuente ilimitada de postulantes, ya que desde hace cinco años a la fecha éste se ha estabilizado, sin embargo siguen aumentando los competidores institucionales. La selección, entendida en cuanto tal, se hace cada vez más difícil dado que hay ofertas similares de otras universidades no adscritas la Consejo de Rectores (es decir creadas con posterioridad a 1980), que no requieren pasar estos complejos procesos de selección de estudiantes.

Pese a lo expuesto el ámbito de la selección de alumnos a la educación superior podría ser visto como un tema parcial, técnico y árido, destinado en lo medular a revisar sólo una parte pequeña, como son las pruebas de selección, de un gran asunto como es la Universidad. Sin embargo, no es menos cierto que tras el proceso de selección y admisión convergen una serie de elementos importantes de revisar, algunos de los cuales se refieren a que: (i) la selección no es un acto neutro, sino que se realiza en referencia a determinados criterios y valores; (ii) la selección significa, en el caso de los postulantes, la posibilidad de acceder a una profesión, y por ende a una serie de consecuencias para el resto de su existencia; (iii) los procesos de selección conllevan explícitamente consecuencias sobre el financiamiento de las diversas instituciones de educación superior; y (iv) de manera indirecta, las pruebas sustentan un juicio sobre la calidad de la enseñanza en los diferentes establecimientos de enseñanza media y a la pertinencia del sistema educacional medio.

De esta manera el análisis del problema del sistema de selección y admisión de postulantes a la Universidad es una materia que necesariamente congrega criterios técnicos, políticas sociales y efectos personales. Ello la hace de suyo un área de gran relevancia tanto para los actores individuales, como para los sociales, entendiéndose por éstos últimos a las instituciones universitarias, el mundo laboral y la sociedad como tal, pues generarán consecuencias sobre tales dimensiones.

2. Del bachillerato a la Prueba de Aptitud Académica

Hasta el año 1966 las universidades chilenas seleccionaban sus estudiantes mediante un conjunto de pruebas conocido como "Bachillerato" (Nota 6) y la combinación de sus puntajes con otros como las calificaciones de enseñanza media y las pruebas especiales, que algunas carreras las incorporaban como requisito. Las pruebas propias del Bachillerato comprendían una parte común que abarcaba las áreas de Comprensión y Redacción, Historia de Chile e Idiomas, y una parte específica que medía conocimientos en las áreas de Letras, Matemáticas o Biología, que eran elegidas por el futuro postulante.

El sistema de selección entonces vigente había sido diseñado hacia 1930 para una masa estudiantil pequeña y para un sistema educacional muy reducido. La expansión de la demanda por crecimiento demográfico del estudiantado de educación media, como también el aumento en número de universidades y de programas académicos ofrecidos, provocó que éste adquiriera una complejidad que superó la racionalidad inicial del sistema de selección, colapsando en su operatividad.

En otra dirección, estudios estadísticos revelaron la baja capacidad predictiva del bachillerato o de las combinaciones con otros puntajes y que las modificaciones que fueron introduciéndose al bachillerato en su última década le redujeron sustancialmente su capacidad predictiva (Grassau, 1956). Los problemas derivados del sistema de corrección por medio de examinadores, y del azar en la selección de temas por parte de los alumnos, contribuyeron notoriamente al agravamiento de los problemas anteriores.

La Universidad de Chile formuló un proyecto (Grassau, 1966) tendiente a superar las deficiencias de las pruebas anteriores y que, junto con mejorar sustancialmente los aspectos teóricos, técnicos y administrativos de las pruebas, permitiera su aplicación a contextos ya masificados y con presencia en todo el país, a raíz del desarrollo de universidades y la creación de los Colegios Universitarios regionales.

Junto con el mejoramiento de las cualidades métricas de los instrumentos, el gran ordenador para desarrollar el actual Sistema de Selección y Admisión fue que todos los egresados de enseñanza secundaria tuviesen similares oportunidades frente a una oferta relativamente reducida de vacantes. A partir de este principio de "igualdad de oportunidades" clave en la comprensión y operación de la política de Estado, se examinan otros dos supuestos centrales del sistema de selección de estudiantes, que tiene una dimensión más técnica: la normalidad e independencia de la distribución de las aptitudes, y su estabilidad.

La postulación se realiza a partir de la ponderación de los resultados (estandarizados) de las siguientes pruebas o antecedentes que tienen carácter de obligatorias (los valores en cada una se expresan entre los 300 y 800 puntos, mayor puntaje significa mejor resultado), estas son: calificación promedio de la educación media; resultados en la prueba de aptitud verbal, de la prueba de aptitud matemática y de la prueba de Historia de Chile; a éstos se pueden agregar, los resultados de pruebas de conocimientos específicos de algunas disciplinas. La ponderación (porcentual) que tiene cada una de éstas para cada carrera según universidad se informa a comienzos del período, de manera que al momento de inscribirse en el proceso, los estudiantes saben las exigencias que deben cumplir al respecto.

2.1. Características y etapas operativas del sistema PAA

Existen tres características operativas importantes para comprender el Sistema de Selección de Alumnos. Primero, se trata de un proceso que se realiza anualmente, cuyos resultados sólo tienen validez en ese contexto, en razón de las propiedades técnicas de las pruebas y forma de calcular los puntajes, es decir son independientes entre sí. La segunda característica es su centralización. Tras rendir las pruebas, el alumno realiza una sola postulación marcando hasta doce opciones de carreras debidamente jerarquizadas. En función de los resultados obtenidos en las variables de selección y requisitos establecidos para las carreras, el sistema: (a) le selecciona

en una de ellas, o (b) le deja en "lista de espera" mientras se producen vacantes o, (c) es rechazado, porque otros con mejores puntajes coparon las vacantes.

La tercera característica es que la selectividad que opera - fundamentalmente- de parte de las Universidades. Rasgo diferente de la mayor parte de los sistemas similares en los cuales el postulante tiene opciones de aceptación para sí, independientes unas de otras, abrigándose realmente la elección entre éstas. En este sistema el participante tiene esa posibilidad sólo en el momento de estructurar su postulación, pero no en la aceptación. Una vez que es asignado a una carrera, como se indicó, no continúa en el proceso. Esta hace que sea la Universidad la que, selecciona al postulante. Dejándole a éste sólo la opción de matricularse.

Este proceso de selección consta de varias etapas, las que han permanecido idénticas desde su diseño inicial. La primera es la inscripción anual para rendir las pruebas, a la que pueden concurrir quienes egresan ese año de la enseñanza media (grupo identificado como "de la promoción") que conforman el 60% de los participantes anuales, o quienes lo han realizado años anteriores (conocidos como "rezagados"), que corresponden al 40% de los participantes. Se inscriben anualmente aproximadamente 150.000 personas.

La segunda etapa del proceso es de carácter absolutamente obligatorio para cualquier carrera de este conjunto de universidades y consiste en rendir tres pruebas: una de aptitud verbal, otra matemática y una prueba de Historia de Chile. Del total de inscritos, un 5% no se presenta a rendir las pruebas quedando inmediatamente marginado del proceso, reduciéndose el grupo a algo más de 140.000 participantes. De este grupo a lo menos un 110 a 12% no rinde las tres pruebas obligatorias reduciéndose el grupo inicial a unos 125.000 participantes.

Existe una tercera etapa (que está inmediatamente contigua a la anterior, que corresponde a la rendición de pruebas de conocimientos específicos). Estas son opcionales como requisito para algunas carreras de distintas universidades, pudiendo cada universidad determinar si va solicitar pruebas de este tipo o no y luego estableciendo cuáles y en que proporción o porcentaje participan de la ecuación final.

Para poder postular a la Universidad se requiere alcanzar un puntaje ponderado mínimo de 450 puntos, obtenido de la relación de ambas secciones de la PAA. A partir de ese puntaje se puede continuar participando de la siguiente etapa. De los 1250.000 participantes son eliminados por no alcanzar el puntaje mínimo para postular aproximadamente el 45% del grupo, en consecuencia un número cercano a los 60.000 está en condiciones reales de realizar su postulación definitiva a las universidades.

De este grupo no todos continúan, esta vez porque su opción de ingreso es muy baja o por otros motivos, reduciéndose la masa real de postulantes a unos 50.000 candidatos, los que finalmente la hacen efectiva. Este proceso culmina con la selección definitiva en las carreras, quedando la opción posterior de matricularse en lo que quedó seleccionado o postular nuevamente el año subsiguiente, reiniciando todas estas etapas descritas.

Como se señaló, las universidades fijan de antemano sus cupos, en consecuencia, completan sus vacantes de acuerdo con el puntaje de los postulantes. Esto implica que aquellas carreras y universidades más prestigiosas llenan sus vacantes con puntajes mejores que otras carreras. En el grupo más selecto de carreras se encuentra Medicina, Ingeniería Civil, Odontología, Derecho Arquitectura, Economía., que suelen tener más de 5 postulantes por vacantes, dependiendo de la universidad que se trate, en las más prestigiosas puede alcanzar a más de 10 postulantes por vacante.

El sistema ha operado de esta manera desde hace más de treinta años, ordenando a los postulantes según los puntajes alcanzados, lo que debiera entenderse como un equivalente del potencial académico que éstos tienen.

2.2. El supuesto de "igualdad de oportunidades"

La capacidad académica de los egresados de la educación media se pretende medir con un instrumental que representa una muestra de los respectivos dominios de conductas que el sujeto elicita en situaciones de prueba. Las preguntas o reactivos han sido elaboradas y seleccionadas especialmente, de forma de obtener los mejores estimadores acerca de la capacidad de aprender en los dominios conductuales examinados. De esta manera, los sujetos se enfrentan a conjuntos de preguntas, algunas de las cuales son (supuestamente) igualmente familiares o bien igualmente inusitadas. El punto importante es que la prueba busca ser igual para todos, tanto en sus reactivos como en los procedimientos para su administración.

Este supuesto no sólo se ubica en la teoría y en la instrumentación de las pruebas de aptitud, sino que trasciende al plano de la política con una especial connotación. En efecto, el carácter "igualitario" aseguraría la igualdad de oportunidades a los egresados de la enseñanza media. Este no deja de ser un argumento insuficiente, toda vez que se amplifica el hecho de que se trata de la misma única medición, llevando inapropiadamente a la conclusión de que los procesos educacionales en la enseñanza media se han desenvuelto con homogéneas características.

Es así como, el supuesto debe ser considerado cuidadosamente desde el punto de vista del contenido de los tests, toda vez que las pruebas y especialmente sus resultados no son independientes de los factores sociales y culturales. De hecho, el mismo concepto de inteligencia es propio de una cultura determinada. Además este afán igualitarista, creemos, confunde igualdad de medición con una igualdad de procesos. Al ser la Prueba de Aptitud Académica un procedimiento estandarizado, en sus resultados se reproducen las diferencias del proceso educativo y sus condicionantes. En contraposición, tampoco puede afirmarse que las exacerba: tan sólo refleja crudamente un sistema educacional desigual. Más aún, en su diseño la Prueba de Aptitud Académica considera la atenuación de los efectos de factores externos sobre las habilidades básicas de los sujetos, tratando de anular en lo posible los condicionamientos contextuales o idiosincrásicos. Por ello, se centra especialmente en las habilidades más que en los contenidos tanto verbales como no verbales. Así, *"la Prueba de Aptitud Académica no es directamente dependiente del nivel socioeconómico de los candidatos"* (Díaz et al., 1988: 316).

La falacia asumida en ese razonamiento reside en hacer sinónimo "los mejores" en rendimiento con "los mejores" en las condiciones en que cursaron la enseñanza media, igualando resultados en la prueba, con condiciones. Esto conduce a la conclusión que las características del medio educativo en que se desenvolvieron residen estructuralmente en los sujetos, lo que significa un determinismo que ya no cabe en el discurso pedagógico. En efecto, podría pensarse que mejorando las condiciones de los sujetos desmedrados, se lograrían resultados iguales o equivalentes. Esto es confirmado por el efecto de igualación que se produce al nivel del segundo año de la educación superior, si bien debe considerarse el impacto atenuador del efecto "selección", por cuanto no hay representación de todo el conjunto de los egresados de la enseñanza media, sino de un segmento de ellos.

2.3. Normalidad e independencia de la distribución de las aptitudes.

Este supuesto constituye la base del modelo factorial de inteligencia (desarrollado por Guilford) que se adoptó en las pruebas. En efecto, los procedimientos utilizados, como el cálculo de correlación producto-momento y la ortogonalización de ejes por rotación en la construcción de factores, no pueden sino entenderse en términos de la normalidad estadística.

En las pruebas se asume que la población de aptitudes se distribuye normalmente (Nota 7) y se la considera independiente de otras variables como sexo, edad, nivel socioeconómico y cultural, entrenamiento, maduración. La posibilidad de generar factores ortogonales permite modelar el concepto de aptitud de tal forma que sus componentes aparezcan en un estado de separación y relativa autonomía uno

respecto del otro. Así, siguiendo el criterio de la aditividad de la varianza, podría pensarse que una aptitud determinada es la combinación específica que toman diferentes variables que se organizan factorialmente para ello.

Las pruebas están diseñadas para que, dentro de lo posible, la medición sea indiferente a los factores contextuales que podrían estar interfiriendo el desempeño de los sujetos y, por ello, introduciendo error de medición en los puntajes (Díaz, Himmel y Maltes, 1990, 316). Esta suerte de "indiferencia" responde perfectamente al ideal de la integralidad del ser humano, toda vez que no desconoce la interrelación estrecha de las diferentes dimensiones de la persona. El análisis distingue, separa y relaciona para proponer finalmente una síntesis que permita la acción sobre la realidad. Pero a la vez buscar hacerlo independiente de los factores culturales es una pretensión sin destino: la indiferencia cultural relega al ser humano a una abstracción.

2.4. Estabilidad de las aptitudes

Cuando la teoría psicológica asevera y luego sostiene que la aptitud es un rasgo estable (Nota 8), puede desprenderse entonces que ésta puede ser evaluada mediante una única medición. Esa es la propuesta del respaldo teórico del modelo de la Prueba de Aptitud Académica: que las habilidades verbal y matemática según las mide la Prueba de Aptitud Académica son de lento desenvolvimiento, por lo cual los factores que determinan la capacidad general de una persona no deberían experimentar modificaciones notorias en un período relativamente breve. Este argumento de la estabilidad es el que permite hacer juicios predictivos a partir de los puntajes.

Una definición de aptitud es la propuesta por Binghman, a saber, *"condición o conjunto de características que se consideran sintomáticas de la capacidad de un individuo para adquirir, a través de un cierto entrenamiento, un conocimiento, habilidad o conjunto de respuestas (generalmente especificados) como, por ejemplo, la capacidad de hablar un idioma, de interpretar música, etc"* (ap. Avila, 1980).

Esta concepción es la que se asume para la Prueba de Aptitud Académica, y sobre ella se basa la predictibilidad de sus resultados. En efecto, *"la Prueba de Aptitud Académica cumple el propósito de entregar información que permita estimar el desempeño futuro de los sujetos a partir de su comportamiento frente a estímulos representativos de las habilidades consideradas necesarias para cursar con buen éxito estudios superiores"* (Avila, 1991, I, 10).

Algunos resultados obtenidos tras la aplicación de las pruebas por más de un cuarto de siglo permiten relativizar el concepto de estabilidad. Los resultados indican que cuando se rinden por segunda vez las pruebas, se evidencian cambios positivos (relativos) en los puntajes obtenidos (Donoso, 1988; 1989). Esto muestra la influencia de otros factores como entrenamiento, en especial, experiencia, efecto de instrumentación y maduración. La estabilidad de las aptitudes no es materia directa de cuestionamiento, pero hay cambios en los puntajes obtenidos en aplicaciones consecutivas.

Una ampliación teórica del campo de la aptitud permite perfilar más claramente lo que se está midiendo. Snow (1988) distingue dos dimensiones de la inteligencia: inteligencia cristalizada e inteligencia fluida, siendo *"las dos clases de inteligencia (...) independientes durante la adolescencia y la edad adulta"* (p. 828). La inteligencia cristalizada se refiere a la formalización de estructuras del pensamiento para diversas finalidades, de forma de conseguir instrumentos útiles de pensamiento y posterior aprendizaje; la transferencia se refiere no sólo al conocimiento específico sino a las estrategias organizadas como procedimientos (habilidades académicas de aprendizaje). El producto se expresa en el desempeño en pruebas de capacidad escolar o académica y de rendimiento. Por su parte, la inteligencia fluida representa los nuevos o re-novados ensamblajes o acoplamientos flexibles para adaptaciones más extremas en situaciones nuevas. Con relación a las medidas de la inteligencia o

aptitud para el rendimiento escolar, Snow señala que, para el caso del SAT como el ACT norteamericanos (Nota 9), ambos pueden interpretarse en una gran parte como medidas de la inteligencia cristalizada, aunque de hecho ninguno representa exclusivamente a este único constructo.

3. La racionalidad del sistema de selección

Para discurrir la racionalidad del actual sistema de selección es importante comprender que todo proceso de admisión requiere de criterios de selección. Estos serán mayormente importantes cuando la demanda de postulantes supere la oferta de vacantes, ya sea por un problema de costo/oportunidad o simplemente de calidad, en determinadas universidades o para carreras específicas. A partir de ello las formas de clasificación de la demanda se traducirán en los criterios de selección de los postulantes.

El sistema de selección en análisis considera dentro de sus partes las pruebas de admisión, que son el conjunto de instrumentos que le permiten clasificar y ordenar a los postulantes. Las pruebas obligatorias que se aplican son: de Aptitud Académica, parte verbal y matemática, Prueba de Historia y Geografía de Chile. Las opcionales son: de Conocimientos Específicos en Biología, Química, Física, Ciencias Sociales y Matemática. También incluye el promedio notas o calificaciones finales de la educación media (4 años). Para algunas carreras se administran pruebas especiales (arquitectura, arte, psicología)

La finalidad de las pruebas es estimar el desenvolvimiento futuro de un individuo partiendo de la información contenida en las respuestas a los estímulos que se consideran representativos de las habilidades que se desea medir (Avila, 1978), lo que revela un claro propósito predictictivo. Una creencia muy arraigada en nuestro medio nacional asume que la universidad debe seleccionar a los mejores alumnos de acuerdo a ciertos criterios (como son las aptitudes verbal y matemática), estableciendo que con ello se mejoran las probabilidades de éxito de los estudiantes. Señala Aranda (1985) que *"las universidades se han interesado por admitir en sus aulas a aquellos alumnos que puedan enfrentar con éxito las exigencias académicas... esta aspiración, válida tanto ayer como hoy, hace necesaria la existencia de un sistema de selección"* (p. 20).

Lo anterior significa fundamentar el derecho que asiste a una institución de educación superior para seleccionar los candidatos que postulan a ser alumno, de acuerdo a criterios originados en las características propias de cada carrera, de los niveles de exigencia planteados, de la capacidad para atender a un número determinado de alumnos, entre otros.

Adicionalmente se argumenta que es preciso disponer de un mecanismo que regule y ordene la demanda para una oferta menor en número (vacantes). El sistema es apropiado y funcional cuando hay una oferta cerrada de vacantes y carreras, frente a una demanda que la excede y cuando se impide la existencia de ofertas alternativas. Sin embargo pierde funcionalidad global cuando la demanda por vacantes iguala la oferta, salvo para algunas carreras muy definidas en donde la selección se rige por criterios absolutos de calidad (como podría ser medicina).

Esta situación experimenta cambios a partir de la reforma de la educación superior de 1981. La oferta de vacantes en variedad y cantidad se ha equilibrado con la demanda. Consecuentemente, el mercado de postulantes no es un mercado cerrado, de donde el concepto de "selección de alumnos" se hace más fluido y relativo, toda vez que no siempre las instituciones de educación superior realmente seleccionan alumnos, sino que ha habido una transferencia tal que ahora el alumno puede hacer con mayores recursos la selección de la universidad o carrera a que desea ingresar (Nota 10).

Coherente con la concepción del Estado Docente, éste busca un equilibrio entre la "igualdad de oportunidades" y el uso de los recursos, como financiador de la educación superior y como responsable superior del desenvolvimiento social y

cultural de la nación. La gratuidad de los estudios universitarios hasta la década de los 80 llevó a que la inversión en el nivel tuviera características de regresividad. La reforma del sistema de 1981, buscó revertir esta características por intermedio de la privatización de la educación superior y la disminución del aporte fiscal directo.

4. Discusión del modelo teórico de inteligencia sustentado por el sistema

Existen diversas maneras de enfocar el constructo "inteligencia" dado que no se trata de un concepto unívoco, además del hecho de que es un concepto sesgado culturalmente. Adicionalmente, no se dispone de un referente fáctico delimitable operacionalmente por un conjunto de indicadores, sobre el cual haya consenso. Se encuentra que "inteligencia" es propiamente un constructo, para el cual coexisten diversas teorías: evolutiva como la de Piaget, fisiológica (Hebb), del aprendizaje como la de Ferguson, estadísticas como Thurstone y Guilford (Reese y Lipsitt, 1980), de procesos, que es la que propone Sternberg (Sternberg y Powell, 1982) y de áreas como Gardner (1995)

El modelo teórico sobre el que se sustenta la Prueba de Aptitud Académica es la teoría de la estructura del intelecto, desarrollada por J.P. Guilford y que corresponde a una noción fundamentalmente estadística (Guilford, 1959, 1982). A partir de los enfoques factoriales de L.L. Thurstone, Guilford desarrolla un modelo de inteligencia tridimensional y de estructura cúbica, formado por unos ciento veinte factores. En este modelo no se encuentra ningún factor común o general. Estos factores independientes se encuentran formados por el cruzamiento de las formas en que pensamos (las operaciones), lo ideado (contenidos) y los resultados de la aplicación de una determinada operación a un determinado contenido (productos).

Un importante planteamiento teórico alternativo es en nuestros días el propuesto por Sternberg y asociados (Sternberg, Conway, Ketron y Bernstein, 1981; Sternberg y Powell, 1987) y también por H. Gardner (1995). En el caso de Sternberg, el enfoque procede a partir del análisis factorial de un conjunto de respuestas emitidas por expertos, y que aísla tres factores, a saber: Inteligencia Verbal, Solución de Problemas e Inteligencia Práctica. En todo caso, el hecho de proceder de las respuestas de un conjunto de expertos no implica necesariamente que las opiniones "factorizadas" de éstos sean coincidentes con la realidad. Sin embargo, este es un conflicto aparentemente insuperable por el momento. Por su parte Gardner (1995) en la teoría de las inteligencias múltiples habla de siete áreas de desarrollo, que forman parte del espectro de inteligencia que las personas poseen, con distintos niveles de desarrollo y complejidad.

El modelo de inteligencia de la Prueba de Aptitud Académica considera, sin embargo, sólo dos factores, a saber la inteligencia verbal y la inteligencia matemática, ya que son "un perfil general que es indispensable para proseguir cualquier estudio de nivel superior. Hoy en día la habilidad para razonar es la condición sine qua non del concepto de inteligencia, pues "razonar implica las capacidades para deducir, abstraer, conceptualizar e inferir" (DAPAA, 1994: 3); esto también es expresado en Díaz, Himmel y Maltes (1988: 315).

5. Estructura de las pruebas y sus características

Las pruebas del sistema de selección han sido confeccionadas considerando los aspectos técnicos relevantes de la teoría de la medición, especialmente las características de confiabilidad y validez que son propias de este tipo de instrumentos. Además, se las administra bajo condiciones de estandarización, lo que colabora a minimizar el error de la medición.

Bajo otra óptica, puede pensarse no sólo en términos de lo que revela sino también de lo que oculta un sistema de selección como el actual. Asociado a ello, se encuentra la característica de gran credibilidad social del sistema de selección y ciertamente de las pruebas que forman parte del mismo.

5.1. Pruebas de aptitud y pruebas de conocimientos

Se consideran dos grupos de pruebas, de aptitud y de conocimientos específicos. Técnicamente, se refieren a distintas cuestiones, con implicancias para los fines de la selección y predicción del desempeño académico de los futuros estudiantes universitarios.

Existen diferencias entre aptitud y conocimiento específico, las que se expresan en características de las pruebas y de sus mediciones. El término aptitud está referido principalmente a rasgos estables a partir de los cuales pueden hacerse predicciones en desempeños futuros de los sujetos; asimismo, se entiende que las aptitudes no son entrenables en el corto plazo, si bien hay estudios sobre los rendimientos de los rezagados que arrojan conclusiones no siempre coincidentes (Rojas, 1985; Rojas et al. 1988; Donoso, 1988, 1989).

Por su parte, una prueba de conocimientos está referida a un muestreo de conductas que se ejercen sobre unidades de información en un dominio disciplinario determinado. A diferencia de las pruebas de aptitud, las pruebas de conocimientos no se caracterizan por la estabilidad de los resultados; su principal aporte proviene de la información que proporcionan acerca del grado de conocimiento acerca de una materia con que un postulante pretende ingresar a la universidad.

En el caso propio de las pruebas de conocimientos del sistema de selección, sus contenidos han sido muestreados a partir del currículo oficial de la enseñanza media, teniendo a la vista las necesidades del sistema de educación superior; es decir, aquellos contenidos que son más significativos para los programas de las carreras que la requieren.

5.2. Medición: características y propiedades

El concepto de la medición en psicometría, como es sabido, está definido teóricamente en términos de confiabilidad y validez. Operacionalmente, el acto de medir se concibe como el registro de respuestas (marcas hechas de manera estándar e invariable) cuyos cómputos proporcionarán los mejores estimadores posibles para determinar el grado de dominio o destreza de un sujeto sobre cada una de las dimensiones bajo examen (Nota 11).

La teoría psicométrica distingue tres tipos de confiabilidad: de formas paralelas, como estabilidad, y como consistencia interna. Las pruebas de aptitud académica han sido estudiadas acuciosamen- te en cuanto a su consistencia interna, a partir del número de ítemes, la varianza de cada uno de los mismos, y la varianza total de la distribución (Nota 12). Los resultados obtenidos indican que la confiabilidad en la parte verbal de la Prueba de Aptitud Académica es muy alta, manteniéndose alrededor de 0.94. En cuando a la parte matemática, el índice de consistencia interna es aún mayor, alcanzando a 0.97 (Díaz et al. 1990) (Nota 13).

Por su parte se distinguen cuatro tipos de validez: de contenido, concurrente, predictiva y de constructo. Estas dimensiones han sido examinadas en distintos trabajos, de forma que sólo se expondrán los casos de validez predictiva y de constructo.

En el caso de la validez predictiva, los análisis acerca del potencial predictor de las pruebas de aptitud arrojan resultados que se ubican dentro de los estándares internacionales, con especial relevancia del peso de las notas de enseñanza media y la parte matemática de la Prueba de Aptitud Académica. La capacidad predictiva se extiende de manera importante a los dos primeros semestres de los estudios universitarios.

La validez de constructo se refiere a la calidad o "bondad" conque la prueba mide el constructo hipotético. Un trabajo de Díaz et al. (1987) realiza esta evaluación, concluyendo que existe una apropiada y correcta articulación entre los conceptos o constructos y sus correspondientes referentes empíricos, a saber, las preguntas de las

pruebas. Concluyen diciendo que "se puede afirmar que las dimensiones evaluadas por la Prueba de Aptitud Académica representan de manera adecuada y técnicamente válida los principales procesos cognitivos necesarios para el logro de los objetivos propuestos en los planes y programas de las carreras impartidas por las Corporaciones de Educación Superior" (p. 34). Esta afirmación debe entenderse en la perspectiva de que se da por sentado un concepto de aptitud e inteligencia, cuestión discutida en secciones anteriores.

También es importante considerar este sesgo, por cuanto la docencia universitaria no escapa al modelo general transmisivo de la enseñanza.

La estandarización de una prueba se entiende en términos de los procedimientos que deben seguirse en la administración de la misma, de forma de asegurar el control de cualquier variante que pudiese contaminar los resultados. En este sentido, la aplicación de las pruebas de selección y, específicamente, las Pruebas de Aptitud Académica se caracterizan por ser rigurosas en este sentido.

Los eventuales errores que pudiese haber en los puntajes individuales entre el "puntaje real" del sujeto y su "puntaje de prueba", por lo tanto, encontrarán su origen en cuestiones completamente dependientes del sujeto mismo (como tensión de prueba o fatiga). Este es un argumento más a favor de la igualdad de oportunidades que estarían siendo garantizadas.

El término *dificultad* implica que las pruebas contienen determinados porcentajes de ítemes fáciles, medianos y difíciles, a fin de lograr una máxima discriminación dentro del grupo. Se combinan ítemes de diferente grado de dificultad procurando asignarles una dificultad media del 50%, que es el grado ideal para obtener esa buena discriminación (Avila, 1991, I, 19)

5.3. Lo que ocultan las pruebas

Por su propia estructura, las pruebas de aptitud revelan ciertas dimensiones de los sujetos en cuanto aspectos medidos, aunque no revelan otras. Bajo las series de puntajes, estadísticamente exentas hasta lo posible de error, están presentes dimensiones que, para el paradigma de inteligencia culturalmente dominante, son estructurales y no coyunturales, las que afectan a los estudiantes y sus rendimientos. Bajo esta concepción, las fuentes de error no serían atribuibles a las pruebas de aptitud en ninguno de sus aspectos (Nota 14) sino a los propios sujetos, conclusión que ciertamente puede discutirse.

En primer lugar, el que las pruebas de inteligencia y aptitud exijan a los examinandos que produzcan respuestas pero no que produzcan preguntas oculta una parte importante del intelecto de los sujetos. De esta manera, *"estas pruebas carecen de una mitad vital de la inteligencia, preguntar"* (Sternberg, 1987c: 11). Agrega el autor que es extraño que las pruebas de inteligencia sólo exijan responder preguntas, en vez de pedir hacerlas tanto como responderlas. Agrega que de esta manera se trata con sólo la mitad de lo que está implicado en la relación de la inteligencia con las preguntas, y esa mitad que es razonablemente la menos importante.

Una segunda dimensión oculta es que los resultados de la Prueba de Aptitud Académica reflejan crudamente un sistema educacional desigual, y también reflejan la baja calidad de la educación media nacional. En efecto, los puntajes estandarizados de las pruebas de aptitud como de selección ocultan los reales problemas de formación de los estudiantes de la enseñanza media. Por ejemplo, el hecho de que entre las partes matemática y verbal de la Prueba de Aptitud Académica, sea la matemática el mejor predictor tiene que ver seguramente con la baja capacidad lingüística de los estudiantes egresados de la educación media.

En la parte matemática, es notoria la falla en cuanto a procesos superiores de pensamiento formal; ello puede obedecer, como sugieren Díaz et al. (1990, 326), a la estructura jerárquica no sólo de la disciplina sino del currículo, tal que los procesos

superiores no pueden ser logrados sino sobre la base de los que les anteceden, lo que no siempre se logra cubrir por completo en el curso de la enseñanza media. Esto implicaría, desde un punto de vista de la epistemología genética, que los estudiantes que egresan de la enseñanza media en promedio no han alcanzado la fase de las operaciones formales.

Por otra parte, considerando los puntajes brutos, los bajos puntajes alcanzados en las pruebas específicas no indican otra cosa sino el fracaso de la educación media en el logro de los objetivos que le han sido fijados por el currículo oficial.

La serie coordinada por Avila (1991, Tomos I a VII) evidencia las diferencias estructurales que muestran los resultados de los estudiantes, dependiendo del tipo de colegio de proveniencia. Así, hallamos como norma resultados notablemente superiores en los colegios particulares pagados, seguidos por los particulares subvencionados y los liceos municipales en último lugar. También debe considerarse, para una apreciación más justa, que mientras los colegios pagados representan cerca del 7% de la matrícula de la enseñanza media, los municipalizados alcanzan a cerca del 80% (cfr. MINEDUC, 1990).

6. Las calificaciones de enseñanza media

Las calificaciones de enseñanza media tienen dos dimensiones en este análisis: por una parte, resumen en un único valor las apreciaciones evaluativas hechas al alumno durante su enseñanza media; por otra, son un componente importante dentro del sistema de selección y en su característica más notable que es la capacidad predictiva.

Es parte del saber común en el campo educativo que las calificaciones que ponen los profesores a los estudiantes son, técnicamente hablando, débiles, careciendo de calidad de muestreo significativo del dominio de contenidos, construidas sin respaldo técnico, con exigencias taxonómicas de bajo nivel (generalmente conocimiento simple y aplicación mecánica). Todo ello hace que una calificación no sea considerada confiable ni válida tanto del punto de vista psicométrico como del edumétrico. Sin embargo, al considerarse el total de calificaciones obtenidas por un sujeto durante el curso de cuatro años se obtiene un promedio final que se funda en varios cientos de registros. La combinación del número de registros a pesar del estrecho rango de puntuación que se les puede asignar, hacen que estas apreciaciones finalmente redunden en un promedio que es un buen estimador parametral.

Desde los inicios se ha notado una progresiva reducción del rango de las calificaciones por elevación de los valores inferiores, lo que ha impactado en la capacidad predictiva de las Notas de Enseñanza Media.

La capacidad predictiva de las calificaciones de Enseñanza Media ha sido apreciada desde los inicios del sistema de selección vigente. En 1985, Cristina Rodríguez expresaba que *"las calificaciones de Enseñanza Media son buenos predictores del rendimiento en la universidad, constituyendo en la mayoría de los casos el antecedente que más aporta a su explicación"* (Rodríguez, 1985: 47). Estudios posteriores, sin embargo, han hallado que junto con un decrecimiento de los índices de rendimiento en el logro de los objetivos de las diferentes asignaturas de la enseñanza media, se encuentra una tendencia cada vez mayor a hacer subir artificialmente las calificaciones de los estudiantes, sin que ello represente un mejoramiento sustancial de la calidad de los aprendizajes (Díaz, Himmel y Maltes, 1990). La reducción del rango o recorrido de los promedios implica una disminución de la varianza, con los consiguientes efectos sobre la capacidad predictiva que pueda tener la variable.

La reducción de la capacidad predictiva de las Notas de Enseñanza Media, sin embargo podría ser paliada por la consideración de una variable adicional que se refiere al lugar que ocupa el estudiante entre los alumnos de su colegio de origen. Himmel y Maltes (1985) informan que este elemento se usa con frecuencia en otros

países y "con resultados excelentes como elemento de selección."

Por otra parte existen opiniones relativas a que las calificaciones de la enseñanza media representan mucho más de lo que se supone y menos de lo que se espera. En lo primero, se puede pensar que las calificaciones representan, por ejemplo, el esfuerzo invertido por el estudiante en el logro del aprendizaje, o la inteligencia propia del mismo, los componentes socioeconómicos y culturales que lo ponen en ventaja o desventaja, la calidad de la educación que imparte el establecimiento, y otras.

En cuanto a lo segundo, ya la constatación de la reducción del rango de las calificaciones, en especial por el alza del límite inferior, indica que las mejores calificaciones no implican mejores aprendizajes. Esto es más grave cuando se compara con los rangos de variación de las calificaciones en el primer año de la educación superior, en que los límites inferiores que llegan al cero de la escala.

7. Discusión final

7.1. Respecto del sistema de selección

Las nuevas condiciones bajo las cuales se estructura la demanda por vacantes de pregrado en las universidades chilenas hace que este sea un tema abierto, posible de reordenarse bajo múltiples factores, siendo el financiero uno de los más importantes.

El tema del ingreso a la universidad en Chile ha perdido parte importante de su carga dramática que tuvo hasta fines de los años 90. En la actualidad existen muchas otras opciones universitarias que se van consolidando fuertemente y que pone efectivamente en tela de juicio el sistema de selección vía PAA como se le llama. Sin embargo mientras exista financiamiento ligado a puntajes en estas pruebas (o en otros instrumentos similares) van a ser factores a considerar dentro del proceso de selección.

Con todo, tendencial y persistentemente el tema selección en su conjunto ha perdido gravitación dentro del ámbito cada vez más presente de mejorar los procesos de producción del conocimiento o de calidad de la docencia (entendida ésta en su sentido más amplio). Es decir es muy posible que el acento futuro esté cada vez más presente sobre los procesos de producción, traspaso y recreación del conocimiento que realizan los estudiantes, más que en la garantía inicial de un determinado puntaje entregado por un conjunto de pruebas que miden parcialmente un conjunto reducido de habilidades y aptitudes.

De todas formas el fenómeno del sistema de selección vía PAA forma parte de la cultura universitaria y es empleado como clasificador de muchas otras dimensiones del hacer de la universidad en el plano docente, en el reconocimiento social como centro de calidad, así como también lo es para los establecimientos y los entes formadores, como finalmente lo es también para quienes obtienen buenos puntajes. Todo ello concurre hacia un marco que rectificar o cambiar radicalmente se torna complejo, por el conjunto de interrelaciones que presenta y la cantidad de empresas sociales asociadas al fenómeno analizado.

7.2 Respecto de las pruebas

Desde el punto de vista del modelo de inteligencia sobre el que se construye y bajo los principios de la teoría psicométrica, la Prueba de Aptitud Académica es consistente. En efecto, su rango de medición abarca conductas de los dominios que configuran operacionalmente su modelo de inteligencia, y cumple cabalmente los requisitos de confiabilidad y validez que señala la psicometría.

Si bien esta consistencia es muy alta en la Prueba de Aptitud Académica, es preciso entenderla bajo la óptica de que el modelo de inteligencia (verbal y matemática) que le subyace es el modelo culturalmente dominante, supuesto para el cual no parece haber evidencia definitiva.

Bajo estas condiciones, se entiende que el modelo se auto- pronostique: hipotéticamente no podría sino cumplirse porque se observan las variables del modelo y no otras, lo que lo ubica en una dimensión de profecía autocumplida, "la predicción que asegura su propio cumplimiento" (Gould, 1988: 151). Por ello es que no debería producir admiración que los puntajes de las pruebas predigan las calificaciones de los estudiante, al menos en el primer año universitario, ya que tanto los puntajes como las calificaciones se basan en un esquema que consiste en preguntas y respuestas bien estructuradas, definidas y preseleccionadas. Es por ello que *"tampoco hay que asombrarse que los puntajes de tests predigan mucho menos en relación a situaciones no académicas que a situaciones académicas"* (Sternberg, 1987c: 13).

La discusión se hace más compleja por cuanto la psicometría otorga un viso de objetividad y certeza a sus afirmaciones, haciéndolas extensibles al resto de la realidad social y cultural. La psicometría, sin embargo, no está libre de su condicionamiento histórico - cultural, aparte de los propiamente científicos. Sería interesante revisar los orígenes de la teoría y la medición de la inteligencia.

Si, además, la Prueba de Aptitud Académica mide inteligencia cristalizada, de acuerdo a Snow (1988)-, queda la pregunta por la medición de la inteligencia fluida, sus efectos sobre el mejoramiento de la predicción, la medición de potencial, y la rentabilidad social de la selección.

Lo anterior es especialmente importante si se considera que la inteligencia, culturalmente entendida, es inteligencia en un contexto o ecosistema dado, de los muchos en los que se mueven los sujetos (cfr. Brofenbrenner, 1990). Siendo la educación superior un ambiente altamente diferente al Liceo, si se miden apropiadamente los logros, esto es, la inteligencia cristalizada, podría suceder que esta inteligencia fuese parcialmente disfuncional a la educación superior, disfuncionalidad que podría ser compensada por la consideración de la inteligencia fluida.

Asumiendo el criterio de que una teoría se justifica si sirve como modelo explicativo del fenómeno objeto de la misma, podría pensarse en la necesidad o conveniencia de que se incorporaran tanto a la teoría como al diseño de la Prueba de Aptitud Académica o de la batería de selección, otras dimensiones de la inteligencia como ha sido sugerida por diferentes teóricos, como es la inteligencia fluida, la inteligencia como procesos, la dimensión ecológica del desempeño, etc.

Una dimensión sobre la cual aún no se conoce suficientemente es la del carácter ordenador de numerosas dimensiones de la vida que se ha dado a la Prueba de Aptitud Académica, tema sobre el que parece conveniente llevar a cabo mayor investigación

Notas

1. Con el aporte de Dirección de Investigación de la Universidad de Talca (Proyecto 463-10)
2. Por universidades antiguas se entiende a la Universidad de Chile, Pontificia Univer- sidad Católica de Chile, de Concepción, Técnica del Estado (que pasó a denominarse Universidad de Santiago de Chile), Católica de Valparaíso, Técnica Federico Santa María, del Norte (en la actualidad Universidad Católica del Norte) y Austral de Chile.
3. Las doce profesiones de carácter exclusivamente universitario son: Leyes, Psicología, Bioquímica, Ingeniero Agrónomo, Ingeniero Forestal, Ingeniero Civil (y sus menciones). Médico Cirujano, Cirujano Dentista, Economista, Médico Veterinario, Químico Farmacéutico, Arquitecto. Después se agregaron: Periodista y Profesor de Educación Básica, Media, y de Párvulos, educación especial y Diferencial, incorporadas como modificaciones a la Ley Orgánica Constitucional de Enseñanza.

4. Sólo la educación parvularia o preescolar está controlada en parte importante por instancias ligadas directa o indirectamente al Ministerio de Educación.
5. El Consejo de la Universidad de Chile, en 1966, acordó poner a disposición de las universidades del país la Prueba de Aptitud Académica. El día 11 de Enero de 1967 se aplicó por primera vez a nivel nacional
6. Se hace referencia al Bachillerato de la Universidad de Chile (fue el más importante). Existió también el de la ex Universidad Técnica del Estado, el de la Universidad Católica de Chile y de la Universidad Católica de Valparaíso.
7. Por "normalidad" se entiende el concepto estadístico en el sentido propio de la distribución probabilística de las medias muestrales, atendiendo además a que cada factor, aptitud o habilidad puede comprenderse como un dominio de conductas.
8. "Estabilidad" es la presencia invariante en el tiempo de un rasgo en un sujeto. Operacionalmente, significa que las diferencias de puntajes obtenidos en distintas y sucesivas aplicaciones de una prueba que mida ese rasgo serán mínimas. Este concepto se encuentra ligado a la medición. En los hechos, lo que se tienen son pruebas que miden establemente.
9. El SAT, Scholastic Aptitude Test, y el ACT, American College Testing Program, son pruebas utilizadas para selección de alumnos a los estudios superiores en los Estados Unidos. La Prueba de Aptitud Académica sigue el patrón del SAT.
10. Hasta la admisión 1997 las universidades de Chile y Católica de Chile, concentraban aproximadamente el 65% de los "mejores puntajes" y, por ello, los recursos provenientes del Aporte Fiscal Indirecto (DFL 4, 1981). Las veintitrés universidades restantes alcanzan, en conjunto, algo más del 30%.
11. La precisión del concepto, posterga otras dimensiones que no pueden (por ahora) ser recogidas y registradas de las maneras que fijan los instrumentos disponibles.
12. Un indicador frecuente para estimar la consistencia interna es la fórmula desarrollada por Kuder y Richardson, conocida como KR-20 y el coeficiente alfa de Crombäch
13. Los datos corresponden a 1989, estudios parciales realizados años siguientes por las distintas universidades confirman estos datos.
14. Podría sin embargo, aludirse al tema del sesgo cultural que implica tanto la forma de los reactivos (ítemes de selección múltiple) como los términos usados y sus cargas semánticas.

Referencias

Aranda, Fernando (1985). "Ingreso a la Educación Superior. El Sistema Nacional", Cuadernos del Consejo de Rectores de las Universidades Chilenas. No. 24, enero-junio de 1985, pp. 1-23

Avila, Emilio (Coord.) (1991). Informe de los Resultados de las Pruebas de Admisión a la Educación Superior. Santiago: Universidad de Chile, Depto. de Administración de la Prueba de Aptitud Académica, Unidad de Estudios, 1991.

Ashwort, A. Harvey, R. (1994) *Assessing Quality in Further and Higher Education.* Jessica Kingsley Publishers. London.

Brunner, José Joaquín (1993). "La Reforma de 1980 diez años después", En: F.E.S. (1993), op. cit, pp. 229 -238

CONSEJO DE RECTORES (1985). "Editorial", Cuadernos del Consejo de Rectores de las Universidades Chilenas. N? 24, Enero-Junio de 1985.

CONSEJO DE RECTORES (1992). Informe de la Comisión de Estudio sobre la Prueba de Aptitud Académica. Santiago, Octubre de 1992.

Diaz, Eliana; Himmel, Erika. Maltes, Sergio (1990). "Evolución Histórica del Sistema de Selección a las Universidades Chilenas, 1967-1989." En Lemaitre, María

José (ed), 1990.

Diaz, Eliana; Himmel, Erika; Donoso, Graciela y Maltes, Sergio (1987), Validez de Constructo de la Prueba de Aptitud Académica. Santiago: Universidad de Chile, Dirección General Académica y Estudiantil; Pontifica Universidad Católica de Chile, Vicerrectoría Académica.

Donoso, Sebastián (1988). "La relación entre los rezagados y los alumnos de la promoción en el proceso de admisión a la educación superior: lecciones para el sistema educacional", Estudios Pedagógicos, No. 14, pp. 41-57. Valdivia, Universidad Austral de Chile.

Donoso, Sebastián (1989). "Admisión a la Educación Superior: magnitud de diferencias entre los postulantes de la promoción y rezagados", *Estudios Pedagógicos,* No. 15, pp. 15-32. Valdivia, Universidad Austral de Chile.

Gardner, Howard (1995) *Inteligencias múltiples. La teoría en la práctica.* Editorial Paidós Ibérica, Barcelona.

Grassau, Erika (1956). Análisis estadístico de las pruebas del Bachillerato de la Universidad de Chile. Santiago: Universidad de Chile.

Grassau, Erika (1966). Los exámenes de Admisión a la Universidad. Universidad de Chile: Instituto de Investigaciones Estadísticas.

Himmel, Erika y Maltes Sergio (1978), "Análisis de la selección y de la predicción del rendimiento académico 1977 en la Pontificia Universidad Católica de Chile", Vicerrectoría Académica, Informe No. 5.

Himmel, Erika y Maltes Sergio (1978a) "Modelos de análisis del rendimiento académico universitario. Pontificia Universidad Católica de Chile", Vicerrectoría Académica, Informe No. 6.

Himmel, Erika y Maltes Sergio (1979) Estabilidad de la capacidad predictiva de los factores de selección en la Pontificia Universidad Católica de Chile. Vicerrectoría Académica, Informe No. 7.

Himmel, Erika y Maltes Sergio (1985) "Sistemas Alternativos de selección universitaria", *Revista del Consejo de Rectores.* Santiago. Vol. 25.

Himmel, Erika y Maltes Sergio (1986) "El Sistema de Selección a las Universidades Chilenas", En Lavados, H; E. Hill; V. Apablaza (Eds) El Sistema Educacional Chileno. Santiago: Corporación de Promoción Universitaria, pp. 317-360

Lehman, Carla. Editora (1990), Financiamiento de la Educación Superior: Antecedentes y Desafíos. Santiago, Centro de Estudios Públicos, Colección Foro de la Educación Superior.

Lemaitre, Ma. José. Ed. (1990). La Educación Superior en Chile: un Sistema en Transición. Santiago: Corporación de Promoción Universitaria -CPU, Colección Foro de la Educación Superior.

Rodriguez, Cristina (1985), "Resultados de la Aplicación del Sistema en Chile", Cuadernos del Consejo de Rectores de las Universidades Chilenas. No. 24, Enero-Junio de 1985, pp. 24-58

Rojas, R. Bocchieri, M.A. Schublin, A. Cisternas, L.1988). Estabilidad de los Resultados de la Prueba de Aptitud Académica y de las Pruebas de Conocimientos Específicos. Santiago, Universidad de Chile, Dirección General Académica, 1988.

Snow, Richard (1987), "Educación e Inteligencia", En: Sternberg, R. op cit. 1988,

791-920

Sternberg, Robert J. (1987a). *Inteligencia Humana, I. La naturaleza de la inteligencia y su medición.* Buenos Aires: Paidós,

Acerca de los Autores

Sebastián Donoso

P.O.Box 747 Talca. Chile
Phone: 5671 -200251
E-Mail: sdonoso@pehuenche.utalca.cl

Profesor Asociado del Instituto de Investigación y Desarrollo Educacional
Universidad de Talca, Chile.

Profesor de Diseño de Políticas Educacionales, autor de artículos y libros acerca de las características y resultados del Sistema de Selección de alumnos a las Universidades.

Gustavo Hawes

Profesor Asociado y Director del Instituto de Investigación y Desarrollo Educacional

Universidad de Talca, Chile.

P.O.Box 747 Talca
Chile

Phone: 5671 -200255
E-Mail: ghawes@pehuenche.utalca.cl

Profesor de Evaluación de Políticas Educacionales, autor de artículos y libros acerca de las características y resultados del Sistema de Selección de alumnos a las Universidades.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
aacosta@cucea.udg.mx

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

This article has been retrieved 847 times since May 10, 2000

Education Policy Analysis Archives

Volume 8 Number 22 May 10, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

The Influence of Scale on School Performance: A Multi-Level Extension of the Matthew Principle

Robert Bickel
Marshall University

Craig Howley
Ohio University and AEL, Inc.

Abstract

In this study, we investigate the joint influence of school and district size on school performance among schools with eighth grades (n=367) and schools with eleventh grades in Georgia (n=298). Schools are the unit of analysis in this study because schools are increasingly the unit on which states fix the responsibility to be accountable. The methodology further develops investigations along the line of evidence suggesting that the influence of size is contingent on socioeconomic status (SES). All previous studies have used a single-level regression model (i.e., schools or districts). This study confronts the issue of cross-level interaction of SES and size (i.e., schools and districts) with a single-equation-relative-effects model to interpret the joint influence of school and district size on school performance (i.e., the dependent variable is a school-level variable). It also tests the equity of school-level outcomes jointly by school and district size. Georgia was chosen for study because previous single-level analysis there had revealed no influence of district size on performance (measured at the district level). Findings from this study show substantial cross-level influences of school and district size at the 8th grade, and weaker influences at the 11th grade. The equity effects, however, are strong at both grade levels and show a distinctive pattern of size interactions. Results are interpreted to draw implications for a "structuralist" view of school and district restructuring, with particular concern for schooling to serve impoverished communities. The authors argue the importance of a

 BEST COPY AVAILABLE

notion of "scaling" in the system of schooling, advocating the particular need to create smaller districts as well as smaller schools as a route to both school excellence and equity of school outcomes.

Research on the role of school and district size as an influence on school performance has a long history and a large literature (see, for example, Barker & Gump, 1964; Guthrie, 1979; McDill, Natriello, & Pallas, 1986; Smith & DeYoung, 1988; Fowler, 1991; Walberg & Walberg, 1994; Khattari, Riley & Kane, 1997; Stiefel, Berne, Iatarola, & Fruchter, 2000). The varying methods used to study the issue have, of course, generated conflicting results (Rossmiller, 1987; Caldas, 1993; Lamdin, 1995; Rivkin, Hanushek & Kain, 1998). In consequence, size has often been relegated to the status of an obligatory but uninteresting control variable. Not infrequently, it has simply been ignored altogether (Barr & Dreeben, 1983; Burtless, 1996; Gamoran & Dreeben, 1986; Farkas, 1996; Wyatt, 1996; Hanushek, 1997, 1998). A recent school effectiveness review by eleven production-function virtuosos, for example, devoted just three of its 396 pages to school size (Betts, 1996, pp. 166-168). Consequences of variability in school size, moreover, were, in passing, judged to be uncertain. District size is considered even less interesting than school size by most researchers interested in school performance.

The study reported here, by contrast, builds on a line of evidence that has related the size of both districts and schools to aggregate student achievement. Previous research developing this line of evidence, however, has constructed only single-level analyses (schools or districts). The present study deploys a multi-level method (Boyd & Iversen, 1979; Iversen, 1991) to link effects at the two levels. In other words, this new work constitutes a first step from an empirical consideration of "size effects" toward an empirical consideration of "scale effects" (cf. Guthrie, 1979).

School System Scale: A Timely Issue

A great deal of skepticism exists about the role of size as a structural condition of US schooling. Educators have generally disparaged the role of structure and focused attention on the role of process. This focus of interest is easy to fathom. Both school teachers and administrators devote themselves to the processes of teaching and administration; the structural features of their practices are, for the most part, tacit. Teachers and principals encounter schools and districts as the particular stages on which they personally enact their work and deploy professional processes. Whatever structural variety might distinguish one such "stage" from the next, teachers and principals do not often personally experience it. Superintendents, by contrast, are better positioned to develop a sense of structural differences among schools and districts, but such an appreciation might be almost as exceptional among superintendents as it is among other educators, since process also consumes most of a superintendent's time.

This propensity to focus on process has a philosophical dimension, as well. A structuralist view confines free will to an apparently smaller range of influence as compared to a view that privileges process. Education, and the culture of education, pays considerable homage to free will (cf. Bruner, 1996). In the grandest tradition, education is seen as the route to a "larger life" open to everyone equally (e.g., Prichard Committee, 1990). James Coleman was among the first to point out that equal educational opportunity was more problematic than previously imagined, of course, and due to structural reasons. The school effectiveness literature ensued and dramatically valorized process as the profession's response to a sociological perspective on structure; school reform has had a procedural focus ever since (cf. Dorn, 1998).

Recent research and current events, however, have combined to challenge the conventional disposition to privilege process over structure. First, nearly a decade of research on school size (in particular) has developed a preponderance of evidence to suggest that smaller school size would improve schooling in impoverished communities (Howley, 1989; Irmsher, 1997; Raywid, 1999). Second, school-shooting tragedies have curiously and sadly brought the issue of school size

to popular attention. Possibly as a result of these awful events, the US Secretary of Education and the Governors of Georgia and North Carolina have recently spoken in favor of small schools. Surprisingly, the Secretary praised the resistance of rural communities that have fought fiercely for decades to preserve their small schools in the face of consolidation (Riley, 1999). It has, of course, been a losing battle, with some fortunate exceptions.

The recent attention has not even begun to challenge the privileged position that process enjoys, of course, and many observers continue to believe that administrative arrangements like "schools-within-schools" and "houses" can replicate the processes presumed to characterize small scale. Both Mary Anne Raywid (1996) and Deborah Meier (1995) argue persuasively that the conditions of smallness entail characteristics tantamount to structural difference: separate administration, separate budgets, distinctive authority, unique cultures, and so forth. Simulations, it turns out, have difficulty reproducing these structural features of small scale.

Nonetheless the rhetorical change is itself dramatic. No longer does size appear merely as a footnote to effectiveness studies or as a container of essentially interesting processes, but as a distinct phenomenon. School size now matters in discourse, anyhow.

School district size, however, continues to be regarded as a much less interesting issue than school size. The size of a district would seem to have no direct and little if any net influence on student achievement. As a variable, district size seems quite remote from student learning. Thus, most studies have considered district size almost purely as an administrative issue bearing on resource allocation (e.g., Bidwell & Kasarda, 1975; Meyer, Scott, & Strang, 1987). There have been a few exceptions within these studies, of course. Bidwell and Kasarda (1975) studied district size and concluded its influence on school performance was complex and contradictory:

> The total effects of [district] size were slight because its consequences for output, transmitted mainly by the structural and staff qualifications variables, were of roughly equal strength in a positive and in a negative direction.... It was associated with well-qualified staff and low administrative intensity (and, therefore, we have argued, with minimal diversion of human resources away from front-line tasks). But large size also meant more students to teach and thus higher ratios of students to teachers. (p. 69)

However, beginning with a 1988 study (Friedkin & Necochea, 1988), a new line of evidence has developed the hypothesis that the influence of both school and district size on aggregate performance is contingent on socioeconomic status. The direction of the effect has implicated small size (of schools and districts separately analyzed) as productive for the performance of schools or districts serving more impoverished communities, but larger size as productive for more affluent communities. Howley (1996) replicated the California study in West Virginia and reported similar results. Recent work (to be considered shortly) has extended the single-level findings to Georgia, Montana, Ohio, and Texas—with nearly identical results.

Relevant Literature

Researchers' tendency to overlook the interaction of school and district size with other variables (such as poverty) may be a disabling limitation of most studies that investigate the influence of school and district size on achievement, including quite recent efforts (e.g., Stiefel et al., 2000; Mik & Flynn, 1996; Riordan, 1997). This oversight tends to perpetuate the view that one size must fit all circumstances, or that some universally "best size" must exist (e.g., Lee & Smith, 1997; Stevenson, 1996). On this dubious view, size-related benefits and size-related costs are inadvertently construed as being enjoyed equally by all students (Conant, 1959; Haller, 1992; Haller, Monk, & Tien, 1993; Hemmings, 1996). Stiefel and colleagues (2000), using a somewhat more refreshing approach, recently found that small

regular 9-12 high schools have a budget- per-graduate that is no greater than the budget-per- graduate of other 9-12 high schools, and, in some cases a much cheaper budget-per-graduate. (The Berne study, however, uses a small sample of schools from a single large city (n=121) and leaves aside the question of the difference between budgeted and actual costs. The conclusions about small school size, unfortunately, rest on data from just 19 small high schools, of which only 8 are "regular" schools!)

Within the past decade, however, a growing body of empirical research has held that size is negatively associated with most measures of educational productivity. These conclusions encompass measured achievement levels, dropout rates, grade retention rates, and college enrollment rates (e.g., Walberg & Walberg, 1994; Stevens & Peltier, 1995; Fowler, 1995; Mik & Flynn, 1996). The drift of the past decade of this research, then, is to portray the optimal or best size as somewhat smaller than it was after James Conant proposed 400 students as the absolute minimum size for a suitably "comprehensive" high school (Conant, 1959; Lee & Smith, 1997).

Seldom have policy makers or researchers asked "Better for whom?" or "Better for what?" or "Better under what conditions?" Asking such questions, of course, may be seen as leading to unbearable complications. Again, in this welter of interest, indifference, and outright evasion, the role of district size is seldom considered, though both Herbert Walberg's (urban) and John Alspaugh's work (rural) remain notable exceptions (e.g., Alspaugh, 1995; Walberg & Walberg, 1994).

Size-by-Socioeconomic Status Interaction Effects

The joint or interactive, rather than independent, effects of size and socioeconomic status (SES), may also have contributed to renewed interest in smaller schools and districts. If smaller schools and districts are shown to benefit some settings, the new conventional wisdom (i.e., "smaller is better") gains support.

Specifically, interaction effects reported in some studies suggest that the well-known adverse consequences of poverty are tied to school size and, to some extent to district size, in substantively important ways. In brief, as size increases, the mean achievement of a school or district with less-advantaged students declines. The greater the concentration of less-advantaged students attending a school, the steeper the decline.

Investigations of the interaction hypothesis are relatively new, and multiple replications have only recently been undertaken and completed (see Howley & Bickel, 1999, for a recent synthesis of results in four states). Replications are important because without them, confidence in findings would be comparatively weak; research done in other locations could well yield different, and perhaps sharply conflicting, results.

The additional replications, however, now extend the scope of findings to Georgia (Bickel, 1999a), Montana (Howley, 1999a), Ohio (Howley, 1999b), and Texas (Bickel, 1999b). Previous work concerned California (Friedkin & Necochea, 1988); Alaska (Huang & Howley, 1993, in a study in which students were the unit of analysis), and West Virginia (Howley, 1996). These states represent considerable variety salient to the structure and operation of schooling in the United States—rural and urban mix, ethnic mix, magnitude of influence of State Education Agency, district organization types, school and district size, and funding inequity (Howley & Bickel, 1999).

The school-level findings in these single-level analyses are robust. In every study, an interaction effect has been confirmed. The effect varies from very strong (California, Georgia, Ohio, Texas, and West Virginia) to weak, (Montana) (Note 1). The overall conclusion is that smaller schools help maximize achievement for schools serving impoverished communities, but that larger schools serve the same function for more affluent communities.

Robust district-level interaction effects, however, were discovered in the four recent studies only in Ohio. Somewhat weaker direct negative effects of district size were reported for Texas; still weaker direct and interactive effects were evident in Montana. No district- level interactions were found in the Georgia study (Bickel, 1999a). The recent findings about district-level effects differed from the earlier

findings for California and West Virginia, where substantial district-level interactions were evident (Friedkin & Necochea, 1988; Howley, 1996).

Equity Effects

In addition to reviving interest in school size as a variable of importance in educational research, this work has begun to sensitize researchers, policymakers, journalists, and (perhaps most notably) citizens to equity concerns associated with school size. One-size-fits-all is no longer a unanimous judgment. Some researchers and policymakers have indeed begun to ask, "Best-size-for-whom?" (Henderson & Raywid, 1994; Devine, 1996).

In the five replications of the Friedkin and Necochea work (i.e., West Virginia, Georgia, Montana, Ohio, and Texas) Howley and Bickel also hypothesized equity effects of size. This hypothesis proceeds logically from confirmation of the interaction hypothesis. Namely, if small size improves the odds of academic success in small schools and districts (a sort of "excellence effect" of size), then the usual relationship between SES and performance must be to some extent disrupted in them as compared to larger schools and districts. Simple zero- order correlational analysis was used to measure the magnitude of relationship between SES and achievement in smaller versus larger units (schools or districts divided at the median in these separate data sets).

The equity effects of size are more consistent and more impressive, in fact, than the excellence effects. At all grade levels, in all five states, for both schools and districts, for a variety of alternative measures of SES, and for quite different sorts of achievement tests (i.e., both criterion-referenced and norm-referenced), the amount of variance in achievement associated with SES is substantially reduced in smaller units. In most cases, the magnitude of the relationship (Note 2) among the smaller units is about half what it is among the larger units (Howley, 1996; Howley & Bickel, 1999).

The Challenge of Cross-Level Interactions

Although the "excellence effects" of school size and the "equity effects" of both school and district size seem clear from the analyses reported by Howley and Bickel (1999), failure to confirm interaction "excellence effects" for districts in some states is intriguing. The line of evidence about school and district size has not, however, thus far included examinations of possible links between school size and district size. As a result, if unacknowledged multi-level contextual effects were present, previous studies would have ignored some portion of the structural influence of size on achievement. If the cultivation of high levels of achievement is a complex matter dependent on multiple influences, then we ought to suspect the existence of cross-level influences.

Further, discovery of such cross-level influences could be considered evidence that a structural notion of organizational scale was relevant to the enterprise of schooling—most particularly to the cultivation of academic achievement. If such cross-level relationships existed, administrators and policy makers would be well advised to coordinate their view of school size with a view of district size—and eventually with classroom size, and individual student performance, at one end of the spectrum, and size of the state and even national systems at the other end. The phenomenon of scaling could be seen as a structural characteristic of state school systems (see Thiétart & Forgues, 1995, for an interesting discussion of scaling as a feature of nonlinear dynamic systems in a chaotic state).

Methods

The present study addresses these issues by extending the consideration of "excellence effects" and "equity effects" of school and district size to a multi-level analysis with cross-level interaction terms. We chose to examine these relationships with the data for Georgia precisely because no effects of district size—either direct or interactive—had been discovered in the single-level analyses conducted by Bickel (1999a). On the basis of district-level effects that are inconsistently evident across states, we hypothesize the presence of cross-level interactions that could not be

detected in the previous single-level analysis.

> The Georgia dataset on which all analyses in this report are based is available for download here in any one of three formats:
>
> - SPSS (409K filesize),
> - Excell (1.65M), or
> - ASCII text (460K).

We might as easily have chosen any of the other states, but the use of individual states is advisable for two reasons, the first theoretical and the second practical. First, from the perspective of scale, each state constitutes a uniquely structured system. In this sense, combining dissimilar states is more likely to misrepresent reality than to provide a fuller picture of it. Second, since comparable achievement measures are not available for schools and districts across the four states for which we have assembled recent data, the merging of data sets would necessarily inflate measurement error.

A Single-Equation Relative-Effects Model

To study further previously identified equity effects, we specifically ask, in this two-level analysis, if there are cross-level interaction effects that remain significant in regression equations constructed to include school and district size, as well as school and district SES, and which also control for the proportion of students who are African American, the proportion of students from ethnic minorities, and pupil-teacher ratio (a proxy for class size). Our focal interaction terms are the products of (1) district size and school SES and (2) school size and district SES. Our model also includes the two original interaction terms: (1) the product of district size and district SES and (2) the product of school size and school SES.

We use a procedure developed by Boyd and Iversen (1979) and Iversen (1991). It employs ordinary least squares estimates (Note 3) of partial regression coefficients for school-level variables, district-level variables, and school-by-district interactions in the same equation. In effect, we are combining school-level and district-level regression models, and including school-by- district interactions, which reflect variability in district-level effects from school to school (Bryk & Raudenbush, 1992, pp. 70-74). The dependent variables in these equations are always school-level performance measures.

We adopt the single-equation relative-effects version of the model, since school-level and district-level variables are likely to be closely correlated. In this model, school-level variables are centered with respect to their group means (i.e., district means) and district-level variables are centered with respect to the grand mean. Centering all independent variables in this way helps to avoid inflated estimates of standard errors due to multicollinearity (Cronbach, 1987). Centering also enables us to unambiguously partition the percentage of variance in a dependent variable accounted for by each set of independent variables in our multilevel models (Iversen, 1991). Four such distinct sets of independent variables exist in our model: (1) the set of individual-level (school) variables, (2) the set of group-level (district) variables, (3) the set of single-variable interactions by level (e.g., the product of school size and district size), and (4) a set of within and cross-level interactions of different variables. Within the fourth set of variables are found the focal interactions of this study—the two cross-level interactions of SES and size: (1) the product of district size and school SES and (2) the product school size and district SES.

Examination of residuals plotted against the independent variables shows that the residuals are not uniformly distributed with respect to SPANSIZE for the 8th grade outcome measures. The same is true for FREEPCT when using the eleventh grade outcome measures. As a result, we used weighted least squares to remedy these departures from homoscedasticity, thereby restoring the efficiency of the estimators (Gujurati, 1995, pp. 381-390).

Data Sources and Variables

Official representations describe Georgia as a state with an educational system encompassing approximately 1800 public schools (e.g., Georgia Department of Education, 1999). The data set we are using, for school year 1996-97, contains complete information on 1626 regular public schools. For this study we selected for analysis data about the universe of schools with grade 8 or grade 11 test scores. Grade 8 is the grade level in Georgia with scores prior to the wave of early-school leaving that transpires at the high school level (generally grade 10), whereas grade 11 data portray the relationships that prevail subsequent to this too-familiar exodus.

The choice of these grade levels for analysis is therefore strategic. First, students from impoverished backgrounds become dropouts more frequently than students from more affluent backgrounds. Second, this being the case, the demography of schooling at grade 11 will differ somewhat from the demography at grade 8, namely in the fact that the proportion of impoverished students will have declined. Third, the probable effect of these changed conditions, we hypothesize, will be to weaken grade 11 results. The reason for this inference is that if smaller sizes positively influence achievement in impoverished schools, demographic changes in larger schools serving impoverished students will, in effect, cast off the cause of their negative influence—by removing disproportionate numbers of impoverished students. (Note 4)

Dependent variables. Dependent variables are school-level percentile rank scores for eight subtests of the widely used Iowa Test of Basic Skills (grade 8) and school-level percentage of students passing the first administration of the Georgia High School Graduation Test (grade 11). School-level means vary dramatically with both tests, from as low as the first percentile to as high as 93rd for the ITBS and from 11 to 100 percent passing (on the grade 11 Graduation Test).

Seven of the ITBS subtests are designed to measure achievement in reading comprehension, mathematics, reading vocabulary, social studies, language arts, science, and research skills. The eighth subtest is a composite measure, intended to provide a global gauge of achievement.

The High School Graduation Test is used in this study because the ITBS is not administered above grade 8 in Georgia. The Graduation test gauges achievement in English, mathematics, social studies, and science. In addition, students receive a composite score. First administration passing percentages for the five scores are used as our outcome measures for the eleventh grade.

Independent variables. Our main predictor variables, (each measured at the school level, at the district level, and as the interaction between the school and district level) include the following: (1) number of students per grade level in thousand-student units as our measure of size (SPANSIZE); (2) proportion of all students eligible for free or reduced-price meals (FREEPCT); (3) proportion of African-American students (BLACKPCT); (4) proportion minority (i.e., nonwhite) students (MINORPCT); and (5) student-teacher ratio (S/SRATIO), a proxy for class size. We include student-ratio, in particular, to address the possibility that any findings might principally be the result of differences in class size, rather than differences in school or district size.

In order to test for the existence of cross-level interactions between size and SES, we include four interaction terms: (1) school SPANSIZE by school FREEPCT, which is the same as the school-level interaction term that had proven significant in previous single-level analyses; (2) district SPANSIZE by district FREEPCT, which is the same as the district-level interaction term that had proven non-significant in previous single-level analyses of Georgia data; (3) district SPANSIZE by school FREEPCT, which is one cross-level interaction term of interest in this multi-level analysis; and (4) school SPANSIZE by district FREEPCT, the other cross-level interaction term of interest in the present study.

Results

Tables 1 and 2 provide descriptive statistics (means and standard deviations) for our dependent and independent variables for grade 8 and 11, respectively. SPANSIZE, at both the school and district level is measured in units of 1,000 students. A standard deviation of ".NNN," in the case of district size, for instance, is

therefore equivalent to the product of ".NNN" and 1,000. Tables 3 through 10 report regression results (Note 5) for the eight achievement measures that predict school performance at the 8th grade level. The first panel in each table apportions explained variance in three columns to (1) individual-level (school-level), (2) group-level (district-level), and (3) individual-by-group (school by district) interactions. The second panel reports, in a single column, the variance attributable to interactions among SES and size variables, at both levels (i.e., individual and group), yielding the four interaction terms specified in the concluding paragraph of the methods section.

In the reporting of results below, only selected tables are presented, which nonetheless convey the findings from the complete set of analyses. The complete set of tables in Rich Text Format can be downloaded from this point.

Table 1

Descriptive Statistics: Grade 8

Dependent Variables
Schools

	Mean	St. Dev.
READING COMPREHENSION	47.02	12.88
MATHEMATICS	52.26	12.42
READING VOCABULARY	43.82	15.05
LANGUAGE ARTS	54.20	12.72
SOCIAL STUDIES	51.31	12.04
SCIENCE	51.07	13.88
RESEARCH SKILLS	53.01	12.60
COMPOSITE	51.25	13.71

Independent Variables

	Mean/(St. Dev.)	
	Districts	**Schools**
SPANSIZE	0.219	0.259
	(0.101)	(0.124)
FREEPCT	48.18	45.28
	(17.48)	(22.93)
BLACKPCT	34.47	37.29
	(25.25)	(29.66)
MINORPCT	2.91	4.14
	(4.22)	(5.41)
S/RRATIO	16.13	16.25
	(1.51)	(1.86)
	N=158	**N=367**

Table 2
Descriptive Statistics: Grade 11

Dependent Variables
Schools

	Mean	St. Dev.
ENGLISH	92.87	5.18
MATHEMATICS	85.33	9.77
SCIENCE	70.66	15.22
SOCIAL STUDIES	75.14	12.97
COMPOSITE	63.89	16.41
	Independent Variables Mean/(St. Dev.)	
	Districts	**Schools**
SPANSIZE	0.233	0.280
	(0.139)	(0.114)
FREEPCT	48.18	33.49
	(19.76)	(21.26)
BLACKPCT	35.42	38.03
	(25.30)	(29.80)
MINORPCT	2.53	3.84
	(3.42)	(5.03)
S/RRATIO	17.03	17.74
	(2.34)	(3.15)
	N=155	**N=298**

Recall that previous single-level analyses reported statistically significant and negative SPANSIZE by FREEPCT interaction effects. These conspicuous effects meant that as school (and in some states, district) size increased, the mean achievement costs associated with less-advantaged students increased. Tables 1 through 8 again confirm interaction effects, but the interactions portrayed there are quite clearly shown to represent a complex phenomenon that escaped notice in single-level analyses. These more complex effects were predictably masked in the earlier single-level analyses, since those analyses examined schools and districts separately. The following written report of the findings may be difficult to follow, but the Tables themselves actually picture a consistently complex set of relationships prevailing between schools and districts as those complex relationships influence school-level performance. We encourage readers to refer to the Tables as they read the following discussion.

Eighth Grade "Excellence Effects"

Combining schools and districts in a multilevel analysis, the single-level SPANSIZE by FREEPCT interaction effects that were so conspicuous in the previous single- level research are not evident at all at the 8th grade. However, several interesting (and uniquely specified) single-level and cross-level interactions are present in the equations. Overall this means that the effects of size on achievement depend on multiple influences, and not merely school- or district-level SES. One size is shown more clearly than ever before not to fit all cases, and, at the same time, these results suggest that the influential features of circumstance vary to such an extent that each setting can be understood as unique. We present this conclusion prematurely in order to help readers take a wider perspective on the presentation of detailed findings that follows.

Single Variables Within and Across Levels. First let us consider the results given in panel 1 of Tables 3 through 10 (the unique influence of single variables at each of two levels separately and then jointly across levels). We will interpret the results of Table 10 (composite achievement) only, as the results given there can be viewed as not only encompassing the generality of the findings reported in Tables 3 through 9, but as representing a summative indicator of school performance. Readers are, however, directed to those other Tables to observe the somewhat variant results among the various ITBS subtests. We will first consider the single variables as

unique school-level and district-level influences (Note 6):

(1) Both FREEPCT (-) and BLACKPCT (-) exhibit uniquely significant ($p < .001$ and $< .01$, respectively) school-level influences in the equation, accounting for 26.4% of the variance in school-level performance. Neither SPANSIZE nor S/RRATIO (our proxy for class size) show any net direct influence at the school level.

(2) FREEPCT (-) and MINORPCT (+) exhibit uniquely significant ($p<.001$ and $p<.01$, respectively) district-level influences in the equation, accounting for 31.3% of the variance in school performance.

Table 10
Weighted Regression Results with Corrected Standard Error
Grade 8: Composite Score

Unstandardized and (Standardized) Regression Coefficients

	Individual-Level	Group-Level	Individual by Group Interactions
SPANSIZE	-6.401	27.174	-308.619**
	(-.050)	(.094)	(-.167)
FREEPCT	-0.401***	-0.340***	-0.005*
	(-.418)	(-.460)	(-.098)
BLACKPCT	-0.119**	-0.001	-0.003**
	(-.207)	(-.002)	(-.212)
MINORPCT	0.112	0.347**	0.014
	(.040)	(.121)	(.033)
S RRATIO	0.457	-0.660	-0.460
	(.043)	(-.060)	(-.068)
Variance Explained	26.4%	31.3%	10.8%

Within-Level and Cross-Level Interactions

SCHOOL SPANSIZE by SCHOOL FREEPCT	0.141 (.024)
DISTRICT SPANSIZE by DISTRICT FREEPCT	-0.332 (-.023)
DISTRICT SPANSIZE by SCHOOL FREEPCT	-4.304*** (-.211)
SCHOOL SPANSIZE by DISTRICT FREEPCT	-1.046*** (-.237)
Variance Explained	10.7%

Residual Intraclass Correlation .056
School/District Ratio 2.32
Standard Error Inflation 6.88% (Corrected)

Partial Derivatives for Y with Respect to (1) SCHOOL SPANSIZE and (2) DISTRICT SPANSIZE

Y wrt 1 = - 308.619 x (DISTRICT SPANSIZE) - 1 046 x (DISTRICT FREEPCT)

Y wrt 2 = - 308.619 x (SCHOOL SPANSIZE) - 4 304 x (SCHOOL FREEPCT)

*p <.05
** p <.01
***p <.001

These two single-level results show that a substantial portion of the variance in school performance (i.e., mean ITBS percentile rank in a school) actually is

accounted for by district-level influences. Poverty contributes a negative influence that is about 4 times the magnitude of the positive influence of MINORPCT. The direct influence of district size and district student teacher ratio, we note, are once again nonsignificant.

We next consider the individual by group interactions reported in column 3 of panel 1 (Table 10). This column reports cross-level interactions for each of the major variables separately. That is, these reported interactions compute the interactive (joint) influence of SPANSIZE, FREEPCT, BLACKPCT, MINORPCT, and S/SRATIO at the two levels. Results, which account for a unique 10.8% of the variance in school-level performance, are summarized as follows:

(1) The unique interactive influence, across levels, of SPANSIZE (-) is highly significant ($p<.001$).

(2) The unique interactive influence, across levels, of FREEPCT (-) is somewhat significant ($p<.05$).

(3) The unique interactive influence, across levels, of BLACKPCT (-) is also significant ($p<.01$).

(4) There is no unique interactive influence, across levels, of MINORPCT or S/RRATIO.

To interpret these interactive results, recall that all independent variables are centered for the regression analyses. Values of the variables that fall below the mean are negative and values that fall above the mean are positive. The product of two negative values at the district level (e.g., low district poverty) and school level (small school size) will yield positive values of the interactive variable, just as the product of positive values at both levels will yield positive results. In this Georgia data set, the existence of small schools in small districts, and the existence of large schools in large districts are conditions uniquely associated with lower school performance. (Note 7) Similar inferences can be drawn in the case of FREEPCT (though the influence here accounts uniquely for less than 1% of school performance) and BLACKPCT. It is crucial for readers to keep in mind that the influences on school performance discussed thus far are not interpretable in isolation from the totality of size influences. This research is developing a model of cross-level influence of size on school performance. In this model, however, we can see that single-variable influences within and across levels account for almost 70% of the variance in school performance.

Variables Interacting Within and Across Levels The single variables—whether uniquely at different levels, or jointly across levels—present a substantial but still incomplete view of influences on school performance. These influences, in this analysis, are completed by an analysis of interactions between variables, both within and across levels. We turn next, therefore, to a consideration of these influences, given in the second panel of Tables 3 through 10. Again, discussion centers on Table 10 (composite achievement) which, in the case of interactions between pairs of focal variables (SES and size), very closely parallels results presented in Tables 3 through 9. We observe the following results (again, directionality is given parenthetically):

(1) The single-level interactions of FREEPCT and SPANSIZE, whether school- or district-level influences, are not statistically significant.

(2) The interaction (-) of SPANSIZE as a district-level influence and FREEPCT as a school- level influence is highly significant ($p<.001$).

(3) The interaction (-) of SPANSIZE as a school- level influence and FREEPCT as a district-level influence is highly significant ($p<.001$).

The two significant interactions together account for an additional 10.7% in the variation of school performance. Thus, the two-level model accounts for 79.2% of

BEST COPY AVAILABLE

the variance in the performance of Georgia schools with an 8th grade. In other words, just 20% of the variance in school performance is the result of other influences—including school processes (such matters as curriculum and instruction).

The first interaction, the statistically significant and negative interaction of district-level SPANSIZE by school- level FREEPCT, shows two things. First, as district sizes increase, the mean achievement cost associated with increases in the proportion of less-advantaged students at the school level increases as well. (Note 8) Second—as in the previously reported single-level analyses—the converse also pertains: As district sizes decrease (negative values of district size as a centered variable), the mean achievement cost associated with decreases in the proportion of less-advantaged students (i.e., negative values on school-level poverty) at the school level increases as well. In other words, more affluent school-communities appear to be better served by being in larger districts, but less affluent school-communities appear to be better served by being in smaller districts. Put most simply, district poverty and large school size are shown to jointly hurt predicted school-level performance, just as district affluence and small school size are shown to do. The relationship is interactive—it cuts two ways.

The second interaction, the statistically significant and negative interaction of school SPANSIZE by district FREEPCT follows the preceding interpretation. First, as school sizes increase, the mean achievement cost associated with increases in the proportion of less-advantaged students at the district level also increases. Second, as above, the converse is true as well: As school sizes decrease, the mean achievement cost associated with being in a district with decreases in the proportion of less-advantaged students also increases. The simple form of this statement, again, would be: school poverty and large district size are shown to hurt predicted school-level performance, just as school affluence and small district size are shown to do. Again, this interactive relationship cuts two ways

Eleventh Grade "Excellence Effects"

Tables 11-15 present the regression results using the five eleventh grade outcome measures. As predicted, the 11th grade results are less consistent than the 8th grade regressions (Tables 3 through 10). Interestingly, the cross-level interaction of school SPANSIZE by district FREEPCT is highly statistically significant, alone accounts for as much as 15% of the variance in school-level performance, and exhibits the expected negative sign in each equation. As with the 8th grade results, this means that as school sizes increase, the mean achievement cost associated with being in districts with increasingly less- advantaged students also increases. As before, large schools in low-income districts encounter a decided achievement disadvantage. Overall, the 11th grade "excellence" effects of size are considerably muted, and they have their mark most particularly with the cross-level interaction of SPANSIZE and FREEPCT. (Note 9)

In general, the 11th grade results account for less variance than the 8th grade results. In the case of the composite score (Table 15), for instance, the model explains about 50% of the variance in school-level performance. The greatest proportion of variance accounted for by our model appears for mathematics (about 66%); the low is English (less than 30%). Mathematics, we observe, is a highly differentiated school subject at the high-school level, with the first course in algebra serving in the famous "gatekeeper" role (Silva & Moses, 1990) (Note 10). In other words, structural influences (poverty, race, size and the interactions among them) might exert a stronger influence on school performance than they would in less differentiated subjects such as English.

Table 15
Weighted Regression Results with Corrected Standard Error
Grade 11: Composite Score

Unstandardized and (Standardized) Regression Coefficients

	Individual-Level	Group-Level	Individual by Group Interactions
SPANSIZE	3.688	19.952	-133.985
	(.027)	(.052)	(-.100)
FREEPCT	-0.413***	-0.187*	-0.003
	(-.304)	(-.222)	(-.066)
BLACKPCT	-0.257***	-0.116**	-0.001
	(-.378)	(-.206)	(-.067)
MINORPCT	0.321	0.262	-0.015
	(.088)	(.070)	(-.028)
S/RRATIO	-0.915*	-0.145	0.096
	(-.126)	(-.013)	(.060)
Variance Explained	28.7%	10.0%	1.7%

Within-Level and Cross-Level Interactions

SCHOOL SPANSIZE by SCHOOL FREEPCT	0.281 (-.075)
DISTRICT SPANSIZE by DISTRICT FREEPCT	-0.423 (-.025)
DISTRICT SPANSIZE by SCHOOL FREEPCT	-0.027 (-.001)
SCHOOL SPANSIZE by DISTRICT FREEPCT	-1.357*** (-.456)
Variance Explained	8.0%

Residual Intraclass Correlation .066
School/District Ratio 1.96
Standard Error Inflation 5.96% (Corrected)

Partial Derivatives for Y with Respect to (1) SCHOOL SPANSIZE and (2) DISTRICT SPANSIZE

Y wrt 1= - 1.357 x (DISTRICT SPANSIZE)
Y wrt 2 = COEFFICIENTS NOT STATISTICALLY SIGNIFICANT

*p <.05
** p <.01
***p <.001

Interpreting the Effect Sizes of Size

The regression equations provide a prospective tool with which to estimate the effects of projected changes in size (of schools and districts) on school performance in Georgia relevant to the independent variables that describe a school's context. In order to interpret these predicted effects of size on school performance, we adapt the technique pioneered by Friedkin and Necochea (1988).

Those researchers differentiated their regression equations in order to infer a rate of change in achievement attributable to size, relative to a school's or district's poverty level. Their procedure found the partial derivative (Note 11) of school or district performance with respect to socioeconomic status. The partial derivative was then evaluated to find the rate of achievement change associated with changes in school or district size for schools or districts of a certain SES. This is the technique also used in the work recently reported by Bickel and Howley (e.g., Howley & Bickel, 1999).

Since our goal here is to provide a fuller quantitative account of the relationship between size and SES we have computed partial derivatives of the regression equations that give the rate of change in the dependent variable (school performance) with respect to size (school or district), holding poverty (FREEPCT) constant (at two levels of influence). It is important to remember that the dependent

variable in the partial derivatives represents a rate: change in school performance per change in size.

Because this is a two-level analysis, however, two equations are necessary. One equation describes the predicted influence of changes in school size on school performance, and in this analysis that rate turns out to be a function of district-level variables. The other equation describes the predicted influence of changes in district size on school performance (in this case as a function of school-level variables). Think of this relationship as follows: Y wrt 1 is a rate of change in school performance per change in the size of a school. But this rate, in cross-level analysis, depends on district-level characteristics. Y wrt 2 is a rate of change in school performance per change in district size; this rate depends in cross-level analysis on school-level characteristics. Both equations can be standardized to give rate of change in standard deviation units if desired.

Of most importance to this analysis, however, is the prediction of total change resulting from the joint influence of variables at both levels. Computing this rate of change requires that the two partial derivatives be combined. To effect this combination, we calculate the total differential. The total differential predicts the magnitude of influence of changes in size (of both schools and districts) on school performance (which is always the dependent variable in these analyses), all else equal. Let us begin by explaining the partial derivatives. In the immediately subsequent section, however, we provide an explanation of and illustrate the use of the total differential, as it constitutes the most important interpretation of size effects jointly interaction with poverty.

Partial Derivatives. In Tables 3-15 we report two partial derivatives, one for each level of influence (school and district) separately. Partial derivatives give the rate of change in a dependent variable produced by focal variables (SPANSIZE and FREEPCT, in the present case), holding constant all other variables (i.e., BLACKPCT, MINORPCT, and S/RRATIO). Readers need to understand how they may use these additional equations. (Note 12) We will use the 8th grade composite statistics (Table 10) to illustrate our procedure, and we explain both the creation of partial derivatives and the calculation of the total differential. First, taking the partial derivative of Y with respect to SPANSIZE at the school level ("Y wrt 1" in Table 10) tells us that the rate of change in Y with respect to SCHOOL SPANSIZE, holding constant the other independent variables, is equal to:

$$f\,x1'(y) = [(-308.619)(\text{DISTRICT SPANSIZE})] - [(1.046)(\text{DISTRICT FREEPCT})]$$

Similarly, using the same outcome measure, taking the partial derivative of Y with respect to SPANSIZE at the district level tells us that the rate of change in Y with respect to DISTRICT SPANSIZE, holding constant the other independent variables, is equal to:

$$f\,x2'(y) = [(-308.619)(\text{SCHOOL SPANSIZE})] - [(4.304)(\text{SCHOOL FREEPCT})]$$

The first partial derivative enables us to see that, all else equal, if we increased the value of DISTRICT SPANSIZE by, say, one quarter standard deviation unit (.025 = .25 x .101), the predicted outcome measure would decrease by 7.7 points. Similarly, if DISTRICT FREEPCT were increased by one quarter standard deviation unit (4.4 = .25 x 17.5), the outcome measure would decrease by 4.6 points. These effects, of course, are additive, and changes of equal magnitude, but in the contrary directions, would yield no net effect.

The second partial derivative enables us to determine the effect on 8th grade composite scores of an increase or decrease in SCHOOL SPANSIZE and SCHOOL FREEPCT. A one quarter standard deviation unit increase in SCHOOL SPANSIZE (.031 = .25 x .124) yields a 9.6 point decrease in the outcome measure. A one quarter point standard deviation unit increase in SCHOOL FREEPCT (5.73 = .25 x 22.9) yields a 24.7 point decrease in the outcome measure.

The Total Differential

Information about the composite relationship between size and achievement is provided by the total differential. The total differential (dy) is the sum of the products of the partial derivatives and their differentials, dx_1 and dx_2, where dx_1 represents a change in SCHOOL SPANSIZE and dx_2 represents a change in DISTRICT SPANSIZE. The total differential, then, is the sum of the changes in measured achievement due to changes in SCHOOL SPANSIZE and DISTRICT SPANSIZE, contingent on SCHOOL FREEPCT and DISTRICT FREEPCT (all else equal):

$$dy = \{[fx_1'(y)](dx_1)\} + \{[fx_2'(y)](dx_2)\}$$

The values of dx_1 and dx2 represent proportional changes (e.g., -.10 or +.10) in school or district size (SPANSIZE). To illustrate the calculation of the total differential, we computed hypothetical values of dx_1 and dx_2 tied to real-life values in the Georgia data set. We divided the SPANSIZE into the difference between SPANSIZE and the difference between the value of SPANSIZE for cases n + 1 and case n. That is, using the subsequent case in the data set as a reference point, we inferred rates change for school and district size in the subject case. This procedure produces arbitrary changes, but these arbitrary changes vary only within the range of variation that the Georgia school system exhibits.

In keeping with Dowling's (1980) admonition that differentials should be realistically small, we then eliminated cases with values for dx_1 or dx_2 greater than one-half standard deviation above or below their mean. (Note 13) The absolute value of dx_1 for all remaining cases was less than .068, and the absolute value of dx_2 was less than .026. We then randomly selected ten of the remaining schools for inclusion in Table 16.

Table 16
Total Differential: Illustrative Values for Randomly Selected Cases

Grade Eight Composite Scores

DISTRICT SPANSIZE	SCHOOL SPANSIZE	DISTRICT FREEPCT	SCHOOL FREEPCT	dx_1	dx_2	dy
.0829	.0835	79.47	71.79	-.047	-.010	8.47
.1562	.2187	73.38	74.90	-.019	-.009	6.08
.2285	.3427	20.21	0.90	-.066	-.005	6.63
.1541	.1527	24.34	27.10	.013	.018	-3.88
.1437	.2770	70.84	66.20	-.029	-.013	8.21
.1469	.1497	61.07	59.70	-.108	-.006	13.62
.1311	.1270	66.60	61.20	.005	.010	-3.56
.1825	.2120	29.76	19.50	.000	.005	-0.70
.0980	.0944	55.89	48.10	-.062	.013	2.46
.1566	.3060	47.39	55.30	.016	.016	-6.75

Notes.
Values of variables are given *uncentered*. Equations are derived from and computed with centered values.

Total differential computed as: $dy = \{[f_{x1}'(y)](dx_1)\} + \{[f_{x2}'(y)](dx_2)\}$

Values of partial differentials, dx_1 and dx_2 computed as follows(cases selected for $|dx| \leq 0.5\sigma$):

$$[(SPANSIZE_{case(n+1)}) - (SPANSIZE_{case(n)})] / (SPANSIZE_{case(n)})$$

The first four columns in Table 16 describe the focal variables (district and school size and subsidized meal rates). The fifth and sixth columns provide the (hypothetical) proportional changes in school size (dx_1) and district size (dx_2). The values of the total differential—the predicted change in each school's mean Composite Test score attributable to these composite changes in size—contingent on these proportional changes in school and district size appear in the column headed "dy" ("total differential").

Observe that Table 16 illustrates the inverse relationships between school performance (8th grade composite, in this case) and changes in SPANSIZE at both the school level and the district level. The first two cases, for instance, show a positive influence of joint school and district size in a uniformly impoverished school and district. Case seven shows the decline in similar circumstances of a joint increase in size. And case nine shows the somewhat more modest increase in test scores resulting from a joint reduction in school size and increase in district size.

Eighth and Eleventh Grade "Equity Effects"

Most people understand inequity in school finance. Affluent communities almost always enjoy better-funded schools, and improvements in financial equity would require that schools in impoverished communities be much better funded than they are. In other words, mitigating financial inequity requires that we break the link between poverty and school finance. Some educators (we among them) believe that no ethical principle justifies the privilege enjoyed by more affluent citizens in this regard. Why should the rich enjoy the best-funded schools? The rich commonly argue that it is their right, and the argument prevails.

Inequity in achievement presents much the same case. Which children, in general, enjoy the highest achievement? More affluent children do. Some observers, of course, believe that since the constructs "affluence" and "ability" correlate well, this state of affairs is actually very fair. The rich might well argue that inequity of outcomes in their favor is also their right. Others (we among them) note that—among affluent and impoverished people alike—a great range of abilities exists, and that in all adult occupations a similarly great range of abilities persists. On this view, the low achievement of impoverished children is not nearly so fair as it at first might seem (e.g., Gardner, 1983). In this view, public schooling can and should do much more to nurture the learning of impoverished students, in particular among all students. As with financial equity, equity in achievement means breaking—or at least substantially mitigating—the prevailing bond between SES and achievement. (Note 14)

Table 17
Multi-Level Georgia Equity Effects[a]

Larger v. Smaller Schools and Districts with Grades 8 and 11

Composite

	Grade 8		Grade 11	
	Districts			
	L[b]	S	L	S
Schools L	.76	.72	.77	.74
S	.63	.35	.54	.16

Reading Comprehension (8)/ English (11)

	Grade 8		Grade 11	
	Districts			
	L	S	L	S
Schools L	.84	.74	.69	.59
S	.71	.36	.28	.16

Mathematics

	Grade 8		Grade 11	
	Districts			
	L	S	L	S
Schools L	.71	.59	.72	.65
S	.46	.29	.48	.25

Science

	Grade 8		Grade 11	
	Districts			
	L	S	L	S
Schools L	.82	.73	.73	.71
S	.70	.37	.46	.27

Notes:
a) Variance (R^2) in school performance attributable to school-level subsidized meal rates.
b) L = Larger half; S = Smaller half.

Table 17 gives the variance in achievement associated with SES in four groups by the medians of district size and school size (2 grades and 4 tests). Within each panel, by grade level, we report the observed variances proceeding left to right and top to bottom in each of the 8 contrasts for: (1) large schools in large districts, (2) large schools in small districts, (3) small schools in large districts, and (4) small schools in small districts.

In each of these 8 (2 grade levels by 4 tests) four- way contrasts, large schools in large districts show the highest proportion of variance in achievement associated with SES: between 71% and 84%, whereas the lowest proportion of variance is exhibited among small schools in small districts: between 16% and 27%. Moreover, the order of declining variance follows an identical pattern in each of the 8 contrasts: large-large, large-small, small-large, and small-small. In 6 of 8 cases, the largest magnitude of decline within the evident sequence (large-large, large- small, etc.) of decreasing variance comes in the change from small schools in large districts to small schools in small districts.

In other words, Table 17 suggests that the predicted equity effect of reducing district size but not school size would be practically significant; the predicted equity effect of reducing school size but not district size would also be practically significant and perhaps somewhat larger; and the combined strategy of reducing both school and district size would be predicted to yield substantial equity and excellence effects (given the previous multi- level regression analyses).

Some rural states (e.g., Montana; see Howley 1999b) structure their school

systems in just this way. That is, such systems have chosen to sustain small schools within small districts. The Montana system doubtless has plenty of room for "improvement," but on the terms of accountability (and the value of more equal outcomes), Montana is an exemplar. Please note that Montana has a substantial American Indian population (13%), whose children also attend small, predominantly public, schools and districts.

In rural areas, the phenomena of school and district size seem mutually dependent; larger rural schools often prevail in larger rural districts (e.g., as in West Virginia; see Howley, 1996). District "reorganization" has often been a first step toward eliminating small schools (DeYoung, 1995; Peshkin, 1982). This strategy would be predictably harmful to the achievement of students in impoverished rural communities. In the southeast US a single high school now often serves entire rural counties, covering large geographical areas.

The situation in urban areas is equally bad, though in somewhat different ways. The huge big-city districts were created, not just to improve schools, but to destroy a resource (school jobs) that could be controlled by ward politics. Usually portrayed as a "progressive" change, an important motive of city fathers was to wrest power back from the hands of working-class urban communities (e.g., Tyack, 1974; Erie, 1988). Today, most urban districts are nightmares and wildernesses of bureaucracy and outright fear (e.g., Devine, 1996). Jobs are as much a political issue as ever in many large cities, but the power to dispense them has shifted to the nexus between political regimes and school bureaucracies, with the bureaucracy often in the better position. No wonder so many thoughtful educators champion the re-establishment of smaller schools in cities (e.g., Meier, 1995; Klonsky, 1995).

Difficult as it is, in both rural and urban locales, to defend or re-establish small schools, that task leaves the structural challenge incomplete. Seldom are reductions in district size—especially in the case of large city districts—seriously considered.

Our principal "clear and simple" recommendation therefore is to suggest the wisdom, of reorganizing districts that are now far too large. Policy makers should start imagining ways to re-create districts that are everywhere sufficiently small to respond well to students, families, and (especially) communities. One way to enable this decision making might be for communities to enjoy the right to charter public school districts as well as public schools (and, naturally, to receive the requisite state- level support to succeed). The policy issues are surely difficult, but no more difficult than those that have already led to the counterproductive structuring that presently prevails. To do nothing or little leaves the burden of coping with the enormity to impoverished students, families, and communities—exactly where it currently rests.

Misuse of the findings

Our findings cannot be interpreted to warrant the construction of huge schools, however, even for relatively comfortable communities; in general, we advise an upper limit of about 250 students per grade for 9-12 high schools and about 100 students per grade for elementary schools— and these rule-of-thumb upper limits apply to communities where the poverty rate is zero (Howley, 1997; but see Irmsher, 1997, and Raywid, 1999, for quite similar recommendations based on recent reviews of the literature).

Recently we learned that our research was being used to help justify construction of a school in a semi-rural area of an eastern state proposed to house 2,000 elementary students in grades 3-6. In view of extant and easily accessible research syntheses such as those by Irmsher and Raywid, proposals to create schools of this size—particularly elementary schools—are, we believe, capricious and professionally irresponsible.

We are unhappy (but not surprised) to learn that our work has been deployed to support such proposals; but we also understand the role that bad state-level policy plays in shaping such decisions as this (see Purdy, 1997, for a clear example in a rural state where the state influence is heavy-handed). The administration in this district experienced considerable angst when community members there contacted us and we voiced our objections to the misuse of our research publicly. In fact, however, we are used to being contacted by community members resisting such

efforts and equally used to not hearing from members of our own profession as they make construction plans. Despite uproar in the community and defeat of the bond issue, plans for the mega-school (to be organized in "houses") apparently continue. The superintendent in this case has reportedly vowed revenge on the interfering outside researchers! We regret the angst that emerges in these situations, but we believe the present study provides evidence to support our evolving position on the issues.

Conclusion

Small size is good for the performance of impoverished schools, but it now seems as well that small district size is also good for the performance of such schools in Georgia, where district size, in single-level analyses, had revealed no influence. Because of the consistency of school-level findings in previous analyses, we strongly suspect that the Georgia findings characterize relationships in most other states. This claim can, of course, only be evaluated by additional replications, and we hope other researchers will see merit in such work.

The equity effects reported here, however, extend the evidence of the previous single-level studies to the interaction of school and district size. Larger schools in larger districts seem to propagate inequality of outcomes by comparison to smaller schools and smaller districts. In fact, smaller schools in larger districts demonstrate a useful equity effect, as well. For large schools in smaller districts, however, the improvements in equity might be so slight as to be called "negligible."

The equity effects are so striking, and appear so instrumental in association with the "excellence" effects of small size in impoverished communities, that further investigation into this mitigating influence would seem crucial. How does the principle evident in the findings apply to individual students? In what settings? To what extent? What structural features of small size enable such an effect? How do impoverished students fare in schools that are, overall, rather affluent? Is an overall upper limit to school size and district size worth establishing by policy? How should such upper limits be set? What policies can succeed in recreating smaller districts in big cities and the rural southeast?

These are interesting and important questions, we think, but the conclusions of this study would seem to require rather wide debate and reconsideration of the size issue, across the spectrum of poverty and wealth, and not just in the case of impoverished communities. We note that America's elite sends its children to Andover and Exeter and other such fine high schools, where enrollments seldom exceed 1,500. What do they know that the rest of us have yet to learn, we wonder?

Notes

This work is based on research funded by the Rural School and Community Trust. It does not, however, represent the opinions or positions of the Rural Trust. We are grateful for the support; the errors and opinions are our own.

The two authors are equal contributors to the work reported here.

1. Unlike the other states, Montana has retained many small schools, and this historic decision is a likely cause for the weak interaction effects. Bickel (1999b) also reported no interaction effect among the 132 Texas schools that house all students in grades K-12 (cf. Franklin & Glascock, 1998).
2. Magnitude of the relationship was measured as the proportion of variance in achievement associated with SES.
3. See the Appendix for a discussion of the problem that intraclass correlation poses to the use of ordinary least squares regression. The Appendix describes the conditions needed to use OLS in multi-level analysis and shows that our data set meets these conditions.
4. This logic, of course, is also supported by the findings previously reported for the single-level four-state analyses, in which reported effects are strongest at grade 8 or 9, and always weaker at grades 11 or 12. On the basis of past experience, then, we would have reason to suspect similar results.

5. We report statistical significiance levels as a gauge to practical significance. Because the data set includes practically all schools in Georgia, the relationships that emerge are those that prevail, and, we maintain, should not be considered as subject to sampling error.
6. Directionality of the influence is given in parentheses following the variable name. The effect of centering is not reflected in Tables 1 and 2.
7. These findings are conceptually consistent with previously reported school-level analyses, which found that, among impoverished communities, smaller schools reduced the achievement costs of poverty and that large ones magnified such costs; but the converse was true as well, in those cases: Among affluent communities, smaller schools increased the achievement costs of affluence and larger ones reduced such costs.
8. "Mean achievement costs" represent declines in predicted achievement. Therefore, another way to put this interactive relationship is this: (1) as poverty and district size continuously increase, predicted school performance continuously declines; and (2) as poverty decreases and district size decreases, predicted performance also continuously declines.
9. We might also observe that other cross-level interactions appear significantly in the equations reported in several of these Tables: Table 11 (math: FREEPCT), Table 13 (social studies: SPANSIZE), and Table 14 (science: BLACKPCT). Cross-level structural influences are weak at the 11th grade but still evident.
10. Robert Moses's "Algebra Project" construes algebra as the course that governs access to the academic track in life; failing algebra, or never taking it in the first place, marks one as academically inept.
11. A "derivative" can be understood as the calculus tool for determining the "slope" of a curved line (which, in geometrical terms, is the tangent of the curve at a given point). The slope of such a line is constantly changing (just as the effects of school or district size, or their joint effects, constantly change with respect to poverty levels), and the derivative provides the formula for calculating this changing slope. To find this changing rate, one "takes the derivative" of the formula that describes the line. A partial derivative holds one variable constant during differentiation (the process of "taking the derivative") so that the influence of that variable can be subsequently evaluated. This process of "holding an influence constant" is similar to calculating a partial correlation coefficient.
12. Consult Howley (1996) for a complete description of the derivation of partial derivatives in the single-level analyses.
13. Dowling's counsel is important because we are dealing, in using calculus techniques that estimate changing rates, with how these rates of change at the margin (i.e., the usual addition or loss of a few students) under normal conditions, and not, in fact, in such catastrophic alterations as are produced by consolidations of two or more schools (where size may well increase by hundreds of students). Calculus is the mathematics of smooth curves and not of disruption and disjunction.
14. In practical terms, one is unlikely to break the bond completely, because the negative effects of poverty can be eliminated only when a society finds them intolerable and actively cultivates the well-being of the poor. Even in the current economic boom, however, such a realization has not overtaken the US, and in general, the gap between the affluent and the impoverished is growing ever wider here. Also, some observers balk when they realize that breaking the bond must apply not just to the poor, but to the affluent as well.

References

Alspaugh, J. (1995, October). *A comparison of four enrollment groups of K-8 and K-12 Missouri rural school districts.* Paper presented at the annual meeting of the National Rural Education Association, Salt Lake City, UT. (ERIC Document Reproduction Service No. ED 389 501)

Barcikowski, R. (1981) Statistical power with the group mean as the unit of analysis.

Journal of Educational Statistics, 6(3), 267-285.

Barker, R., & Gump, P. (1964). *Big school, small school*. Stanford, CA: Stanford University Press.

Barr, R., & Dreeben, R. (1983). *How schools work*. Chicago: University of Chicago Press.

Betts, J. (1996). Is there a link between school inputs and earnings? Fresh scrutiny on an old literature. In Burtless, G. *Does money matter*? Washington, DC: Brookings.

Bickel, R. (1999a). *School size, socioeconomic status, and achievement: A Georgia replication of inequity in education*. Randolph, VT: Rural Challenge Policy Program.

Bickel, R. (1999b). *School size, socioeconomic status, and achievement: A Texas replication of inequity in education*. Randolph, VT: Rural Challenge Policy Program.

Bidwell, C., & Kasarda, J. (1975). School district organization and student achievement. *American Sociological Review, 40*(1), 55-70.

Boyd, L., & Iversen, G. (1979). *Contextual analysis: Concepts and statistical techniques*. Belmont, CA: Wadsworth Publishing Co.

Bruner, J. (1996). *The culture of education*. Cambridge, MA: Harvard University Press.

Bryk, A., & Raudenbush, S. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Burtless, G. (Ed.). (1966). Does Money Matter: Washington, DC: Brookings.

Caldas, S. (1993). Re-examination of input and process factor effects on public school achievement. *Journal of Educational Research, 86*, 206-214.

Clones, D. (1998). *Development report card for the states* (12th ed.). Washington, DC: Corporation for Enterprise Development.

Conant, J. (1959). *The American high school today*. New York: McGraw-Hill.

Cronbach, L. (1987) S atistical tests for moderator variables: Flaws in analyses recently proposed. *Psychological Bulletin, 102*(3), 414-417

Devine, J. (1996). *Maximum security*. Chicago: University of Chicago Press.

Dorn, S. (1998). The political legacy of school accountability systems. *Education Policy Analysis Archives, 6*(1) [On-line serial]. Available World Wide Web: http://olam.ed.asu.edu/ epaa/v5n12.html

Dowling, E. (1980). *Mathematics for economists*. New York: McGraw-Hill.

Erie, S. (1988). *Rainbow's end: Irish Americans and the dilemmas of urban machine politics*. Berkeley, CA: University of California Press.

Farkas, G. (1996). *Human capital or cultural capital? Ethnicity and poverty groups in an urban school district.* Hawthorne, NY: de Gruyter.

Fowler, W. (1995). School Size and Student Outcomes. *Advances in educational productivity, 5*, 3- 26.

BEST COPY AVAILABLE

Fowler, W. (1991). School size, characteristics, and outcomes. *Educational evaluation and policy analysis, 13*, 189-202.

Franklin, B., & Glascock, C. (1998). The relationship between grade configuration and student performance in rural schools. *Journal of Research in Rural Education, 14*(3), 149-153.

Friedkin, N., & Necochea, J. (1988). School system size and performance: A contingency perspective. *Educational Evaluation and Policy Analysis, 10*(3), 237-249.

Gamoran, A., & Dreeben, R. (1986). Coupling and control in educational organizations. *Administrative Science Quarterly, 31*: 612-632.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Georgia Department of Education. (1999). *A messsage from the state superintendent of schools: Welcome to the 1997-1998 Georgia public education report card.* Atlanta, GA: Author. (Available World Wide Web: http://168.31.216.185/RCP/LindaLtr.html)

Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.

Gujarati, D. (1995). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

Guthrie, J. (1979). Organizational scale and school success. *Educational Evaluation and Policy Analysis, 1*, 17-27.

Haller, E. (1992). High school size and student indiscipline: Another aspect of the school consolidation issue. *Educational Evaluation and Policy Analysis, 14*, 145-156.

Haller, E., Monk, D., & Tien, L. (1993). Small schools and higher order thinking skills. *Journal of Research in Rural Education, 9*, 66-73.

Hanushek, E. (1998). Conclusions and Controversies about the Effectiveness of School Resources. Federal Reserve Bank of New York Economic Policy Review, 4: 11-28.

Hanushek, E. (1997). Assessing the effects of school resources on student performance: An update. Educational Evaluation and Policy Analysis, 19: 141-164.

Hemmings, B. (1996). A longitudinal study of Australian senior secondary school achievement. *Issues in Educational Research, 6*, 13-37.

Henderson, R., & Raywid, M. (1996). A "small" revolution in New York City. *Journal of Negro Education, 63*, 28-45.

Howley, C. (1999a). *The Matthew Project: State report for Montana*. Randolph, VT: Rural Challenge Policy Program.

Howley, C. (1999b). *The Matthew Project: State report for Ohio*. Randolph, VT: Rural Challenge Policy Program.

Howley, C. (1997). Dumbing down by sizing up: Why smaller schools make more sense—if you want to affect student outcomes. *The School Administrator, 54*(9), 24-26, 28, 30.

Howley, C. (1996). The Matthew Principle: A West Virginia Replication? *Education*

Policy Analysis Archives, 3(1), (Entire Issue.) [Available online: http://epaa.asu.edu/epaa/v3n18.html]

Howley, C. (1989). *What is the effect of small-scale schooling on student achievement?* (Digest EDO-RC-89-6). Charleston, WV: ERIC Clearinghouse on Rural Education and Small Schools. (ERIC Document Reproduction Service No. ED 308 062)

Howley, C. (1989). Synthesis of the effects of school and district size: What research says about achievement in small schools and school districts. *Journal of Rural and Small Schools, 4*, 2-12.

Howley, C., & Bickel, R. (1999). *The Matthew Project: National report.* Randolph, VT: Rural Challenge Policy Program.

Huang, G., & Howley, C. (1993). Mitigating disadvantage: Effects of small-scale schooling on student achievement in Alaska. *Journal of Research in Rural Education, 9*(3), 137-149.

Irmsher, K. (1997). *School size* (ERIC Digest). Eugene, OR: ERIC Clearinghouse on Educational Management. (ERIC Document Reproduction Service No. ED 414 615)

Iversen, G. (1991). *Contextual analysis*. Newbury Park, CA: Sage.

Khattari, N., Riley, K., & Kane, M. (1997). Students at risk in poor rural areas: A review of research. *Journal of Research in Rural Education, 13*, 79- 100

Klonsky, M. (1995). *Small schools: The numbers tell the story: A review of recent research and experiences*. Chicago, IL: Small Schools Workshop. (ERIC Document Reproduction Service No. ED 386 517)

Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.

Lamdin, D. (1995). Testing for the effect of school size within a district. *Educational Economics, 3*, 33-42.

Lee, V. and Smith, J. (1997). High school size: Which works best, and for whom? *Educational Evaluation and Policy Analysis, 19(3), 205-227.*

McDill, E., Natriello, G., & Pallas, A. (1986). A population at risk: Potential consequences for tougher school standards for student dropouts. In G. Natriello (Ed.), school dropouts: Patterns and policies. New York: Teachers College Press.

Meier, D. (1995). The power of their ideas: Lessons for America from a small school in Harlem. Boston: Beacon Press.

Meyer, J., Scott, W., & Strang, D. (1987). Centralization, fragmentation, and school district complexity. Administrative Science Quarterly, 32(2), 186-201.

Mik, M., & Flynn, M. (1996). School size and academic achievement in the HSC examination: Is there a relationship? Issues in Educational Research, 6, 57- 78.

Peshkin, A. (1982). The imperfect union: School consolidation and community conflict. Chicago: University of Chicago Press.

Prichard Committee for Academic Excellence. (1990). The path to a larger life: Creating Kentucky's educational future. Lexington, KY: University of Kentucky Press. (ERIC Document Reproduction Service No. ED 323 055)

Purdy, D. (1997). An economical, thorough, and efficient school system: The West

Virginia School Building Authority "economy of scale" numbers. Journal of Research in Rural Education, 13, 70-82.

Raywid, M. (1999). Current literature on small schools (EDO-RC-98-8). Charleston, WV: ERIC Clearinghouse on Rural Education and Small Schools. (ERIC Document Reproduction Service No. ED 425 049)

Raywid, M. (1996). Taking stock: The movement to create mini-schools. New York: ERIC Clearinghouse on Urban Education. (ERIC Clearinghouse Accession No. UD 031 001)

Riley, R. (1999, October 13). Schools as centers of community (Remarks as prepared for delivery by Secretary of Education Richard W. Riley to the American Institute of Architects). Washington, DC: US Department of Education. (Available World Wide Web: http://www.ed.gov/Speeches/10-1999/991013.html)

Riordan, C. (1997). Equality and achievement. New York: Longman.

Rivkin, S., Hanushek, E., & Kain, J. (1998). Teachers, schools, and academic achievement (Working Paper Number 6691). Washington, DC: National Bureau of Economic Research.

Rossmiller, R. (1987). Achieving equity and effectiveness in schooling. Journal of Educational Finance, 4, 561-577.

Rural Trust. (1999, November). Riley recognizes community role in school design and use. Rural Policy Matters, p. 2.

Silva, C., & Moses, R. (1990). The algebra project: Making middle school mathematics count. Journal of Negro Education, 59(3), 375-91.

Singer, J. (1987) An intraclass correlation model for analyzing multilevel data. Journal of Experimental Psychology, 18(4), 219-228.

Smith, D., & DeYoung, A. (1988). Big school vs. small school: Conceptual, empirical, and political perspectives on the re-emerging debate. Journal of Rural and Small Schools, 2, 2-11.

Stevens, N., & Peltier, G. (1994). A review of research on small-school participation in extracurricular activities. Journal of Research in Rural Education, 10, 116-120.

Stevenson, K. (1996). Elementary school capacity: What size is the right size? The Educational Facility Planner, 33, 10-14.

Stiefel, L., Berne, R., Iatarola, P., & Fruchter, N. (2000). High school size: Effects on budgets and performance in New York City. Educational Evaluation and Policy Analysis, 22(1), 27-39.

Thiétart, R.A., & Forgues, B. (1995). Chaos theory and organization. Organization Science, 6(1), 19-31.

Tyack, D. (1974). The one best system: A history of American urban education. Cambridge, MA: Harvard University Press.

Walberg, H., & Walberg, H. (1994). Losing local control of schools. Educational Researcher, 23, 19- 26.

Wyatt, T. (1996). School effectiveness research: Dead end, damp squib, or smoldering fuse? Issues in Educational Research, 6, 79-112.

About the Authors

Robert Bickel
Marshall University
Email: bickel@marshall.edu

Robert Bickel is a Professor of Advanced Educational Studies at Marshall University. His recent research is concerned with school size as a variable which moderates the relationship between social class and measured achievement, evaluation of early childhood interventions, and with contextual factors which occasion the at-risk designation.

Craig Howley

Appalachia Educational Laboratory, Inc.
Ohio University
75619 Lively Ridge Road
Albany, OH 45710
Phone: 740.698.0309
Email: howleyc@ael.org

Craig Howley is an education writer and researcher based in Albany, Ohio, and affiliated part-time with Ohio University (Athens, OH) and AEL, Inc. (Charleston, WV). Ongoing scholarly projects in which he is a partner include the following: research on school and district size, rural school busing, and three book projects: an examination of small rural high schools (for Appalachian Educational Laboratory), an extended interpretation of developmentalism as a school ideology (with Aimee Howley) and a textbook (with A. Howley and Ohio University colleagues) on school administration.

Appendix

Ordinary Least Squares Regression and the Problem of Intraclass Correlations

One of the assumptions of ordinary least squares estimators is that residuals are not correlated. However, in a multi-level analysis this assumption may be erroneous. The reason is that first-level observations are located within the groups that constitute the second level of analysis. Grouping of first-level observations (schools) into districts may well mean that schools within a district are more like each other than they are like schools in other districts. The consequence is intraclass correlation, or covariance among residuals for schools in the same district (see Kreft & de Leeuw, 1998, pp. 9-10).

This observation yields the primary objection to traditional contextual models such as ours. Through uncritical use of ordinary least squares, the magnitude of standard errors of regression coefficients may be underestimated and alpha levels artificially inflated (Goldstein, 1995). The observation holds even though ordinary least squares estimators remain unbiased (Barcikowski, 1981).

In the present study, intraclass correlations, which vary by outcome measure and grade level, range in magnitude from .048 to .101. The number of groups or districts is 158 for the 8th grade and 155 for the eleventh grade. With 367 schools reporting 8th grade test scores, and 298 reporting eleventh grade scores, the relative number of second-level observations is large, indeed (Goldstein, 1995).

We conclude that intraclass correlation is a negligible problem. Given this confluence of circumstances—small intraclass correlations and large numbers of districts relative to the number of schools—ordinary least squares will yield estimates which are unbiased and will provide such estimates with very little inflation of regression coefficient variances (Singer, 1987). Furthermore, using a procedure presented by Singer (1987), we have calculated the remaining modest inflation of regression coefficient variances, standard errors, and resulting t-values. We compensated for this statistical artifact when running tests of significance, reducing the magnitude of the affected statistics by the amount they are inflated due

to intraclass correlation.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William Mclnerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina — Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz Garcia (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatistica
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

archives | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 1340 times since May 12, 2000

prior vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 8 Number 23 May 12, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

School-based Standard Testing

Craig Bolon
Planwright Systems Corporation
Brookline, MA (USA)

Abstract

School-based standard testing continues to evolve, yet in some ways it remains surprisingly close to its roots in the first two decades of the twentieth century. After use for many years as a diagnostic and as a filter for access to education, in the closing years of the century it has been pressed into service for state-run political accountability programs. In this role, it is generating vehement controversy that recalls protests over intelligence testing in the early 1920s. This background article explores primary characteristics and issues in the development of school-based standard testing, reviews its typical lack of qualification for political accountability programs, and suggests remedies to address major problems. In general, the attitude toward new techniques of assessment is skeptical, in light of the side-effects and unexpected problems that developed during the evolution of current techniques.

Survival of the Fittest

School-based standard testing began a dream decade in the early 1950s, driven by waves of public anxiety over Soviet "dominos," nuclear weapons, Sputnik and the "missile gap." Now, so many years later, it can be hard to imagine the intensity of fears that the Russians were ahead of everybody else— not just in the size of their standing army but in scientific knowledge, inventions and industry. There was widespread agreement that the U. S. needed to identify talented people and train them for critical occupations. (Note 1)

Of course we know more of the dreary facts today— a Russia of gray poverty and workplace spies, burdened with heavy but narrow investment to produce arms, rockets and nuclear bombs. But in those times, who knew? We saw North Korea fortified with MiG-15s, the Hungarian revolt crushed with Russian tanks, and then the Berlin wall built. Russia had been four years behind the U. S. in testing an atomic bomb but only one year behind with its first thermonuclear blast. And although the U. S. employed the Nazi rocket designers from World War II, Soviet Russia had a space satellite first— winking at us and mocking "the American century."

And so it was, into the breach against Godless communism, (Note 2) that we launched our homespun Scholastic Aptitude and Iowa tests. Few questioned the methods or values. In the climate of those days, school-based standard testing was an engine of progress. (Note 3) It would promote technical expertise and fairly chosen leadership to right the balance and put America first again.

Background

School-based standard testing (Note 4) aims to provide uniform, rapid measurement of some kind of mental capability that is related to education. There are many other assessments related to responsibilities or occupations rather than schools. These include, for example, tests for motor vehicle drivers, aircraft pilots, divers, plumbers and power plant operators. Historical precedents for competence testing can be traced to the ancient civilizations of China (Note 5) and Rome. However, until relatively recently education operated mainly as a craft. Teachers and schools tested their students and applicants, sometimes intensely, but there was rarely interest in tests that would be applied uniformly and rapidly to large groups of students in diverse situations. Key educational credentials were instead the evaluations of students by individual teachers and schools.

It may have been public education, more than any other factor, that inspired interest in school-based standard testing. (Note 6) The U. S., with the strongest history of public schools, also had the strongest early interest in standard testing. Perhaps it should not be surprising that the country which implemented the concepts of standard machine parts and mass production should also be the country that most eagerly adopted standard testing in its rapidly growing education enterprises (see Cremin, 1962, pp. 185-192). The Yankee attitude can be perceived in the pursuit of uniformity and efficiency.

Standard Tests

The distinguishing features of a standard test are uniform administration and some form of calibration. Before routine use, standard tests or component items will be tried out with groups intended to represent populations of test- takers. These trials are used to measure distributions of scores and other properties of a test (Rogers, 1995, pp. 256-257 and 734-741). After calibration, test scores are typically reported by using a formula derived from the calibration (to percentile ranks, for example). Beginning in the 1910s, statistical metrics were developed to characterize test items and report scores (Rogers, 1995, pp. 197-208, 317-325 and 382-388). The IQ score and the SAT scaled score ranging from 200 to 800 are among the well-known metrics.

A quantitative approach helped give standard tests the appearance of objectivity and encouraged a test format that is easily adapted to numerical scoring. Multiple choice and short answer questions quickly became the conventional format. Such questions are scored only as right or wrong. While in principle there is nothing to prevent a standard test from using essays, extended reasoning and scales of partial

credit, reliable scoring of extended answers and essays requires careful training and monitoring of test evaluators and substantially more effort. Rushed and inept evaluation of extended answers can be at least as troublesome as restricting testing to multiple choice and short answer formats.

Standard tests have long been distinguished as having "speed" or "power" formats, meaning that they are strictly timed or that they are loosely timed or untimed (Rogers, 1995, p. 256, and Goslin, 1963, pp. 148-149). The distribution of scores is deliberately widened by strict timing. Many common school-based standard tests, including the Stanford, California and Iowa achievement tests, claim to measure knowledge and skill but are in fact "speed" tests. More recent distinctions are proposed between so-called "norm-referenced" and "criterion-referenced" tests (Rogers, 1995, pp. 653-666). Supposedly a "norm-referenced" test has a calibration relative to a population, while a "criterion-referenced" test has an absolute standard (for example, basic competence to drive a motor vehicle). However, for practical purposes nearly all school-based standard tests are "norm-referenced," because critical decisions about how to use the scores are made after score distributions have been measured. We used to call this "grading on the curve." In fact, wild attempts to produce "criterion-referenced" tests, without knowing how many people can actually pass them, generate some of the horror stories of testing.

Another recent and somewhat misleading distinction is so-called "high- stakes testing," meaning the use of test scores to make decisions that critically affect people. Supposedly this is a new practice. Actually it is quite old; parts of the Chinese civil service were closed to applicants who could not pass required examinations more than twenty centuries ago (Reischauer and Fairbank, 1958, p. 106). Beginning in the nineteenth century, standard tests were developed to place students in French schools according to ability. During World War I, U. S. Army recruits were assigned to combat or support missions on the basis of IQ scores.

According to current psychometric standards, it is improper to use a test for some purpose for which it was not "designed." Ninety years ago, however, intelligence tests were quickly appropriated to identify "morons," "imbeciles" and "idiots," who were then to be sexually restricted. Claims were advanced that experienced testers could readily identify "feeble-minded" people by observation (Gould, 1981, p. 165). We are not as far away from those days as some would like to think. Recent applicants who failed a new, uncalibrated teacher certification test were denounced as "idiots" by a prominent Massachusetts politician. (Note 7) Although some strong advocates of standard testing were once inspired by egalitarian views (such as Conant, 1940), standard tests have long been instruments for social manipulation and control. In an irony of the late twentieth century, tests like the former Scholastic Aptitude series, once praised as breaking the stranglehold of social elites on access to higher education, became barricades tending to isolate a new, test-conscious elite which, as we will see, largely tracks the social advantages of the old elite.

Aptitude, Achievement and Ability

School-based standard testing is largely a phenomenon of the twentieth century. An early product, the "intelligence scale" published by Alfred Binet and Théodore Simon in 1905, was intended to identify slow learners. By the 1920s, the testing movement had split into two camps which remain distinct today (see Goslin, 1963, pp. 24-33). The Binet-Simon scale and its offspring— such as the IQ test produced by Lewis M. Terman in 1916, the Army Alpha and Beta tests organized by Robert M. Yerkes during World War I, and the Scholastic Aptitude Test designed by Carl C. Brigham in 1925— all claimed to measure "aptitude." The essay exams of the College Entrance Examination Board, founded in 1900, the Stanford Achievement tests, first published in 1923, and Everett F. Lindquist's Iowa Every-Pupil tests, developed in the late 1920s and early 1930s, claimed instead to measure "achievement."

Tests of "aptitude" try to measure capacity for learning, while tests of "achievement" aim only to measure developed knowledge and skills. From their earliest days, standard aptitude tests have been clouded in controversy. It has never been clearly shown that "aptitude" can be measured separately from knowledge and

BEST COPY AVAILABLE

skills acquired through experience (see Ceci, 1991; also see Neisser, 1998, and Holloway, 1999, on changes over time). Standard achievement tests, while nominally free of these snares, share assumptions about language and cultural proficiency. Performance on almost any test is strongly influenced by language skills. Likewise, all tests rely to some degree on trained and culturally influenced associations and styles of thinking. Despite longstanding claims of distinct purposes, standard aptitude and standard achievement tests may have more similarities than differences.

Standard achievement test scores tend to correlate with standard aptitude test scores, as shown by Cole (1995) and others. To some observers, such as Hunt (1995), this simply shows that bright people learn well, and vice-versa. To others, it suggests that much of what is being tested might be called test- taking ability (see Hayman, 1997, and Culbertson, 1995). Most content of the widely used school-based standard tests can be viewed as collections of small puzzles to be solved rapidly by choosing options or writing brief statements. Such a pattern of tasks is rarely encountered by most adults in everyday life.

By design, the times allowed to complete standard tests are typically too short for a sizeable fraction of test-takers, putting great stress on rapid work and leaving little opportunity for reflection. For some strictly timed tests favoring men it has been shown that the same tests conducted without time limits favor women (see Kessel and Linn, 1996). Standard test designers may assign high scoring weights to test items written to be ambiguous, so that they will encourage wrong answers (see Owen and Doerr, 1999, pp. 70-72). Right answers are guided in part by trained or culturally acquired associations— intuitions about a test designer's unstated viewpoint. When ambiguous questions are removed, differences in scores between ethnic groups may be reduced. Test designers sometimes say that ambiguous questions "stretch the scale," differentiating the more skilled from the less skilled. Owen and Doerr (1999, pp. 45 ff.) suggest instead that they raise the scores of test-takers who have the favored patterns of associations and thinking.

The stressful properties of a typical standard test make test-taking into a sort of mental gymnastics, an ability that may well have its uses but does not necessarily predict performance in other situations (see Sacks, 1999, pp. 60- 61). We recognize many special skills, such as remembering complex patterns in card games, multiplying numbers in one's head, or solving crossword puzzles. People who do these things deftly may also perform well in other pursuits, or they may not.

Predictive Strengths

Standard tests are promoted on the basis of claims to predict future performance. Their predictive strengths are measured by how well they do this. Despite heavy use of standard tests in circumstances that may critically affect people's lives, there have been remarkably few evaluations of these tests by organizations independent of the test vendors. The underlying substance of predictive evaluations is sometimes shallow. For example, it may be claimed that a standard test required for acceptance to a school program helps to predict the likelihood of graduation, when a key criterion for graduation is the score on a similarly organized standard test.

For a standard test to be useful, it cannot merely predict performance to some degree. It must significantly improve the accuracy of prediction over readily obtained information. Unless it does so, the effort of testing is wasted. (Note 8) During the last forty years, predictive strengths of the SAT, ACT, GRE and similar aptitude tests have been independently evaluated. Scores from these tests improve predictions of first year grades by at most a few percent of the statistical variance over predictions based solely on previous grades, family income and other personal factors. (Notye 9) For later and broader measures of performance, the predictive strengths of these tests evaporate. Sometimes negative correlations have been found— lower performance associated with higher scores. (Note 10) In response to the low predictive strengths of standard aptitude tests, growing numbers of colleges have stopped requiring them as part of applications. (Note 11)

Predictive strengths of standard tests are falsely enhanced when they are used to "track" or group students in schools, providing extra opportunities to some while denying them to others. The favored students stand to gain not only skills and

knowledge but also self-esteem, which has been shown to correlate with higher test scores. (Note 12) Ability grouping based on standard tests is a form of "high-stakes testing" which has been practiced for at least 80 years in U. S. public schools. We can clearly distinguish between the selection procedures of public schools, which have a legal duty to treat every student fairly, and those of taxpaying private institutions, which may not. Of the public schools, we can surely ask, "Why not provide opportunity to everyone?"

Beyond the schoolhouse door, school-based standard tests show hardly any predictive strength for creativity, professional expertise, management ability or financial success. (Note 13) However, these tests stress either generalized test-taking abilities or subjects that are only occasionally relevant to adult life. Tests for competence in specific skills have been used successfully to predict whether workers can perform tasks that require those skills. For example, some temporary employment agencies now administer technical skills tests to new job-seekers before sending them out to interview with potential employers. This practice has increased employer satisfaction with job performance.

Errors of Testing

All measurements are subject to potential error. Compared with physical measurements, the errors in standard test scores are enormous. There are many sources of error. These include:

- Mechanical errors in transcribing short answers or multiple choice answers
- Consistency errors in scoring essays or extended answers
- Computer errors when calculating or reporting results
- Systematic errors from varying difficulty of different test versions
- Random errors arising from the physical or mental states of test-takers
- Bias errors: test designs that favor some groups of test-takers over others
- Content errors: test items that do not accurately cover the intended material

Vendors and promoters of standard tests do not often discuss errors of testing. When they do, they usually bury information in opaque language, tables and formulas found in "technical reports" that may be hard to obtain. Careful reading of such information often reveals defects in the error evaluation as well as large errors.

Test vendors typically present themselves as diligent in reducing or eliminating mechanical, consistency, computer and systematic errors. There are well developed methods for controlling these gross errors. However, such errors do occur. Advanced Systems, a company used by the Massachusetts Board of Education since 1986, was embarrassed by errors in score reporting in Kentucky and lost its Kentucky contract in 1997 (see Szechenyi, 1998, and "Problems," 1998). Gross errors seem to be more common with smaller and newer test vendors than with larger and longer established ones.

The most common error measurement for a standard test is its "reliability." By convention, this describes the range of scores which a test-taker would receive in taking repeated, comparable versions of a test (Rogers, 1995, pp. 61-62, 368-378 and 741-743). A narrow range means high reliability: a test-taker would be likely to receive about the same score on repeated tests. Because training effects occur when tests of a particular type are actually repeated, indirect methods must be used to estimate reliability, such as mathematical models. Details of these methods can be adjusted to change estimates of reliability.

When mechanical, consistency, computer and systematic errors have been well controlled, reliability mainly measures random errors arising from unpredictable, individual circumstances of test-takers. Such errors are often larger than is generally known. As cited by Owen and Doerr (1999, p. 72), the Educational Testing Service has estimated that, on average, individual differences of less than 70 points for its SAT Verbal scores and 80 points for its SAT Math scores are not significant. These margins increase for high scores. Massachusetts (1999a, p. 86, Table 14-4) has estimated there is only about a 56 percent chance that a fourth-grader who is advanced in English language arts, according to its standards, will receive an "advanced" rating on its MCAS fourth-grade English language arts test.

People who are unfamiliar with the large random errors of standard test scores often assume that the scores can be used reliably to rank-order test-takers according to ability. In fact, random errors of testing are so great that scores can be used at most to classify individuals in a few levels. Using only four levels to classify MCAS scores, Massachusetts (1999a, p. 86, Table 14-4) has estimated substantial likelihoods, ranging from 8 to 46 percent, that an MCAS test-taker will be misclassified.

Many types of bias errors have been discovered in standard tests. For example, if the format of a test is changed from multiple choice to essay, different groups of test-takers are favored. A study performed by the Educational Testing Service found that multiple choice questions on its advanced placement tests favored men and European-Americans, while essay questions favored women and African-Americans (cited by Sacks, 1999, p. 205). Grouping test-takers with high essay and low multiple choice scores and those with the reverse pattern, the study showed comparable college grades for the two groups but a sixty point difference in their average Educational Testing Service SAT scores, in favor of the group with high multiple choice scores (Sacks, 1999, p. 206).

People tested using a language in which they are not fluent are likely to do much worse than native speakers of the language. Tests that require reading, in the formats used for most standard testing, assume reading proficiency. Individuals with poor reading proficiency, whatever the cause, are at major disadvantage with respect to others who do not have such limitations. Bias caused by test timing and ambiguous questions has been previously mentioned. Most attempts to compensate for bias involve identifying substantially impaired individuals and providing them extra test time. There is little evidence that test bias is actually corrected with this approach (see Heubert and Hauser, 1999, p. 199).

Perhaps the greatest source of bias and content error in school-based standard testing is the conventional process of standard testing itself, as contrasted with rating actual performance. When an educational assessment should measure success at significant tasks, such as writing a research report or investigating a technical theory, it may be impossible to design a standard test with much accuracy or predictive strength. In the U. S., there has been a movement toward replacing standard testing with criterion-based "performance assessment" (see Appendix 6). A goal of this movement, also called "authentic assessment," is eventually to integrate educational testing with the ordinary processes of teaching and learning. There have been attempts to use performance assessment as part of state testing programs in Kentucky (1990-1997) and California (1991-1995), reviewed by McDonnell (1997, pp. 5-8 and 62-65).

School Accountability

The performance of public schools became an issue in the U. S. almost soon as support for public education began. In 1845 the Massachusetts Board of Education printed a voluntary written examination to measure eighth-grade achievement. Most students could not pass the test. Schoolmasters complained that knowledge tested did not match their curricula. After a few years the test was abandoned (see Appendix 2). In 1874 the Portland, Oregon, school superintendent distributed a curriculum for each of eight school grades. At the end of the school year, he administered written tests on the curriculum. Test scores were published in a newspaper. Based on test scores, less than half the students were promoted that year and the following year. An uprising by parents and teachers then led to dismissal of the superintendent and an end to the practices of publishing scores and denying promotion on the basis of a test score alone. (Note 14) Since those days similar initiatives and reactions have often occurred throughout the U. S..

The U. S. has sponsored a continuing expansion of public education for 350 years. Most people did not expect to graduate from eighth grade until late in the nineteenth century. High-school graduation became a normal expectation only in the 1930s. Today, we are still struggling with rising expectations that include college. At each stage of this growth, critics have condemned the lowering of educational standards and demanded accountability. However, each of these stages can also be seen as intrusion into a formerly elite province of education by large numbers of students who would previously have been excluded. For several years, levels of

performance go down as the system adapts to less prepared students. Over a longer period, curricula change, often abandoning cultural traditions for more practical approaches.

School accountability became a public demand during the first two decades of the twentieth century. (Note 15) Over the ten years from 1905 through 1914 the U. S. accepted the largest flow of immigrants in its history, averaging more than a million per year. Immigration, coupled with stronger school attendance laws, raised school enrollments and increased the fraction of students for whom English was not a native language. Declines in student achievement were noticed and became an object of public concern.

At first standard tests were used to document declining student achievement, but they did not provide a method to improve it. By 1920 many urban school systems had started to use the newly available intelligence tests to measure student aptitude; they grouped students in classes by IQ. (Note 16) Educators hoped to improve performance by providing instruction that was adjusted to student aptitudes. In 1925 a U. S. Bureau of Education survey (cited by Feuer et al., 1992, p. 122, footnote 91) showed that 90 percent of urban elementary schools and 65 percent of urban high schools had adopted this approach. As immigration declined and school attendance became more uniform, student achievement tended to stabilize, and public concern relaxed. Despite warnings from progressives such as John Dewey and Walter Lippmann about a "mechanical...civilization" run by "pseudo-aristocrats" (Dewey, 1922), IQ testing and the multiple choice test format had acquired prestige as techniques to improve public schools.

Strong U. S. demand for school accountability arose again in the 1970s through the 1990s. This time aptitude testing and finances played significant roles. Acceptance of Scholastic Aptitude Test scores as a measure of merit by highly selective colleges was regarded by many people as sanctioning a measure of merit for public schools. Average SAT scores for schools and communities began to circulate as tokens of prestige or shame. During the period from 1963 through 1982, the Educational Testing Service reported a continued decline in its national average SAT scores, followed by a slower recovery, as shown by the scores in Table 1.

Table 1
SAT National Average Scores

Test / Year	1963	1980	1995
SAT Verbal	478	424	431
SAT Math	502	466	482

Source of data: Ravitch, 1996.

These scores, scrutinized year after year, were used by the press, broadcast media and opportunist politicians to stir up a new sense of crisis. Once again, the public schools must be failing.

The charges were false. Accurate tracking of changes over time requires painstaking steps to assure that both the measurements and the groups being measured are comparable at each point. As shown by Crouse and Trusheim (1988, pp. 133-134) and by Feuer et al. (1992, pp. 185 ff.), the groups being measured by SAT scores changed drastically. Increases in scholarships and loans, affirmative action programs, and awareness of long-term financial rewards produced more applications to selective colleges. The number of colleges requiring SAT scores more than doubled. As a result, the number of students taking the SAT series for college applications grew from 560 thousand in 1960 to 1.4 million in 1980, an increase of 150 percent over a period in which public school enrollment grew only 16 percent. Students with lower high-school grades were taking these tests who would not have taken them in previous years. Spreads in scores increased significantly, reflecting more diversity in test-takers. Berliner (1993) shows that SAT scores of students with similar characteristics were actually increasing.

Other school-based standard tests do show changes over this period, but they are not parallel trends. Beginning in 1969, reading, writing, science and mathematics skills have been measured by the National Assessment of Educational Progress

(NAEP), a federal research program. Scores remained roughly steady through 1996, with typical average scores of 280-300 points at the high-school level and typical changes across this period of less than 10 points (see Appendix 1). NAEP reading comprehension scores would probably have fallen and then risen along with SAT Verbal scores if the SAT scores reflected real changes in education. Actually NAEP high-school reading scores were flat within a band of ± 1% over the entire 1971-1996 period. There may have been declines in science during the 1970s, but changes in NAEP procedures make them uncertain. During the past 20 years, at the high-school level there appear to have been modest gains in science and math and a slow but persistent decline in writing skills (while SAT Verbal scores were rising). Overall patterns of NAEP scores indicate little change in educational achievement. However, these research results do not generate flashy headlines or sound bites, and they are usually ignored.

The other major cause of concern during the last three decades of the twentieth century has been the increasing cost of public schools (see Appendix 1). Proportionately spending rose even faster from 1950 to 1970, but that was also a period of rapid growth in school enrollment, the "baby boom" generation, and a period of anxiety over the possibility of nuclear war. Annual, inflation-adjusted public school spending grew from about $1,570 per student in 1950 to $3,720 in 1970 and $7,140 projected for 2000 (all in 1998 dollars). Total public school spending climbed even during the 11 percent enrollment drop from 1970 to 1980. By demanding accountability the public has in part been seeking value in return for its reasonably generous support.

"School Reform"

Accountability is a political concept, not an educational one. The public figures who talk about it loudest today want "school reform," a familiar war cry in U. S. politics. (Note 17) The measures many current "school reformers" promote are:

- Frequent school-based standard testing with "high goals"
- Publication of scores for individual schools or districts
- Denial of school activities and diplomas to students with low scores
- Removal of principals and teachers in schools with low scores

Some politicians go further. (Note 18) In 1983, the Reagan administration embraced a system that would circulate test scores to colleges and employers, maintaining permanent national dossiers of people's test records. The Bush administration proposed legislation in 1991 including these concepts, but it was defeated in Congress. Just what such a program might do to people never seems to have been a concern for the "school reform" promoters.

In the name of "school reform," without any federal mandate, state legislatures and politically controlled state education boards have been increasing the use of standard tests in public schools and the punishments for low test scores. Typical of the state-run "school reform" programs are the following measures:

- Statewide standard achievement tests in several or all school grades
- Statewide standard tests for course credit, promotion and graduation
- "Curriculum frameworks," or required curricula, "aligned" to standard tests
- Access to advanced courses and special programs based on standard test scores
- Athletic team participation and student privileges based on standard test scores
- Special diplomas, honor programs and scholarships based on standard test scores
- Classification of school performance based on standard test scores
- Publication of test scores or classifications by school or by district
- Publicity about school testing requirements, changes and schedules
- Financial support for "test preparation" consultants and materials
- Financial incentives for administrators and teachers to achieve high test scores
- Removal of administrators and teachers in schools with low test scores
- State seizure or closure of schools with low test scores

Also associated with "school reform" are movements to support religious schools via "school choice" and financial "vouchers" and initiatives to create privately run "charter schools."

In 1980 eleven states required minimum scores on their standard tests to receive a high-school diploma. By 1997 seventeen states enforced such a requirement (National Center for Education Statistics, 1999, Table 155). During the years 2000-2005 several states, including Alaska, California, Delaware, Massachusetts, New York and Texas, are planning one or more of the following "school reform" initiatives:

- Add standard tests for course credit, promotion or graduation.
- Raise or begin enforcing required scores.
- Dismiss principals of low scoring schools.
- Place low scoring schools in receivership.

About two-thirds of the current states with high-school graduation tests are southern or southwestern states; they tend to have larger fractions of poverty and low-income households than the national averages. The students who are denied high-school diplomas typically come from the most disadvantaged households in those states.

Texas has a program often pointed to by "school reform" advocates as a model (see Appendix 4). The program is politically controlled by the governor and state legislature. It has changed several times since its inception in 1984. The key feature for the last ten years is a test system called TAAS, which includes high school graduation requirements. Under this system, there have been reports of weeks spent on test cramming and "TAAS rallies." School ratings are raised by "exempting" students. Schools are allowed to contract for "test preparation" consultants and materials, and some have spent tens of thousands of dollars. There have been reports of falsifying results. In April, 1999, the deputy superintendent of the Austin school district, which had shown dramatic score improvements, was indicted for tampering with government records. In Houston three teachers and a principal were dismissed for prompting students during test sessions ("TAAS scandal," 1999). Official Texas statistics claim reductions in school dropouts, but independent studies consistent with U. S. government data show persistent increases, with 42 percent of all students failing to receive a high school diploma as of 1998 ("Longitudinal Attrition Rates," 1999). Students identified by Texas as black or Hispanic are disproportionately affected. In some schools 100 percent of students with limited English proficiency drop out (IDRA, 1998). Illiteracy remains a major problem in Texas, and it appears to be worsening.

New York has recently released part of the initial results from its new high-school graduation tests. Based on currently required scores, they show that diplomas are likely to be denied at about twice the statewide rate to students in New York City who complete high school (see Appendix 3). The city has the largest concentrations of poverty in the state. In five years New York will increase the required scores by abolishing so-called "local" diplomas. The probable result will be an even more severe impact on students from poverty and low-income households.

State-run "school reform" has operated largely on the basis of beliefs, not evidence. There is little evidence that these programs actually work as intended. Feuer et al. (1992), show that claims for improved achievement, as measured by test scores, are often hollow. They are commonly a result of training students to take the standard tests (also see Sacks, 1999, pp. 117-151). When a new series of tests is substituted, scores typically return to levels, measured against national norms, that are similar to scores when the previous series of tests began.

If "school reform" has caused substantial improvement in student achievements, measurements performed by the National Assessment of Educational Progress (NAEP) ought to reveal it. This longstanding federal research program has taken care to provide broad coverage of educational content, to maintain consistency in its testing over time, and to avoid test formats with sources of bias such as hectic pacing and heavy dependence on reading proficiency in tests other than reading (see

Feuer et al., 1992, pp. 90-94). Test formats use multiple choice, short answer, extended answer and essay questions, with scales of partial credit. Since participating schools change, there is little opportunity or incentive for students to be taught the tests. From about 11,000 to 44,000 students participated in each of the test series given from 1982 through 1996.

Most of the geographically segmented data published for the NAEP are grouped by regions rather than by states. The Northeast region includes Connecticut, Delaware, District of Columbia, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island and Vermont. From 1982 through 1996 none had a major "school reform" program; only one of the twelve had a high-school graduation test (only New York; see National Center for Education Statistics, 1999, Table 155). The Southeast region includes Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia and West Virginia. From 1982 through 1996 all had major "school reform" programs and eleven of these twelve had high-school graduation tests (all except Kentucky; see National Center for Education Statistics, 1999, Table 155). Average NAEP scores reported for these two regions from 1982 through 1996 are shown in Table 2.

Table 2
NAEP Regional Average Scores, 1984 and 1996

Reading Scores	Northeast			Southeast		
	1984	1996	Change	1984	1996	Change
Grade 11	292	291	-1	285	279	-6
Grade 8	260	261	+1	256	252	-4
Grade 4	216	220	+4	204	206	+2

Writing Scores	Northeast			Southeast		
	1984	1996	Change	1984	1996	Change
Grade 11	291	290	-1	287	273	-14
Grade 8	273	264	-9	267	260	-7
Grade 4	212	213	+1	204	200	-4

Math Scores	Northeast			Southeast		
	1982	1996	Change	1982	1996	Change
Age 17	304	309	+5	292	303	+11
Age 13	277	275	-2	258	270	+12
Age 9	226	236	+10	210	227	+17

Science Scores	Northeast			Southeast		
	1982	1996	Change	1982	1996	Change
Age 17	284	296	+12	276	288	+12
Age 13	254	255	+1	239	251	+12
Age 9	222	234	+12	214	224	+10

Source of data: National Center for Education Statistics, 1997.

If a case can be made for improvement that may have been caused by "school reform" it is in math and science, where both regions had score improvements but those of "school reform" states were better. However, "school reform" states had worse changes in reading and writing scores. The Northeast, without major "school reform," improved scores an average of 2.8 points, while the Southeast, under major "school reform," improved scores an average of 3.4 points. With the random errors in scores estimated for NAEP, the difference in these results has no statistical significance (National Center for Education Statistics, 1997, pp. iii-vi). At the high-school level, the changes measured in "school reform" states were somewhat

better in math, the same in science, somewhat worse in reading and substantially worse in writing. Despite great hopes for "school reform," there is no general evidence of benefit.

"School reform" is strongly associated with high dropout rates and low rates of high-school graduation. Nationally about 32 percent of public school students aged 15 through 17 are enrolled below normal grade levels, a figure that climbed steadily during the years 1979 through 1992. (Note 19) Statistics on school dropout cannot be evaluated readily, since government reporting procedures have been changing, possibly to conceal unfavorable trends (see Appendix 4). Table 3 estimates normal high-school graduation rates for the class of 1996 as percentages of ninth-grade enrollments in the fall of 1992. (Note 20) It compares nine southern and southwestern states under major "school reform," requiring minimum scores on standard tests for graduation, with nine northeastern states that did not have major "school reform" programs:

Table 3
High-school graduation rates by state, 1996
(Percentage normal high-school graduation, class of 1996)

States under "school reform"		States without "school reform"	
Alabama	58%	Connecticut	74%
Florida	58%	Maine	72%
Georgia	55%	Massachusetts	76%
Louisiana	58%	New Hampshire	75%
Mississippi	57%	New Jersey	83%
North Carolina	62%	New York	62%
South Carolina	54%	Pennsylvania	76%
Texas	58%	Rhode Island	71%
Virginia	76%	Vermont	90%

Source of data: National Center for Education Statistics, 1996 and 1999.

Only one southern or southwestern state with major "school reform" had a normal graduation rate **above** two-thirds, while only one of the northeastern states had a rate **below** two-thirds. The worst northeastern state is New York, which has a longstanding Regents examination for high-school graduation but during the 1992-1996 period was also awarding "local" diplomas (see Appendix 3).

Reform Schools and Private Interests

By the early 1990s, with reform schools entrenched for ten years or more in several states, a perverse competition began, which might be called *Our Standards Are "Stiffer" Than Yours*:

- We make tests harder.
- We mandate more tests.
- We raise minimum scores.
- We enforce more punishments.

See Heubert and Hauser (1999, pp. 59-67) and Sacks (1999, pp. 98-99 and 114). As with most of "school reform," the process is political (see Appendix 4 and Appendix 5). Typically, it is known that test scores ramp up for a few years and then flatten out. Otherwise there is little organized review of whether the testing and punishment systems actually produce harm or benefit for anyone. Nevertheless, state governors and legislators vie for TV spots and news headlines with commitments to "raise standards." In states without major "school reform," politicians are prepared to exploit anxiety over somehow being left behind. (Note 21)

Many states are trying "school reforms" faster than their school systems can adapt. Seeking to change educational content and testing practices at the same time

worsens these problems. It has become common first to impose a test and then to "align" the curriculum, obviously putting the cart before the horse. Even states with a relatively stable curriculum and incremental changes in testing, such as North Carolina, have fallen prey to this disease (McDonnell, 1997, pp. v and 8-11). Some "school reformers" like the Pioneer Institute in Boston utilize the resulting chaos in political karate, aiming to promote "charter schools" which are actually private business ventures fed by tax revenues. James A. Peyser, Executive Director of Pioneer Institute, is currently Chairman of the Massachusetts Board of Education. Charles D. Baker, Jr., a member of the Pioneer Institute Board of Directors, is also a member of the Massachusetts Board of Education. Former and current directors of the Pioneer Institute founded Advantage Schools, Inc., of Boston, a for-profit business that has opened two Massachusetts charter schools and fourteen charter schools in other states.

These cross-interests and educational mistakes need to be made familiar to the public. They are usually ignored by the large newspapers and broadcast media unless a tragedy occurs. (Note 22) In contrast to the strong interest over test scores, our press, broadcast media and politicians show only sporadic interest in the education process. Effective innovations such as team teaching, "looping" and open classrooms are being neglected or forgotten (see Tyack and Cuban, 1995, pp. 86-107). Science and math have been emphasized, but long-term surveys of achievement suggest that progress in these areas has occurred partly at the expense of writing skills. Only computer technology gets much attention, but its limits are becoming apparent. While classroom computers are convenient for exploring the Internet and organizing assignments, they have otherwise taught students few skills.

By conventional standards of psychological testing, (Note 23) major test vendors have been earning revenue from highly questionable uses of their products. While technical manuals may advise that their achievement tests are not "validated" for uses such as school rating or promotion tests, they sell large volumes of these tests to jurisdictions using them for purposes other than individual counseling. For example, the Stanford Achievement Test series, published by Harcourt Brace Educational Measurement, is being used by the state of California to rate and compare school districts (see Appendix 5). The Iowa test series, from the Riverside Publishing division of Houghton Mifflin, is being used by the city of Chicago as promotion tests (see Roderick et al., 1999). When so used, these tests effectively set the curriculum and the standards of performance for public schools, without meaningful public input or control. Parents and taxpayers are poorly informed about test validation and about strong effects these tests have in setting educational standards.

Taking a cue from Horace Mann, who fought for school standards and then moved to Congress a century and a half ago (see Appendix 2), many modern politicians have sought to use "school reform" as a platform for advancement. The "school reform" movement has enough momentum that few state officeholders and candidates openly oppose it. Candidates for state offices often use "school reform" backgrounds to support their campaigns. In 1996 Governor Wilson of California attempted to mount a campaign for President; Governor Bush of Texas is doing the same this year. Wilson left office after the defeat of his 1998 plan (proposition 8) to create state-appointed "governing councils" for all California public schools, in charge of budgets. Taking a moderate approach, such as supporting smaller class sizes and improved facilities, has sometimes won out over "back to basics" appeals, as it did in the victory of Tom Vilsack over Jim Ross Lightfoot in the 1998 election for governor of Iowa.

The Social Context

School-based standard testing does not occur in a social vacuum. It has consequences, and the techniques it uses reflect interests and values. Insight and candor about these consequences, interests and values are rare today; they must often be inferred from behaviors. In previous times, the advocates of standard testing were less guarded about their intents.

It has become well known that early promoters of standard aptitude tests were profoundly racist and sexist. Goddard, Terman, Thorndike, Burt, Yerkes and Brigham all believed that these tests identified African-Americans, native

Americans, immigrants from southern and eastern Europe, or women as typically less able than white men whose ancestors came from northern and western Europe. (Note 24) Goddard, Terman and Brigham were advocates of the "eugenics" movement, (Note 25) favoring IQ tests followed by sexual restriction of the "feeble-minded." An echo of their attitudes can be heard in the enthusiasms for standard tests sometimes expressed in the U. S. today, reducing access by African-Americans and Hispanic-Americans to universities and professional schools. Few of the modern promoters of standard tests flaunt prejudices that were once openly displayed. Relative success on these tests by Jews and by the offspring of Asian immigrants has greatly tempered hubris over "Nordic superiority."

The myth of measuring innate talent has been exposed. Multifactor studies link high scores on aptitude tests with advantages in family income, language and cultural exposure, motivation, self-confidence and training (see, for example, Goslin, 1963, pp. 137-147, Duncan and Brooks-Gunn, 1997, pp. 132-189, and Brooks-Gunn et al., 1996). Key research on the inheritance of intelligence, once widely cited, has been probed and found to have been scientific fraud (Gould, 1981, pp. 234-239). After accounting for measurable influences of environment, studies of multiple factors do leave unexplained residues that might be called aptitudes, but they can only be inferred from comparisons across groups. There are no reliable techniques for measuring aptitudes in an individual which are independent of experience, nor has it been shown how many such aptitudes there might be.

Despite exposures of motive and mythology, use of standard testing continues to grow. A century after their origins, school-based standard testing and its scavenger, test preparation, have become industries sustained by powerful institutions and deeply felt personal interests. Their supporters are now often driven by secondary motives that result from widespread testing programs. At least two generations have been able to profit from test-taking success, entering professions and making connections during their college years that might otherwise have been closed to them. They know how to crack the tests; they make sure their children learn; and they can be angered to think that this useful wedge into income and influence might be removed.

Today's standard test enthusiasts range from right-wing extremists to hard-nosed business people to ambitious young professionals to church schools and home schoolers who are looking for validation of their work— in other words, some of our neighbors. Parents who want to keep young children out of the testing game are now beset with legal mandates in many states and with social pressure almost everywhere. Far too few people are asking whether the public schools are really broken and in need of this kind of a fix (see Berliner, 1993, and Berliner and Biddle, 1995).

Among the right-wing, there is a Libertarian perspective from which conventional standard tests are an intrinsic evil because they interfere with local control of schools. Also, it is worth noting that a number of the business enthusiasts for standard testing actually send their own children to private schools where such testing is not emphasized. Berliner and Biddle (1995) have extended such observations into an argument that some testing promoters have a different agenda: using the embarrassment of low test scores in public schools as a weapon to force governments toward corporate schools, which they will operate at a profit.

Much as in the 1920s, its first great decade, school-based standard testing is still sold as a key to discovering talent and measuring ability objectively. When possible its critics are ignored, or they are dismissed as extremists, dreamers or losers. Test development and scoring procedures are wrapped in mystification. "Validation" of tests is widely touted, but it usually means only that people who do well on one test do well on another. Public enlightenment has made progress, but it struggles upstream against a flow of laundry soap, liver pills and snake oil.

What have all the years of more than 100 million school-based standard tests a year (Note 26) brought us? The "one minute" people, perhaps, who judge anything that takes longer as not worth the bother. Try to make life into a rush of standard questions. The idiot-genius computer programmers, fast as lightning. The ones who saddled us with about $200 billion worth of "year 2000" problems, because they didn't think about a slightly bigger picture. The test prep industry, a scrounger that otherwise has no purpose. The product support staff who don't know what to do

when they run to the end of their cheat sheets. The cutback from education to test cramming in the states with standard punishment systems. Don't take chances; teach and learn the test.

Remedies

School-based standard testing has seen more than a century of development in the U. S. (see Appendix 7). No quick or simple remedy can cure the many problems it has caused. Any remedy will require resolute public action. The following priorities are essential:

- Stop using standard test scores to deny promotion or graduation.
- Stop using standard test scores to create financial incentives or penalties.

These are the key weapons of the state punishment systems. The significance and accuracy of standard test scores do not justify these measures. They are viruses that transform schools from education to test cramming. They are all harm and no benefit. If we do not stop the damage being wrecked by these mistaken "school reforms," no other remedies will matter much.

If the catastrophes from "school reform" can be curtailed, we can tackle the worst problems of current school-based standard testing:

- **Conflict of purpose.** We are trying to use the same tests to measure basic competence as to measure high levels of skills and knowledge.
- **Conflict of method.** We say that we want to measure meaningful skills and knowledge, but our test methods stress empty tasks and fast answers.

The root of these conflicts is the same: choosing speed and price over effectiveness. If we want accurate and meaningful results, we must reverse these priorities. Good tests will not be quick or cheap. A test to measure basic competence in a skill or subject must cover a broad range of what we believe basic competence should mean. A test to measure high levels of skills and knowledge must include open-ended tasks that can be performed with many different strategies. We will need to weigh costs and benefits carefully. Even when they do not corrupt education, meaningful tests will take time and resources that could have been spent otherwise.

The "authentic assessment" and "performance assessment" movements seek to combine educational assessments with the learning process. Classic models are the "course project" and the "term paper." While the intents of these movements are understandable, Kentucky and California experiences in the 1990s suggested that such techniques were not mature enough to provide reliable comparisons among schools or school districts, much less to create promotion or graduation tests (Sanders and Horn, 1995). Moreover, we have no school-based achievement tests at all that have been proven to predict meaningful accomplishments by students in the world beyond the schoolhouse door.

Schools probably test too much, yet at the same time they may fail to use tests when tests can help. A key example is poor and late diagnosis of reading disorders. A great fraction of adult activities require proficient reading; most school activities and standard tests do also. We know that some young students have much more difficulty reading than others, although they may otherwise have strong skills. Schools need to identify reading disorders as early as possible and help to remedy them before they become deeply ingrained.

Limited and conflict-ridden as it is, current standard testing shows systematic deficits for students from low-income and minority households. Better testing will give a better picture of how serious these problems are, but it will not cure them. We need plans and resources to address the problems which are already clearly understood:

- **Language.** We should teach standard spoken English as a second language to students from households where it not spoken. We should not disparage dialects or other languages, but we must equip students early with this

essential skill.

- **Motivation.** Other than language, the key barrier for students from low-income and minority households is weak motivation. Home and school partnerships have shown how this problem can be overcome. We must create and strengthen them.

We do not understand all the problems. We do not know how to solve all the problems that we do understand. But we know enough to begin. If not now, then when?

Validity and Relevance

School-based aptitude testing is known to have low predictive strength. Studies have shown that it heavily reflects the income and education levels of students' households and that most of what it can predict is associated with social advantages and disadvantages. If tax-supported or tax-exempt schools use scores on intelligence or other aptitude tests to deny opportunities to some students while providing them to others, they violate the public trust.

For school-based achievement testing, we have few studies of predictive strength (as one example, see Allen, 1996, section IV-B, pp. 118-120). In most circumstances, we simply do not know whether these tests measure anything apart from social privilege that is useful outside a school setting. After adjustment for social factors, can their scores accurately predict future success in occupations, creative achievements, earning levels, family stability, civic responsibility or any of the other outcomes we mean to encourage with public education? Are there alternative assessments that can accomplish these goals? Given the heavy engagement in "school reforms" and the energy spent on their testing programs, it is amazing to see how little attention these matters receive (see related observations by Broadfoot, 1996, pp. 14-15). Academic and foundation-supported scholars specializing in psychometrics have the greatest opportunities to answer these questions, but they have largely ignored them.

Journalists, broadcasters, bureaucrats, politicians, educators and their critics— like most of the public— usually assume that a mathematics test, for example, actually measures some genuinely useful knowledge and skill. Who has shown this to be true, and for which tests? Is there actually a strong and consistent relation, for example, between top scores on a particular high school math achievement test and a successful career as a civil engineer? If there were not, then what does that test measure? Is there a strong and consistent relation between acceptable scores on a social studies test and adult voting participation? If there were not, then how is such a test of use?

Unfortunately, it is far from proven that any method of assessment can escape the biases, the other errors, and the low or unknown predictive strengths outside the schools which plague the current tests. We should take this not as a signal of defeat but as an invitation to humility. The complexities of human behavior are immense, and our current approaches measure them poorly. Rather than try to stretch each student onto a Procrustean bed of so-called "achievement," taking pride in lengthening the beam a bit every few years, we need to promote core competence and recognize the diversity of other skills. If standard tests were to have any useful role, it would most likely be as an aid to help insure that students can exercise skills which have been clearly proven essential for ordinary occupations. Even such a limited objective as this requires both education and test validation well beyond current educational and psychometric practices.

As we question the validity of testing, we may also question the relevance of the education supposedly being tested. Are we using the irreplaceable years of youth to convey significant skills and knowledge, or are we cultivating fetishes and harping on hide-bound answers to yesterday's questions? Somehow, despite decades of claims that our schools are inferior, we in the U. S. have achieved a stronger economy than most other industrial countries. Yet we also have more crime than most of these countries. Is our education responsible for these situations? We have many such issues to address. They present truly difficult questions. None of them will be found on school-based standard tests.

Notes

Comments and suggestions from several reviewers are gratefully acknowledged. Mistakes or omissions remain, of course, the fault of the author.

1. For a viewpoint characteristic of the era, see Rickover, 1959.
2. Pope Pius XI, as spoken in "...the defenders of order against the spread of Godless communism," Christmas Allocution, The Holy See, Rome, 1936. "Godless communism" became a popular phrase among cold-war patriots of the era.
3. Lemann, 1995, recounts the history of draft-deferment testing.
4. Commonly called "standardized testing." The underlying purpose of such tests is to set a standard that is calibrated for a population.
5. Reischauer and Fairbank, 1958, pp. 106-107, describe Chinese origins in the Western (Earlier) Han Dynasty, c. 120 BCE.
6. Schultz, 1973, reviews the industrial model for public schooling.
7. Massachusetts House Speaker Thomas Finneran. See Lehigh, 1998.
8. Goslin, 1963, p. 82 (footnote 2), indicates that the relatively low predictive strengths of aptitude tests for college grades were well known by around 1960.
9. Crouse and Trusheim, 1988, pp. 124-127, review predictive strength for the SAT vs. family incomes and high school grades. Nairn and Associates, 1980, show that SAT scores tend to act as proxies for family income. Tyack, 1974, pp. 214-215, cites an equivalent claim for IQ scores made by the Chicago Federation of Labor in 1924.
10. Sacks, 1999, p. 183 (note 23), cites a negative correlation between GRE aptitude test scores and publishing records for academic historians.
11. Owen and Doerr, 1999, Appendix C, list 284 U. S. colleges and universities where SAT and ACT scores are optional for admission into bachelor's programs.
12. Merton, 1957, pp. 421-436, calls such a phenomenon a "self-fulfilling prophecy."
13. Sacks, 1999, pp. 182-185, cites and summarizes several relevant studies.
14. Tyack, 1974, pp. 35-36 and 47-48, recounts the two examples cited of nineteenth-century school testing.
15. Tyack, 1974, pp. 126-147, shows how demands for accountability were used to cement control of public schools by business leaders and school supervisors.
16. Tyack, 1974, pp. 194 and 206-216, recounts the rapid spread of standard testing in the 1920s.
17. Tyack, 1974, pp. 41-46, recounts the first major U. S. school reform, the system of graded classrooms, inspired by Prussian schools and introduced to the U. S. in the 1840s and 1850s. Tyack and Cuban, 1995, explore the history of twentieth-century school reform movements in the U. S.
18. A Nation at Risk, published by the National Commission on Excellence in Education, U. S. Department of Education, in April, 1983, is cited as inspiring many of these initiatives.
19. See Appendix 1. Precedents from the past are worse. In 1922, New York City reported that nearly half of all students were "above normal age for their school grade," as cited by Feuer, et al., 1992, p. 118.
20. Data from National Center for Education Statistics, 1996, and National Center for Education Statistics, 1999. See 1995 Table 41 for ninth-grade enrollments and 1998 Table 102 for high school graduates. No attempt is made to adjust for immigration, emigration, mortality or population movement between states.
21. An egregious example of these effects can be seen in California from 1994 through 1997, during the Wilson administration. See Appendix 5.
22. Albert L. Powers, "Questionable reform," Carlisle Mosquito, Carlisle, MA, October 29, 1999. Paul Dunphy, "Charter schools fail promises," Daily Hampshire Gazette, Amherst, MA, February 7, 2000. Beth Daley and Doreen I. Vigue, "Firm pulls out of school where boy died," Boston Globe, February 10, 2000.
23. Standards 6.12, 8.7 and 8.12 in Committee to Develop Standards for

Educational and Psychological Testing, 1985, pp. 43 and 53-54. These standards, jointly developed by the American Psychological Association, American Educational Research Association and National Council on Measurement in Education, were also updated in 1999.

24. Brigham, 1923, pp. 87 ff., says "...the foreign born are intellectually inferior," then analyzes inferiority by races and origins.
25. For the proposition that "no feeble-minded person should ever be allowed to marry or to become a parent," Goddard, 1914, p. 561. On "curtailing the reproduction of feeble-mindedness," Terman, 1916, p. 7. On "prevention of the continued propagation of defective strains," Brigham, 1923, p. 210. All three men modified their views in later years.
26. Since at least 1961. See Goslin, 1963, pp. 53-54.

References

Allen, W. B., Project Director (1996). *A New Framework for Public Education in Michigan.* East Lansing, MI: James Madison College, Michigan State University.

Associated Press (1999, June 3). Blacks nearly four times more likely to be exempt from TAAS than whites. *Capitol Times,* Austin, TX.

Berliner, D. C. (1993). Educational reform in an era of disinformation. *Educational Policy Analysis Archives 1*(2), available at http://epaa.asu.edu/epaa/v1n2.html.

Berliner, D. C., & Biddle, B. J. (1995). *The Manufactured Crisis: Myths, Fraud, and the Attack on America's Public Schools.* Reading, MA: Addison- Wesley.

Brigham, C. C. (1923). *A Study of American Intelligence.* Princeton, NJ: Princeton University Press.

Broadfoot, P. M. (1996). *Education, Assessment and Society: A Sociological Analysis.* Philadelphia, PA: Open University Press.

Brooks-Gunn, J., et al. (1996). Ethnic differences in children's intelligence test scores. *Child Development 67*(2), 396-408.

California Department of Education (2000). *Academic Performance Index School Rankings, 1999.* Sacramento, CA: Department of Education, Delaine Eastin, State Superintendent.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology 27*(5), 703-722.

Census Bureau (1992). *Census of 1990.* Washington, DC: U. S. Department of Commerce.

Cole, P. G. (1995). The bell curve: Should intelligence be used as the pivotal explanatory concept of student achievement? *Issues In Educational Research 5*(1), 11-22.

Committee to Develop Standards for Educational and Psychological Testing, Melvin R. Novick, Chair (1985). *Standards for Educational and Psychological Testing.* Washington, DC: American Psychological Association.

Conant, J. B. (1940, May). Education for a classless society. *Atlantic Monthly 165*(5), 593-602.

Cremin, L. A. (1962). *The Transformation of the School: Progressivism in American*

Education, 1876-1957. New York: Alfred A. Knopf.

Crouse, J., & Trusheim, D. (1988). *The Case Against the SAT.* Chicago: University of Chicago Press.

Culbertson, J. (1995). Race, intelligence and ideology. *Educational Policy Analysis Archives 3*(2), available at http://epaa.asu.edu/epaa/v3n2.html.

Daley, B., & Zernike, K. (2000, January 26). State may change MCAS contractor. *Boston Globe.*

Dewey, J. (1922, December 13). Individuality, equality and superiority. *The New Republic 33*(419), pp. 61-63.

Duncan, G. J., & Brooks-Gunn, J., Eds. (1997). *Consequences of Growing Up Poor.* New York: Russell Sage Foundation.

Feuer M. L., et al., Eds. (1992). *Testing in American Schools: Asking the Right Questions* (Publication OTA-SET-519). Washington, DC: U. S. Congress, Office of Technology Assessment.

Goddard, H. H. (1914). *Feeble-mindedness; Its Causes and Consequences.* New York: Macmillan.

Goslin, D. A. (1963). *The Search for Ability.* New York: Russell Sage Foundation.

Gould, S. J. (1981). *The Mismeasure of Man.* New York: W. W. Norton and Co.

Haney, W. M. (1999). *Supplementary Report on the Texas Assessment of Academic Skills Exit Test (TAAS-X).* Boston: Center for the Study of Testing, Evaluation and Educational Policy, Boston College School of Education.

Hayman, R. L., Jr. (1997). *The Smart Culture: Society, Intelligence, and Law.* New York: New York University Press.

Heubert, J. P., & Hauser, R. M., Eds. (1999). *High Stakes Testing for Tracking, Promotion and Graduation.* Washington, DC: National Academy Press.

Holloway M. (1995, January). Flynn's effect. *Scientific American 280*(1), 37-38.

Hunt, E. (1995). The role of intelligence in modern society. *American Scientist 83*(4), 356-369.

IDRA Newsletter (1998, January). Intercultural Development Research Association, San Antonio, TX.

Kessel, C., & Linn, M. C. (1996). Grades or scores: Predicting future college mathematics performance. *Educational Measurement: Issues and Practice 15*(4), 10-14.

Lehigh, S. (1998, June 28). For teachers, criticisms from many quarters. *Boston Globe.*

Lemann, N. (1995, September). The great sorting. *Atlantic Monthly 276*(3), 84-100.

Longitudinal Attrition Rates in Texas Public High Schools, 1985-1986 to 1997-1998 (1999). Intercultural Development Research Association, San Antonio, TX.

Massachusetts Department of Education (1999a). *Massachusetts Comprehensive Assessment System 1998 Technical Report.* Malden, MA: Department of Education, David P. Driscoll, Commissioner.

Massachusetts Department of Education (1999b). *Massachusetts Comprehensive Assessment System, Report of 1999 State Results.* Malden, MA: Department of Education, David P. Driscoll, Commissioner.

McDonnell, L. M. (1997). *The Politics of State Testing: Implementing New Student Assessments* (Publication CSE-424). Los Angeles: National Center for Research on Evaluation, Standards and Student Testing, University of California.

McDonnell, L. M., & Weatherford, M. S. (1999). *State Standards-Setting and Public Deliberation: The Case of California* (Publication CSE-506). Los Angeles: National Center for Research on Evaluation, Standards and Student Testing, University of California.

Merton, R. K. (1957). *Social Theory and Social Structures.* Glencoe, IL: Free Press.

Nairn, A., & Associates (1980). *The Reign of the ETS: The Corporation that Makes Up Minds.* Washington, DC: Center for the Study of Responsive Law.

National Center for Education Statistics (1996). *Digest of Education Statistics, 1995.* Washington, DC: U. S. Department of Education.

National Center for Education Statistics (1997). *NAEP 1996 Trends in Academic Progress* (Publication NCES 97-985). Washington, DC: U. S. Department of Education.

National Center for Education Statistics (1999). *Digest of Education Statistics, 1998.* Washington, DC: U. S. Department of Education.

Neisser, U., Ed. (1998). *The Rising Curve: Long-Term Gains in IQ and Related Measures.* Washington, DC: American Psychological Association.

New York State Education Department (1998). *New York State School Report Card for the School Year 1996-1997.* Albany, NY: Education Department, Richard P. Mills, Commissioner.

New York State Education Department (1999). *New York State School Report Card for the School Year 1997-1998.* Albany, NY: Education Department, Richard P. Mills, Commissioner.

Owen, D., & Doerr, M. (1999). *None of the Above* (Revised ed.). Lanham, MD: Rowman and Littlefield Publishers.

Problems with KIRIS test erode public's support for reforms (1998, February 2). *Lexington Herald-Leader,* Lexington, KY.

Ravitch, D. (1996, August 28). Defining literacy downward. *New York Times.*

Regional Profile, Juarez and Chihuahua (1999). Texas Centers for Border Educational Development, El Paso, TX.

Reischauer, E. O., & Fairbank, J. K. (1958). *East Asia: The Great Tradition.* Boston: Houghton Mifflin.

Rickover, H. G. (1959). *Education and Freedom.* New York: E. P. Dutton and Co.

Roderick, M., et al. (1999). *Rejoinder to Ending Social Promotion: Results from the First Two Years.* Chicago: Consortium on Chicago School Research, Designs for Change.

Rogers, T. B. (1995). *The Psychological Testing Enterprise.* Pacific Grove, CA:

Brooks/Cole Publishing Co.

Sacks, P. (1999). *Standardized Minds*. Cambridge, MA: Perseus Books.

Sanders, W. L., & Horn, S. P. (1995). Educational assessment reassessed: The usefulness of standardized and alternative measures of student achievement as indicators for the assessment of educational outcomes. *Education Policy Analysis Archives 3*(6), available at http://epaa.asu.edu/epaa/v3n6.html.

Schultz, S. K. (1973). *The Culture Factory: Boston Public Schools, 1789-1860*. New York: Oxford University Press.

Szechenyi, C. (1998, March 8). Failing grade? Firm with state's assessment contract has troubled past. *Middlesex News*, Framingham, MA.

TAAS scandal widens (1999, April 9). *Lone Star Report*, Austin, TX.

Terman, L. M. (1916). *The Measurement of Intelligence*. Boston: Houghton Mifflin.

Texas Education Agency (1998). *1998 Comprehensive Biennial Report on Texas Public Schools*. Austin, TX: Education Agency, Jim Nelson, Commissioner.

Tyack, D. B. (1974). *The One Best System*. Cambridge, MA: Harvard University Press.

Tyack, D. B. & Cuban, L. (1995). *Tinkering toward Utopia: A Century of Public School Reform*. Cambridge, MA: Harvard University Press.

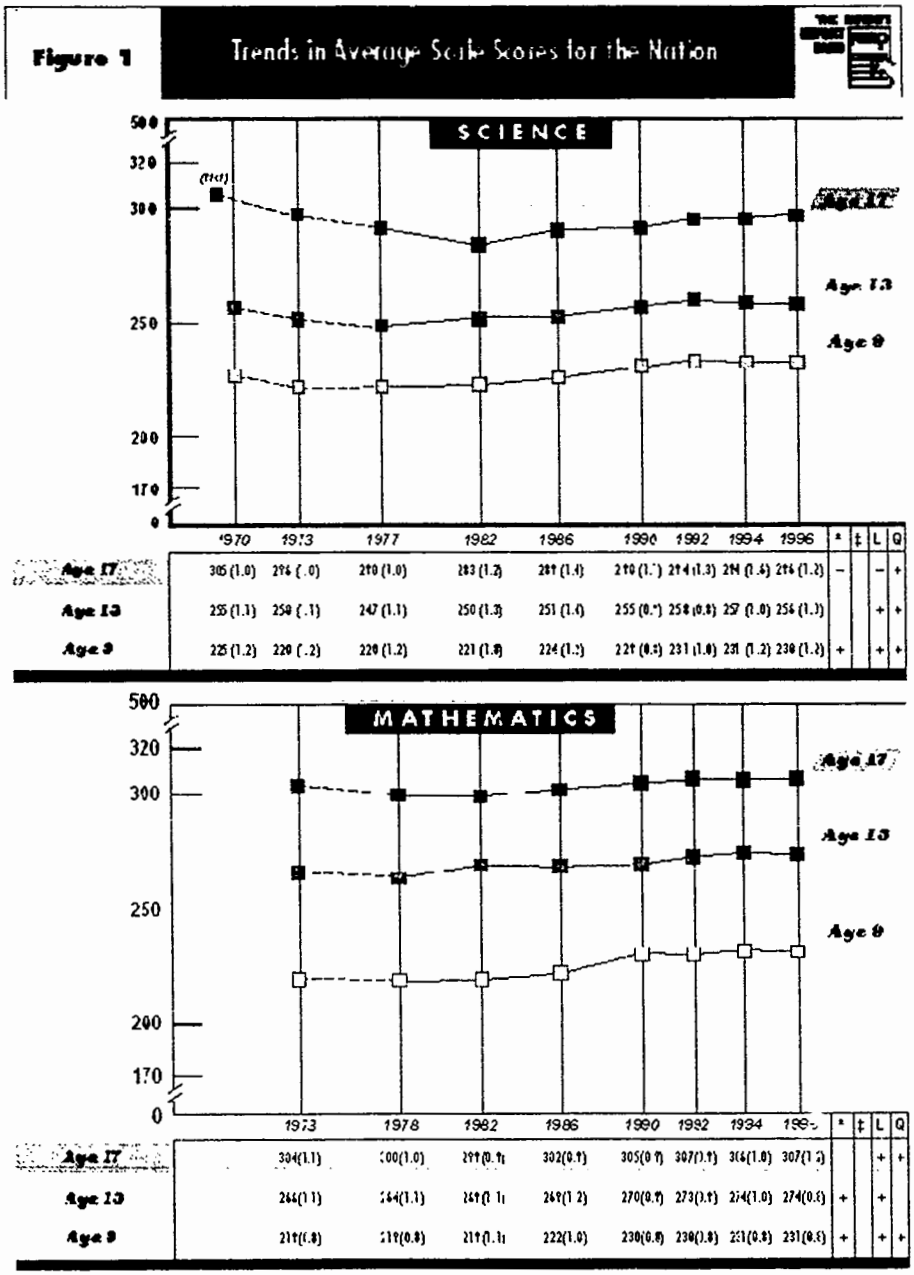
About the Author

Craig Bolon
Planwright Systems Corporation, Inc.
Email: cbolon@planwright.com

Craig Bolon is President of Planwright Systems Corp., a software development firm located in Brookline, Massachusetts, USA. After several years in high energy physics research and then in biomedical instrument development at M.I.T., he has been an industrial software developer for the past twenty years. He is author of the textbook *Mastering C* (Sybex, 1986) and of several technical publications. He is an elected Town Meeting Member and has served as member and Chair of the Finance Committee in Brookline, Massachusetts.

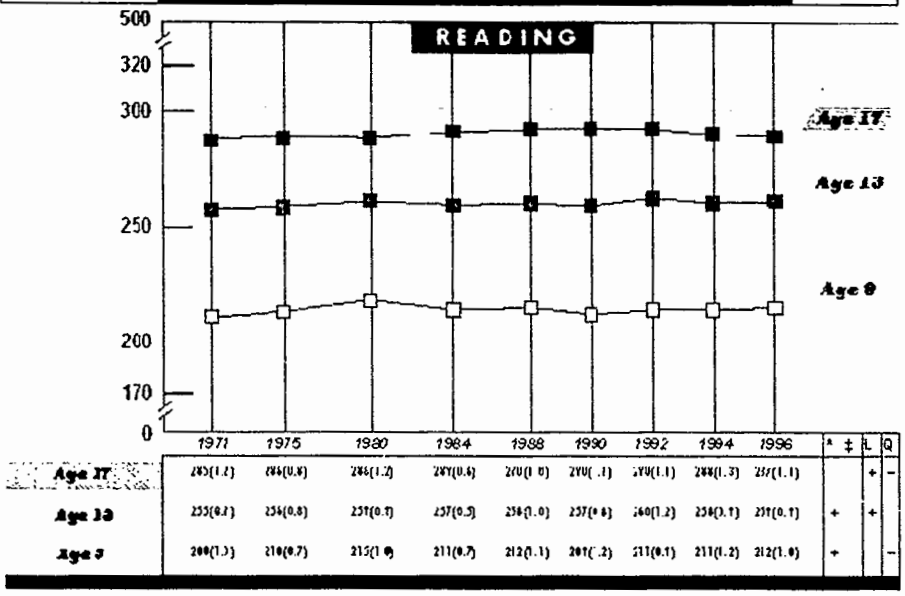
Appendix 1
Information: U. S. Public Education

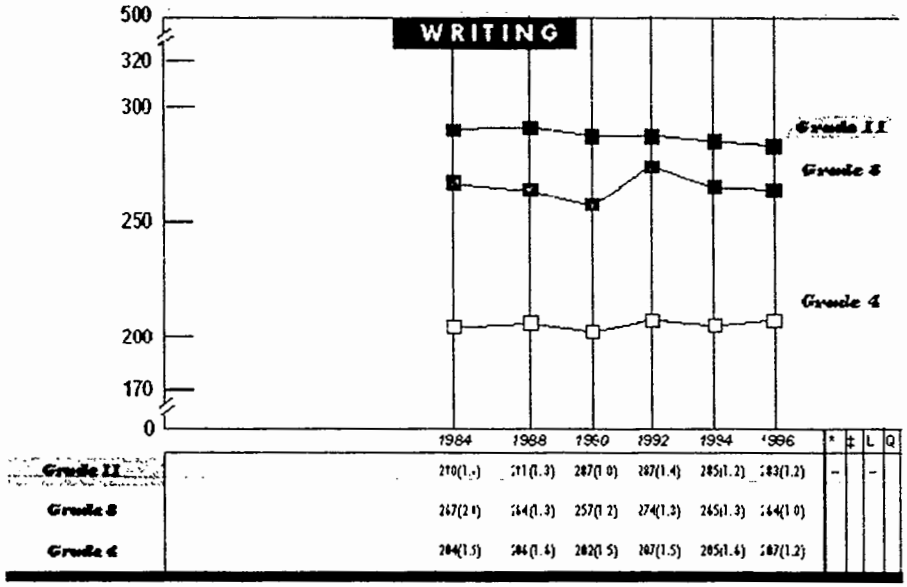
Figure 1 (on two pages, U. S. Dept. of Education, 1997) shows NAEP national average scores from program inception through 1996.



BEST COPY AVAILABLE

Figure 1 (continued) Trends in Average Scale Scores for the Nation





Standard errors of the estimated scale scores appear in parentheses. [– – –] Extrapolated from previous NAEP analyses.
* Indicates that the average scale score in 1996 is significantly larger (+) or smaller (–) than that in the first assessment year.
‡ Indicates that the average scale score in 1996 is significantly larger (+) or smaller (–) than that in 1994.
L Indicates that the positive (+) or negative (–) linear trend is significant.
Q Indicates that the positive (+) or negative (–) quadratic trend is significant.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Long-Term Trend Assessment.

The next chart, Figure 2, with data in Table 4, shows estimated U. S. public school enrollment and spending for the years 1850-2000. Enrollment is for elementary and secondary schools, including kindergarten, in millions. Spending includes local, state and federal outlays, in US$ billions, adjusted to 1998 dollar equivalence by the annualized Consumer Price Index. The last chart, Figure 3, shows U. S. public school enrollment aged 15-17 retained below modal grade, for the years 1971 through 1998. The increase in enrollment below modal grade is caused by increases in retention rates at all grades as well as by later ages of first school enrollment (Heubert and Hauser, 1999, pp. 136-158).

BEST COPY AVAILABLE

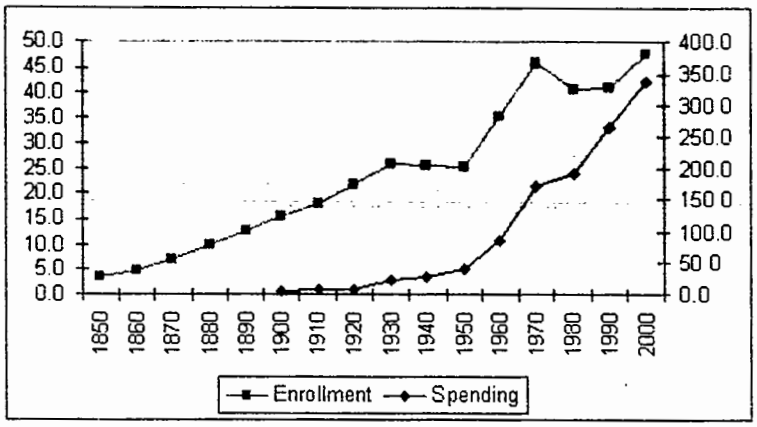


Figure 2. U. S. public school enrollment and spending.

Table 4
U. S. Public School Enrollment and Spending, for Figure 2

Year	Enrollment 1,000,000s	Spending SB (1998)	Spending per student
1850	3.4		
1860	4.8		
1870	6.9		
1880	9.9		
1890	12.7		
1900	15.5	4.2	270
1910	17.8	7.4	420
1920	21.6	8.4	390
1930	25.7	22.6	880
1940	25.4	27.3	1070
1950	25.1	39.5	1570
1960	35.2	86.0	2440
1970	45.9	170.8	3720
1980	40.9	189.9	4650
1990	41.2	265.4	6440
2000	47.4	338.6	7140

Sources: U. S. Department of Education, *Digest of Education Statistics*, 1998 (spending not available in this series before 1900); U. S. Census Bureau, *Census of 1850* and *Census of 1860;* U. S. Bureau of Labor Statistics, *Consumer Price Index: All Urban Consumers* (annual averages, estimated before 1913).

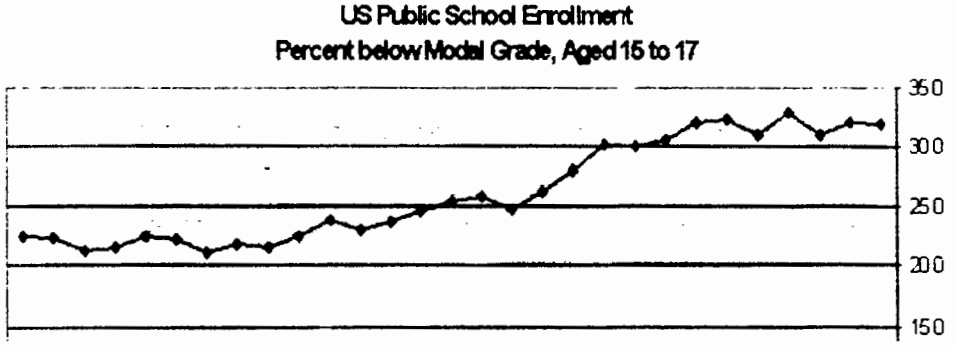


Figure 3. U.S. public school enrollment below modal grade.

Source: "The population 6 to 17 years old enrolled below modal grade: 1971 to 1998," *Current Population Survey Report – School Enrollment – Social and Economic Characteristics of Students*, U. S. Bureau of the Census, Washington, DC, Supplementary Table A-3, October, 1999.

Appendix 2
Information: Massachusetts

The Mather school, the first free public school in the U. S., was founded in Dorchester, Massachusetts, in 1639. In 1647 the Massachusetts General Court enacted a law requiring every town of 100 families or more to provide free public education through the eighth grade, but attendance was not required. In 1821 Boston opened English High School, the first free public high school in the U. S.. It taught English, history, logic, mathematics and science but did not offer the traditional Latin curriculum. An 1827 Massachusetts law required every town with 500 or more families to support a free public high school, and an 1852 law required school attendance to the age of 14, the first such laws in the U. S.. Massachusetts took over 30 years to reach compliance with each.

Massachusetts created a state Board of Education in 1837 to set standards for public schools, then in disarray. Horace Mann, a state senator from Boston and former state representative from Dedham, became the first Secretary to the Board. In 1839, at Mann's urging, Massachusetts created its first state- supported teacher's college, located in Lexington (now in Framingham). In 1845, following disputes over the quality of instruction, the Board of Education issued a voluntary written examination for public school eighth-graders, consisting of 30 short-answer questions. In its first year, the average score was less than one-third correct answers. Scores were soon used to compare schools in the press. Schoolmasters complained that knowledge tested did not correspond to their curricula. After Mann entered Congress in 1848 the examination was discontinued. During the following 138 years the Board of Education did not require testing of students.

In 1986 the Board of Education began statewide student testing called the Massachusetts Educational Assessment Program (MEAP). Among its purposes was to provide comparisons between student achievements in the state and student achievements being measured since 1969 through NAEP, the National Assessment of Educational Progress. Fourth-grade and eighth-grade tests of reading, mathematics and science were given every two years from 1986 through 1996. These tests were designed and administered by Advanced Systems in Measurement and Evaluation, Inc., of Dover, NH. Questions were in multiple choice, short answer and extended answer formats. Only aggregate scores for the state were publicly reported. Scores for individual schools were not disclosed. While Massachusetts average scores were above national averages, from 26 to 32 percent of the 1992- 1996 scores were "below basic," the lowest of four classification levels.

The Massachusetts Education Reform Act of 1993 required revised educational standards and procedures. In January, 1998, the Board of Education began using a new Massachusetts Teacher Test as a part of teacher certification. A communication and literacy skills test and a subject test in one of 41 areas must be passed. These tests, recently renamed the Massachusetts Educator Certification Tests, are being prepared and administered by National Evaluation Systems, Inc., of Amherst, MA, designer of the California Basic Educational Skills tests and the Texas Academic Skills Program tests. They are strictly timed and include multiple choice reading comprehension questions, short answer vocabulary and grammar questions, and a written composition. Testing was

initiated without a tryout period for the test and with relatively little advance notice about test content or consequences. In the first group of candidates, less than half passed both parts of the test. As one result, only white candidates were certified to teach in Massachusetts.

In 1995, the Board of Education released "curriculum frameworks," or required curricula, for mathematics and for science and technology. It later issued frameworks for English language arts and for history and social science. In the spring of 1998, after a tryout period in 1997, the Board began a new student testing program in the fourth, eighth and tenth grades called the Massachusetts Comprehensive Assessment System (MCAS). It includes tests in English language arts, mathematics, science and technology, and history and social science. They are loosely timed and include questions in multiple choice, short answer and extended answer formats. Through 1999, the test for history and social science has been administered only to eighth-grade students. Total testing time is about ten to fifteen hours, depending on the year and number of tests, with about half typically spent on English language arts. Scores are reported in a 200-280 point range; they are classified in four levels, equally spaced in the score range, called "advanced," "proficient," "needs improvement" and "failing." Parents are not permitted to exempt their children from testing. There are alternative procedures, such as small group settings, for special needs students and for students for whom English is not a native language.

Beginning in 1999, aggregate scores for each school in the state were publicly reported. Individual scores are disclosed to schools and parents. Schools also receive an analysis of results for each test item. Both 1998 and 1999 test items have been released to the public. The 1999 tests were offered in Spanish as well as English. Statewide, the results for 1998 and 1999 were similar; combined results from these two years are shown in Table 5.

Table 5
Massachusetts MCAS Average Scores, 1998-1999.

MCAS English Language Arts, statewide, 1998-1999 combined

School Grade	Average Score	Percent Advanced	Percent Proficient	Percent Needs Improvement	Percent Failing
10	229	4	32	34	30
8	237	3	52	31	14
4	230	0	20	66	14

MCAS Mathematics, statewide, 1998-1999 combined

School Grade	Average Score	Percent Advanced	Percent Proficient	Percent Needs Improvement	Percent Failing
10	222	8	16	23	53
8	226	7	23	29	41
4	234	12	23	44	21

MCAS Science and Technology, statewide, 1998-1999 combined

School Grade	Average Score	Percent Advanced	Percent Proficient	Percent Needs Improvement	Percent Failing
10	225	2	21	40	37
8	224	4	24	29	43
4	239	8	44	38	10

MCAS History and Social Science, statewide, 1998-1999 combined

School Grade	Average Score	Percent Advanced	Percent Proficient	Percent Needs Improvement	Percent Failing
10					
8	221	1	10	40	49
4					

Source of data: Massachusetts Department of Education, 1999b.

The Board of Education has released a technical analysis of the 1998 MCAS which includes estimates that its classification levels are consistent (Massachusetts Department of Education, 1999a). These are phrased in terms of the probability that a student who might receive a particular classification level after many repeated tests of some type would be classified at the same level by any one of those tests. Estimated probabilities range from 56 to 92 percent; they are highest for the "failing" level, averaging 85 percent, and lowest for the "advanced" level, averaging 70 percent. This technical analysis considers "validity" only in the narrow sense of comparison with other test results. Strong correlations, from .6 to .8, were found with components of the Stanford Achievement Test series. Significant MCAS score differences between male and female students and large score differences between students of different ethnic backgrounds were found, a pattern that is commonly duplicated by aptitude tests. Neither the methodology for computing the reported scaled scores nor the basis for classifying scores into passing and failing levels has been disclosed to the public.

In 2003, passing scores on tenth-grade tests will be required for a high-school diploma. The Board of Education has also announced plans to remove principals of schools which receive low scores and do not improve. The Board has not reported the fraction of students failing at least one of the tenth-grade tests, but statewide it is obviously more than half. By 2003, Massachusetts may be denying a diploma to a majority of students who complete high school, based on their failure to achieve passing scores on its standard tests.

Like the MEAP tests, the 1998 and 1999 MCAS tests were designed and administered by Advanced Systems in Measurement and Evaluation, Inc., of Dover, NH. Advanced Systems won a 1995 contract estimated at $25 million over competitors Riverside Publishing, publisher of the Iowa Tests of Educational Development, and Harcourt Brace Educational Measurement, publisher of the Stanford Achievement Tests. Advanced Systems has been a target of state investigations for its work in Maine and New Hampshire. In 1997, it lost a contract in Kentucky after being accused of gross errors in test scoring ("Problems," 1998). Scoring errors by the firm have also been reported in Maine. Tests that use extended answer questions, as those in Massachusetts do, must be scored by individual test evaluators. There have been reports of hasty scoring by Advanced Systems test evaluators working under time pressures and of computer programming errors by the company (Szechenyi, 1998).

In the summer of 1999, the Board of Education opened competitive bidding for the MCAS program of 2000-2004. Bids were received from the same companies as in 1995. In January, 2000, Commissioner of Education David P. Driscoll announced that Harcourt Brace Educational Measurement had received preliminary selection, with final negotiations in progress (Daley and Zernike, 2000). Problems with this change in vendors can be expected. A new vendor lacks the time to repeat the review and tryout process of the first MCAS series before testing starts in April, 2000.

Appendix 3
Information: New York

The state of New York began to appropriate funds for support of public schools in 1795. In 1814, all New York municipalities were required to participate in a statewide system of public school districts. At the time, these schools charged tuition to cover differences between operating costs and state funding. In 1867, free public schooling became a requirement of law. The current school year of 180 days was set in 1913, and the current school-leaving age of 16 was set in 1936.

The New York Board of Regents, originally responsible for supervising higher education, began high-school entrance examinations in 1865, later called "preliminary"

examinations. In 1878 it began examinations for graduation from high schools. In the 1880s the Board began inspection visits to public schools. A 1904 reorganization put the Board of Regents in charge of standards for all public education. One response was gradual strengthening of secondary school attendance. Another was development of detailed curricula aimed at preparing students for higher education. Throughout the nineteenth and twentieth centuries, high-school students in New York have been able to obtain a "local" high-school diploma without meeting Regents examination requirements.

In the 1930s the New York City schools began administering the Metropolitan Achievement Test series, designed by The Psychological Corporation, for diagnosis and guidance. In the 1970s the New York Education Department began using this test series, now provided by Harcourt Brace Educational Measurement, for its statewide Pupil Evaluation Program. This program administered tests of reading and mathematics in grades 3 and 6, tests of writing in grade 5, and tests of social studies in grades 6 and 8. During the years 1993 to 1996, the Department gradually changed to the California Achievement Tests, provided by CTB/McGraw-Hill. Throughout these years, the Department also administered the Regents Preliminary Competency Tests of reading and writing in grades 8 and 9.

Beginning in 1999 the Education Department is replacing its elementary and secondary school tests with new Program Evaluation Tests, planned since 1994 and piloted during 1995 through 1998. These tests, developed by CTB/McGraw-Hill, are strictly timed and include questions in multiple choice, short answer, extended answer, essay and laboratory performance formats. Tests for English language arts, mathematics and science are to be administered in grades 4 and 8. Social studies tests are to be administered in grades 5 and 8. Test items are disclosed to the public. Only English language arts and mathematics tests are being given in 1999 and 2000. Tests are currently offered only in English.

The New York Regents high-school graduation examinations are by subject. Beginning with a few subjects, the examination catalog reached a peak of 68 subjects in 1925. After years of consolidation, by the 1960s the catalog was reduced to English, mathematics, science, social studies and certain foreign languages. Subsequent revisions introduced technical education subjects. In 1998 the Education Department announced a new series of statewide tests in English, mathematics, science, global history and geography, and U. S. history and government, starting in 1999. The new Regents examinations have been developed by CTB/McGraw-Hill. They are strictly timed and include questions in multiple choice, short answer, extended answer, essay and laboratory performance formats. Most tests are offered in English only; some have also been offered in Chinese, Hatian Creole, Korean, Russian and Spanish. Scores of 65 on all tests are now required for a Regents diploma, and scores of 55 are required for a "local" diploma. Beginning in 2005, there will be no more "local" diplomas.

A 1987 New York law requires an annual report from the Education Department covering enrollment, student achievement, graduation and dropout rates, and other topics. This is known as the School Report Card. Data tables accompanying these reports show numbers or percentages of students statewide and by school district receiving certain score levels on tests. School Report Card data tables are being released about 9 months after the end of a school year. Statewide percentages of grade enrollment receiving Regents examination scores in specified ranges are shown in Table 6.

Table 6
New York Regents Examination Scores, 1997 and 1998

1997			
Examination score	**55 or more**	**65 or more**	**85 or more**
Comprehensive English	63%	56%	17%
Mathematics I	66%	59%	29%
Biology	51%	44%	15%
US History	56%	48%	15%
Global Studies	57%	48%	14%
1998			
Examination score	**55 or more**	**65 or more**	**85 or more**
Comprehensive English	65%	57%	15%
Mathematics I	70%	62%	33%
Biology	51%	44%	16%
US History	60%	52%	17%
Global Studies	65%	56%	17%

Source of data: New York State Education Department, 1998 and 1999.

Although data tables for the 1999 Regents examinations are not yet available, a summary for the English language arts test has been released. It shows that statewide 78 percent of grade enrollment has received a score of 55 or more on this examination, much higher than the percentage on the previous Comprehensive English examination. However, in the New York City schools only 55 percent of grade enrollment has passed this examination, with 35 percent yet to attempt it.

Appendix 4. Information: Texas

The Republic of Texas enacted laws to support free public education in 1845, in anticipation of statehood later that year. It also created a state fund to provide part of the cost of the public school system. Through the rest of the century public education was limited to eight grades in many rural areas, although high schools were founded in cities. In 1911 Texas reorganized its state education system to provide public high schools in all rural areas.

In 1984 the Texas legislature passed House Bill 72, a public "school reform" law. This revised the state's financial support for education, providing more funds for low-income districts, and it directed the Texas Education Agency to establish school performance standards and administer a statewide high-school graduation test. Until 1990 Texas used a series of tests focused on minimum competence. In that year, as required by law, it began introducing over a four year period a testing program designed to raise the expected level of skills, using a new test series. The Texas Assessment of Academic Skills (TAAS) is a series of standard tests given in the third through tenth grades in reading, writing, mathematics and social studies. These tests are untimed and in multiple choice format except for essays in writing tests. They have been organized by National Computer Systems of Minneapolis, MN, as prime contractor. Harcourt Brace Educational Measurement performs test development; it has involved about 7,000 Texas educators in the process.

TASS tests are available in English and Spanish, and there is an alternate assessment process for students in special education. Satisfactory scores on the tenth-grade tests in reading, writing and mathematics are required for a high- school diploma. Texas also has standard tests on which passing scores are required to obtain credit for certain high-school courses, currently Algebra I, Biology I, English II and U. S. History. In 1999 passing three such tests in the tenth grade was made equivalent to passing the entire TASS series. Starting in 2005 a new Texas law will require high-school graduates to get passing scores on new standard tests of English language arts, mathematics, science and social studies, taken in the eleventh grade.

Since 1994 Texas has used an Accountability Rating System to report school and district performance. Schools are rated as "exemplary," "recognized," "acceptable" or "low

performing." The key criteria are TASS scores, for which large racial and ethnic differences have been documented. For rating purposes, students are classified in four groups: white, African-American, Hispanic and economically disadvantaged. To achieve school ratings, the minimum rating scores are required for each group. There are also requirements for high attendance and low dropout rates. Ratings are published in newspapers. Schools with strong ratings or progress receive financial rewards, currently a total of $2.5 million per year statewide.

Texas public colleges and universities have a standard qualifying examination, the Texas Academic Skills Program test. It is an untimed test of reading, writing and mathematics in multiple choice format, plus an essay, all prepared and administered by National Evaluation Systems, Inc., of Amherst, MA. No one is denied admission based on TASP scores, but passing scores are required to graduate from two-year colleges and to take junior and senior courses at four- year colleges. The test is waived for students with high enough scores on certain other tests.

Racial differences in Texas test scores are well documented (Texas Education Agency, 1998). According to Texas statistics, the percentage of success for TASP is about the same for men and women, but the percentage of success for whites is more than twice that for African-Americans. The success rate on the tenth-grade TAAS series in 1998 was 85 percent for white students, 60 percent for Hispanic students, and 56 percent for African-American students. So far, however, all legal challenges to racial and ethnic differences in Texas standard test scores have failed. New arguments are being used by plaintiffs seeking to overcome the judicial barriers encountered in previous lawsuits. There is no objective evidence to sustain the passing scores set by Texas for the TAAS high-school graduation examinations, and the state provides no program to assure that the tests cover what is taught in the schools (Haney, 1999).

Texas is in denial about the dropout rates its program appears to be causing; official statements claim substantial decreases in dropout rates, to 10-15 percent. U. S. Department of Education enrollment data indicate much higher dropout rates. Haney (1999, p. 22) notes that Texas Education Agency definitions of "drop out" have changed several times in the last ten years. Longitudinal dropout rates in Texas have been surveyed by an independent organization over several years. Their estimates for the school years ending in 1986 through 1999 are shown in Figure 4.

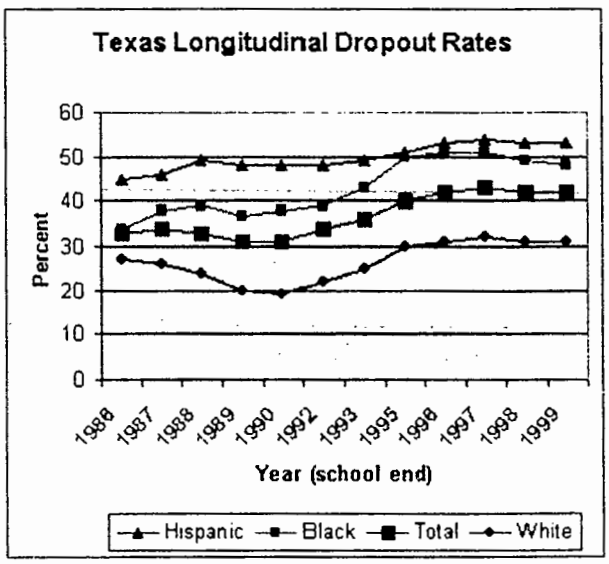


Figure 4. Texas dropout rates.

Source: *Longitudinal Attrition Rates in Texas Public High Schools, 1985-1986 to 1998-1999*, Intercultural Development Research Association, San Antonio, TX, 1999. By permission. Chart prepared by the author. Not shown are data for Asian/Pacific Islander and Native American students. No data were published for 1991 or 1994.

The estimates in Figure 4 are consistent with U. S. Department of Education data. They show that introduction of TAAS in 1990-1995 was associated with a significant

increase in dropout rates which has been sustained in the years since. Although the impact of TASS has been heaviest on African-American students, in some schools 100 percent of students with limited English proficiency drop out (IDRA, 1998). While the impact of TASS on Hispanic students has been less than the impact on African-American students, Hispanic students remain the group with the highest dropout rates.

Some students who do not receive a diploma at normal high-school graduation age continue in school and obtain a conventional diploma later, or they return to school after having dropped out, or they earn a certificate by passing the GED or a similar test, or they arrange to begin higher education without high-school credentials. U. S. Census data suggest that by age 24 half or more of high- school dropouts may have extended their education up to or beyond high-school equivalence. However, there is no consistent source of statistical data on these educational outcomes in Texas, in most other states, or for the U. S. (Heubert and Hauser, 1999, pp. 136-137 and 172).

Under TAAS, there have been reports of weeks spent on test cramming and "TAAS rallies." School ratings are raised by "exempting" students (Associated Press, 1999). Schools are allowed to contract for "test preparation" consultants and materials, and some have spent tens of thousands of dollars. There have been reports of falsifying results. In 1998 the Austin Independent School District produced dramatic TAAS score improvements; then in April, 1999, Deputy Superintendent Kay Psencik and the school district were indicted for tampering with government records. In Houston three teachers and a principal were dismissed for prompting students during test sessions ("TAAS scandal," 1999).

Illiteracy remains a major problem in Texas. Over 80 percent of Texas prison inmates have been found functionally illiterate. The four largest cities— Dallas, Houston, San Antonio and El Paso— have adult illiteracy rates of 12 to 19 percent. Statewide, the Texas adult illiteracy rate is 12 percent, second worst of any state in the U. S. (Census Bureau, 1992). In communities near the Mexican border, where rates are highest, illiteracy among children has increased during the years under TAAS (*Regional Profile,* 1999).

Appendix 5
Information: California

In 1961 California began programs of achievement testing in its public schools, with testing procedures and standards under local school district control. A 1972 state law created the California Assessment Program, under which multiple choice tests for reading, writing and mathematics were administered in grades 2, 3, 6 and 12, with grade 8 added in 1983. By 1987 a writing sample and a test for U. S. history and economics had been added. In 1988 the Board began to offer Golden State Examinations, intended to identify and honor outstanding students in public schools. In 1998 about 2,700 high-school graduates received merit diplomas based on these test scores.

In 1978 California voters passed Proposition 13, radically restricting local funds for schools in most communities. Passage of Proposition 62 in 1986 hobbled the ability of state government to assist with funding for education. Proposition 98, approved in 1988, set a school funding floor at a relatively low level and has tended to prevent further erosion. Since 1978 California has fallen from among the top ten states in many national ratings of education to among the bottom ten. California education initiatives since the 1970s must be viewed in the context of the state's flamboyant and reactionary politics and its drastic change in financial support for public schools.

A 1991 state law authorized a new California Learning Assessment System, and the previous testing program was gradually discontinued. In 1994 the new program died after a veto of legislation by the governor, leaving the state with no statewide testing except the Golden State Examinations. In 1995 new state laws established a Pupil Testing Incentive Program and required statewide standards. The Board of Education began to establish "curriculum frameworks," or required curricula (see McDonnell, 1997). In 1997, before the new testing program had been fully implemented, another new state law replaced it with requirements for revised curriculum standards and nationally normed standard tests, to be designated by the Board of Education. In 1997 and 1998 the Board of Education specified new content standards for reading, writing, mathematics, science, and history and social science (see McDonnell and Weatherford, 1999). Curriculum frameworks and corresponding tests are being revised and developed to correspond.

As required by the 1997 California law, the Board of Education began a Standardized

Testing and Reporting (STAR) Program in 1998. Its major component is annual administration of the Stanford Achievement Tests, published by Harcourt Brace Educational Measurement, to all students in grades 2 through 11. Grades 2 through 8 are tested in reading, writing, spelling and mathematics. Grades 9, 10 and 11 are tested in reading, writing, mathematics, science and social science. There are also "augmentation" tests in language arts and mathematics, intended to reflect the California curriculum, with additional tests in preparation.

By state law, STAR tests are provided only in English, although about forty percent of California's public school students come from Spanish-speaking households. These are strictly timed tests in multiple choice formats plus writing samples. Total testing time is about six hours. Parents may exempt their children from testing. Test items are not being disclosed to the public. California public schools are forbidden by law to use test preparation materials specifically designed for these tests. Their use by parents who can afford them is not restricted.

In April, 1999, the California legislature passed and its governor signed a law called the Public Schools Accountability Act. It requires the state to publish an Academic Performance Index (API) annually for each public school. It also provides extra funding for low performing schools and a system of awards for high performing schools. A total of $100 million was appropriated for awards in 1999. The 1999 law also requires the Board of Education to develop and administer promotion and graduation tests, starting in 2001. After three years, passing scores will be required to enter high school and to obtain a high-school diploma.

For 1999 the Board of Education defined the API on the basis of Stanford Achievement Test scores (California, 1999). It reflects student score ranks, weighted by subject content. Weights for grades 2 through 8 are reading 30 percent, writing 15 percent, spelling 15 percent, and mathematics 40 percent. Weights for grades 9 through 11 are 20 percent each for reading, writing, mathematics, science and social science. A school with all students ranking in the top 20 percent of the distribution of scores will have an API of 1,000, while a school with all students ranking in the bottom 20 percent will have an API of 200. The 1999 API ratings for California public schools are summarized in the Figure 5:

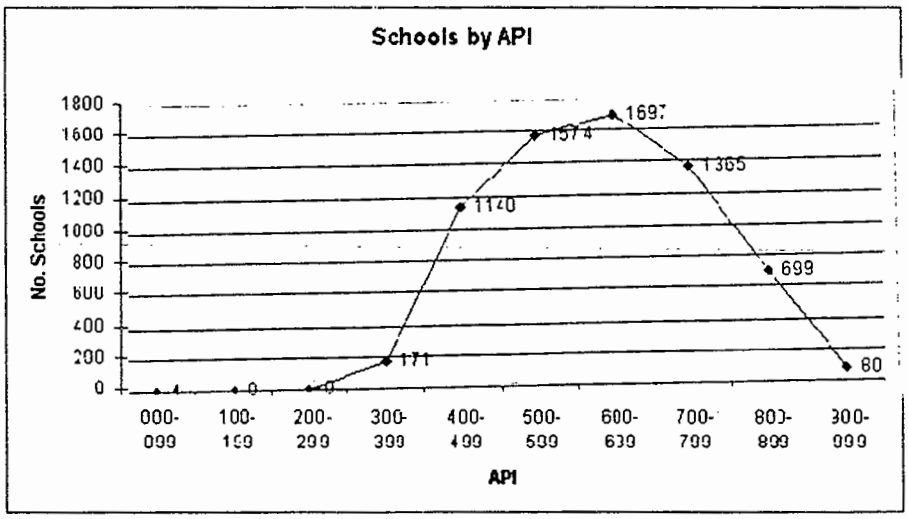


Figure 5. California school ratings.

Source: *Academic Performance Index School Rankings, 1999*, California Department of Education, Sacramento, CA, January, 2000. Chart prepared by the author. Data were grouped into the API ranges shown. Four schools were unrated.

The official goal is to raise all schools to an API of 800. Since the API is essentially comparing scores with averages, this is a "Lake Wobegon" goal, to make "all the kids above average."

Appendix 6
Performance Assessment

"Performance assessment is a broad term. It covers many different types of testing methods that require students to demonstrate their competencies or knowledge by creating an answer or product. It is best understood as a continuum of formats that range from the simplest student-constructed responses to comprehensive demonstrations or collections of large bodies of work over time. This [section] describes some common forms of performance assessment.

"**Constructed-response questions** require students to produce an answer to a question rather than to select from an array of possible answers (as multiple-choice items do). In constructed-response items, questions may have just one correct answer or may be more open ended, allowing a range of responses. The form can also vary: examples include answers supplied by filling in a blank; solving a mathematics problem; writing short answers; completing figural responses (drawing on a figure like a graph, illustration, or diagram); or writing out all the steps in a geometry proof.

"**Essays** have long been used to assess a student's understanding of a subject by having the student write a description, analysis, explanation, or summary in one or more paragraphs. Essays are used to demonstrate how well a student can use facts in context and structure a coherent discussion. Answering essay questions effectively requires analysis, synthesis, and critical thinking. Grading can be systematized by having subject matter specialists develop guidelines for responses and set quality standards. Scorers can then compare each student's essays against models that represent various levels of quality.

"**Writing** is the most common subject tested by performance assessment methods. Although multiple-choice tests can assess some of the components necessary for good writing (spelling, grammar, and word usage), having students write is considered a more comprehensive method of assessing composition skills. Writing enables students to demonstrate composition skills-inventing, revising, and clearly stating one's ideas to fit the purpose and the audience--as well as their knowledge of language, syntax, and grammar. There has been considerable research on the standardized and objective scoring of writing assessments.

"**Oral discourse** was the earliest form of performance assessment. Before paper and pencil, chalk, and slate became affordable, school children rehearsed their lessons, recited their sums, and rendered their poems and prose aloud. At the university level, rhetoric was interdisciplinary: reading, writing, and speaking were the media of public affairs. Today graduate students are tested at the master's and Ph.D. levels with an oral defense of dissertations. But oral interviews can also be used in assessments of young children, where written testing is inappropriate. An obvious example of oral assessment is in foreign languages: fluency can only be assessed by hearing the student speak. As video and audio make it possible to record performance, the use of oral presentations is likely to expand.

"**Exhibitions** are designed as comprehensive demonstrations of skills or competence. They often require students to produce a demonstration or live performance in class or before other audiences. Teachers or trained judges score performance against standards of excellence known to all participants ahead of time. Exhibitions require a broad range of competencies, are often interdisciplinary in focus, and require student initiative and creativity. They can take the form of competitions between individual students or groups, or may be collaborative projects that students work on over time.

"**Experiments** are used to test how well a student understands scientific concepts and can carry out scientific processes. As educators emphasize increased hands-on laboratory work in the science curriculum, they have advocated the development of assessments to test those skills more directly than conventional paper-and-pencil tests. A few states are developing standardized scientific tasks or experiments that all students must conduct to demonstrate understanding and skills. Developing hypotheses, planning and carrying out experiments, writing up findings, using the skills of measurement and estimation, and applying knowledge of scientific facts and underlying concepts—in a word, 'doing science'—are at the heart of these assessment activities.

"**Portfolios** are usually files or folders that contain collections of a student's work. They furnish a broad portrait of individual performance, assembled overtime. As students assemble their portfolios, they must evaluate their own work, a key feature of performance assessment. Portfolios are most common in writing and language arts-showing drafts, revisions, and works in progress. A few states and districts use portfolios for science, mathematics, and the arts; others are planning to use them for demonstrations of workplace readiness."

Source: Michael J. Feuer et al., Eds., *Testing in American Schools: Asking the Right Questions,* OTA-SET-519, Office of Technology Assessment, U. S. Congress, Washington, DC, 1992, p. 19.

Appendix 7
Chronology of Standard Testing in the U. S.

[Listed in brackets are some developments in other countries which had rapid and substantial impacts in the U. S.]

1900 The College Entrance Examination Board is founded at Columbia College in New York.

1905 [Alfred Binet publishes the first intelligence test, to identify slow learners.]

1908 Edward L. Thorndike, a Columbia professor, begins writing a series of standard achievement tests for use in elementary and high schools, completed in 1916.

1916 First publication of the Stanford-Binet IQ test by Houghton Mifflin, developed by Lewis M. Terman, a Stanford professor.

1916 Arthur S. Otis, a student of Terman and later a test editor for the World Book Company, invents the multiple choice format. It is used in the Army Alpha test.

1917 Robert M. Yerkes, a Harvard professor, organizes the Army Alpha and Beta intelligence tests, given to 1.7 million World War I recruits.

1921 The Psychological Corporation is founded in New York by James M. Cattell, Robert S. Woodworth and Edward L. Thorndike.

1923 First publication of the Stanford Achievement Tests by the World Book Company, developed under the direction of Lewis M. Terman.

1925 Carl C. Brigham, a Princeton professor, develops the Scholastic Aptitude Test for the College Entrance Examination Board.

1927 The California Test Bureau is founded in Los Angeles by Ethel M. Clark and Willis W. Clark, a Los Angeles school teacher.

1928 Everett F. Lindquist, a professor at the University of Iowa, begins the Iowa Testing Program in support of a scholarship competition.

1933 First publication of the Progressive Achievement Test series by the California Test Bureau, developed by Willis W. Clark and Ernest W. Tiegs.

1935 Louis L. Thurstone, a professor at the University of Chicago, publishes a theory of factor analysis as applied to psychometric testing.

1935 First publication of the Iowa Every-Pupil Test of Basic Skills by the University of Iowa Testing Bureau, developed under the direction of Everett F. Lindquist.

1936 IBM scores the New York Regents examination using a machine based on the Markograph soft pencil electrical technology invented by Reynold B. Johnson.

1938 The *Mental Measurements Yearbook* is first published by Oscar K. Buros,

a Rutgers University professor.

1940 Houghton Mifflin acquires publishing rights to the Iowa Test of Basic Skills.

1941 The U. S. armed forces begin using the Army General Classification Test and other standardized tests, given to more than 10 million World War II recruits.

1942 First publication of the Iowa Tests of Educational Development by Houghton Mifflin, developed under the direction of Everett F. Lindquist.

1942 The College Entrance Examination Board replaces its traditional essay tests with multiple choice tests.

1943 Everett F. Lindquist first administers the Test of General Educational Development (GED).

[**1944** Great Britain's Parliament approves the Education Act of 1944, beginning the "eleven-plus" examination restricting admission to grammar schools and access to higher education.]

1947 The Educational Testing Service is founded by Henry Chauncey to prepare and administer the Scholastic Aptitude Test (SAT) for the College Entrance Examination Board.

1949 First publication of the Weschler Intelligence Scales by The Psychological Corporation, developed by David Weschler, a professor at NYU Medical College.

1956 Houghton Mifflin introduces electronic scanners developed by Everett F. Lindquist and Albert N. Hieronymous, scoring test sheets on both sides without requiring soft pencil markings.

1958 The Educational Testing Service begins disclosing its SAT scores to test-takers.

1959 The American College Testing (ACT) Program is founded by Everett F. Lindquist and Theodore McCarrel.

1960 Harcourt Brace and Co. acquires the World Book Publishing Co. and its Stanford Achievement Test series.

1968 McGraw-Hill acquires the California Testing Bureau and its CTB Achievement Test series.

1969 Michigan begins a statewide program of standard testing, later expanded to high-school graduation requirements.

1970 Harcourt Brace acquires The Psychological Corporation.

[**1976** Key research findings on the inheritance of intelligence by Cyril Burt, a former professor at University College, London, are exposed as scientific fraud.]

1979 Houghton Mifflin establishes a Riverside Publishing division to publish the Iowa achievement tests, Stanford-Binet IQ test and other school-based standard tests.

1979 New York's legislature passes and its governor signs the Educational Testing Act of 1979, a "truth in testing" law.

1983 The Reagan administration publishes *A Nation at Risk,* embracing a system of school-based standard tests and punitive sanctions for low scores.

1984 Texas begins a statewide program of standard testing, to be required in ten years for high-school graduation.

1985 The National Center for Fair and Open Testing is founded in Cambridge, MA.

1991 The Bush administration's proposed Excellence in Education Act, H.R. 2460, to create federal school and employment tests, is defeated in Congress.

1996 California begins a statewide program of standard testing, to be required in eight years for middle school and high-school graduation, with state receivership for schools with low scores.

1998 Massachusetts begins a statewide program of standard testing, to be required in five years for high-school graduation, with replacement of principals in schools with low scores.

Appendix 8
General Bibliography

David C. Berliner and Bruce J. Biddle, *The Manufactured Crisis: Myths, Fraud, and the Attack on America's Public Schools,* Addison-Wesley, Reading, MA, 1995.

James Crouse and Dale Trusheim, *The Case Against the SAT,* University of Chicago Press, Chicago, 1988.

David A. Goslin, *The Search for Ability: Standardized Testing in Social Perspective.* Russell Sage Foundation, New York, 1963.

Stephen J. Gould, *The Mismeasure of Man,* W. W. Norton and Co., New York, 1981.

Robert L. Hayman, Jr., *The Smart Culture: Society, Intelligence, and Law,* New York University Press, New York, 1997.

Jay P. Heubert and Robert M. Hauser, Eds., *High Stakes Testing for Tracking, Promotion and Graduation,* National Academy Press, Washington, DC, 1999.
Banesh Hoffmann, *The Tyranny of Testing,* Crowell-Collier Publishing Co., New York, 1962.

National Center for Education Statistics (Jay R. Campbell, Kristin E. Voelkl and Patricia L. Donahue, Eds.), *NAEP 1996 Trends in Academic Progress,* NCES 97-985, U. S. Department of Education, Washington, DC, 1997.

Nicholas Lemann, *The Big Test: The Secret History of the American Meritocracy,* Farrar, Straus and Giroux, New York, 1999.

Office of Technology Assessment (Michael J. Feuer et al., Eds.), *Testing in American Schools: Asking the Right Questions,* OTA-SET-519, U. S. Congress, Washington, DC, 1992.

David Owen, *None of the Above: Behind the Myth of Scholastic Aptitude,* Houghton Mifflin, Boston, 1985. David Owen and Marilyn Doerr, *None of the Above: Behind the Myth of Scholastic Aptitude,* Rowman and Littlefield Publishers, Lanham, MD, Revised and Updated Edition, 1999.

Tim B. Rogers, *The Psychological Testing Enterprise: An Introduction,* Brooks/Cole

Publishing Co., Pacific Grove, CA, 1995.

Peter Sacks, *Standardized Minds: The High Price of America's Testing Culture and What We Can Do to Change It,* Perseus Books, Cambridge, MA, 1999.

David B. Tyack, *The One Best System: A History of American Urban Education,* Harvard University Press, Cambridge, MA, 1974.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois- UC

Robert T. Stout
Arizona State University

Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

BEST COPY AVAILABLE

This article has been retrieved 304 times since May 20, 2000

prior vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 8 Number 24 May 20, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Education Reform in Hong Kong: Issues of Consistency, Connectedness and Culture

Chris Dowson
Peter Bodycott
Hong Kong Institute of Education

Allan Walker
David Coniam
The Chinese University of Hong Kong

Abstract

Since the early 1990s, the pace of educational reform in Hong Kong has accelerated and broadened to incorporate almost all areas of schooling. The reforms introduced during this period can be subsumed under what has generally been labelled the quality movement. In this paper, we review and comment on a number of policy reform initiatives in the four areas of "Quality Education," English Language Benchmarking, Initial Teacher Training and the Integration of Pupils with Special Needs into Ordinary Classrooms. Following a brief description of each policy initiative, the reforms are discussed in terms of their *consistency, coherence* and *cultural* fit.

Since the early 1990s, the pace of educational reform in Hong Kong has accelerated and broadened to incorporate almost all areas of schooling. The reforms introduced during this period can be subsumed under what has generally been labelled the "quality movement." This stands in contrast to reform thrusts in previous decades, which tended to target the quantitative aspects of schooling. The shift from quantity to quality has been driven by at least four interrelated reasons. The first is the successful introduction of nine-year compulsory education in Hong Kong. All students in Hong Kong, regardless of background, are now guaranteed access to schooling to at least Secondary 3 (Grade 9). The second reason has been the growing

dissatisfaction from both employers and higher education bodies with student and teacher performance. Related concerns have prompted a search for higher standards and calls for increased accountability. A related argument has been a growing concern for greater economic competitiveness. The third reason has been the perceived need to secure stability and prosperity for all citizens following the change of sovereignty in July 1997. Finally, the quest for quality education in other countries has influenced Hong Kong policy makers and subsequent calls for reform.

In this article, we review and comment on a number of policy reform initiatives introduced in Hong Kong during the 1990s. We do not attempt a thorough review of each policy but rather we set out to describe briefly the initiatives and then analyse them for *consistency*, *connectedness* and *cultural fit*. For the purposes of this paper, *consistency* refers to how the thrust of the reforms and reform components are interpreted. That is, are the reforms consistent, or do they confuse educators through proposing apparently contradictory purposes. *Connectedness* refers to whether reforms or reform components are linked in terms of what they are trying to achieve and how they are achieved. Questions can be asked as to whether the huge array of quality reforms in Hong Kong are coherently connected to each other at the various levels. *Cultural fit* refers to whether the reforms and reform components are appropriate given the unique culture and context of Hong Kong and Hong Kong's educational institutions.

Background to Reform

Soon after assuming office on July 1st, 1997, Tung Chee-wah —the first Chief Executive of the Hong Kong Special Administrative Region of China (HKSAR)—promised an ambitious public spending program, including a massive boost to spending on education. His second policy address in October 1998 included few new initiatives and reiterated the directions established in 1997. The bulk of the policy directives, with the exception of Information Technology, had been in train, to varying degrees, for a number of years.

In 1997, Tung promised expanded investment in basic education through a 7.6% increase in concurrent expenditure and additional capital expenditure of approximately US$2.8 billion. Increased funding was intended to support a number of what have become continued initiatives. The first group of initiatives targeted directly the promotion of "quality education." This included the establishment of a US$650 million Quality Education Fund (QEF), a strong move toward School-Based Management (SBM) and a review of the entire education system. Some of these reforms were spelt out in detail in Education Commission Report Number 7 (ECR7) (Education Commission, 1997). The second suite of initiatives focused specifically on improving the quality of teachers. These included requiring all new teachers to acquire degree status, the upgrading of graduate posts in primary schools and the proposed establishment of a General Teaching Council. The third group of reforms targeted the perennially contentious issue of language enhancement. These included the introduction of the Native—Speaking English Teacher Scheme (NETS), the development of a new Putonghua (Mandarin) curriculum and the development of language benchmarks in English, Chinese and Putonghua. Within these policies, the development of English language benchmarks for all teachers has ignited significant policy debate. Other policies have resulted in increased support to special schools and kindergartens, improved provision for new immigrants to Hong Kong from Mainland China, accelerated movement toward building whole-day primary schools and a massive infusion of Information Technology into schools. One of the major reforms aimed to encourage ordinary schools to admit disabled students with concomitant support, and to establish a two-year pilot study on integration to help formulate a long-term policy on integration (HKSAR Chief Executive's Policy Address, 1997) .

Given the number of reform initiatives placed in train during the 1990s, we will concentrate analysis on four policy areas that are in many ways representative of the current broader reform movement. We do not suggest that these are necessarily the major components but they do exemplify the flavour of the current environment. The four policy areas analysed are:

1. School management
2. English language benchmarking
3. Initial teacher training
4. The integration of students with special needs into regular classrooms

Following a brief description of each policy initiative we will discuss them in terms of their *consistency*, *connectedness* and *cultural fit*.

School management

The ECR7 report (Education Commission, 1997) focused on "ways to improve school management and performance towards the provision of quality school education to better meet the needs of students". Much of the emphasis of the ECR7 Quality Education Reform Initiative drew on an earlier initiative labelled the School Management Initiative (SMI) implemented in 1991 (Education and Manpower Branch and Education Department, 1991). The SMI aimed to devolve responsibility and authority to the school level. While ECR7 continued the trend set in motion by the SMI, it did so with a different emphasis. Whereas SMI primarily aimed to introduce a system of SBM, founded on the body of school effectiveness research, the thrust of ECR7 was to develop quality schools possessing quality cultures, and to introduce a framework to monitor and assure quality. This marked change in nomenclature from "effective schools" to "quality schools" reflects the general policy move in Hong Kong toward quality (Dimmock & Walker, 1998b).

The ECR7 report suggested that many school and system problems centered on the lack of a quality culture. In justifying this claim the report points out that many schools do not have development plans linked to goal achievement; most schools do not have clear targets for both academic and non-academic students; and many do not have appraisal systems to assess the performance of principals and teachers. In addition, there is a perceived lack of support for schools in promoting a quality culture. There is also concern expressed about principal preparation and teacher training programmes, which they saw as inadequate in preparing professionals to cope with the changes required. The report singled out the Education Department (ED) for not adequately promoting quality development in schools, expressed the frustration which many schools felt through inflexible funding arrangements and asserted that there was only scant recognition of the "value-added" efforts made by schools to develop their students' potential. Although ECR7 mainly targeted change at the school level, it is a worthwhile vehicle for reflecting the general quality thrust that has dominated the Hong Kong reform environment of the late 1990s. This is explicitly stated in the policy document.

> While ECR7 focuses mainly on issues of quality school education in the context of public sector primary and secondary schools, in particular ways to improve school management and performance. This move towards the provision of quality education to better meet the needs of students, and the principle behind the various recommendations, is of a generic nature, applicable to all levels of education, and aims to provide a practical framework for the inculcation of a quality culture in the entire education system. (Education Commission, 1997, p.5).

An important area in developing a quality education culture is how teacher education policies are restructured. In terms of schools, perhaps the most far reaching policy has the been the establishment of the Hong Kong Institute of Education and their quest to change the face of teacher education in Hong Kong.

Teacher Education Reform

Teacher education in Hong Kong up until 1995 was largely the responsibility of four Colleges of Education and an Institute of Language in Education (ILE). These institutions provided non-graduate training courses for both primary and secondary teachers. In 1992, the Education Commission Report No. 5 (ECR5) was released. It recommended three reforms that would impact significantly on education at all levels in Hong Kong. The first was the recommendation of an expansion of tertiary

education to provide greater opportunities for graduate teacher training, and the second was an increase in graduate posts in both primary and secondary schools. The third recommendation was to amalgamate the existing colleges and the ILE into a unitary Institute of Education. The mission of the new Hong Kong Institute of Education (HKIEd) was to become a centre of excellence in teacher education and continuous professional development. This would be achieved, initially, through the provision of sub-degree courses and later through degree-level courses.

The amalgamation was completed in 1995, and in 1997 staff of the HKIEd moved into a new purpose-built facility fully dedicated to teacher education. Following a full institutional review in late 1996, the HKIEd was admitted to the governing body of tertiary education in Hong Kong – the University Grants Committee. In November 1997, following the new Chief Executive's address emphasising a commitment to quality education and an all graduate teaching profession, and the release of ECR7, the HKIEd had the first of two new teacher education courses validated by the Hong Kong Council for Academic Accreditation. These were a Postgraduate Diploma in Education (PGDE) and a four-year Bachelor of Education (Honours) for primary teachers. The first intakes of degree-level and PGDE students were admitted in September 1998. Currently the HKIEd offers 53 courses for 9,500 students and has a staff of 400. The HKIEd is an institution born of reform, and as the main teacher education provider in Hong Kong, it continues to reform itself through internal restructuring, the addition of new courses and the upgrading of staff.

The teacher education reform initiative has encountered significant challenges in its implementation and these will be discussed in subsequent sections of the paper. Another reform initiative that continues to create significant debate is the decision to tackle perceived declines in language standards through the compulsory language benchmarking of teachers.

English Language Benchmarking

In late 1995, the Education Commission published Report Number 6 (ECR6) (Education Commission, 1995). This report responded to the concerns expressed by Government, business and educational bodies about declining standards of language skills. The report argues a need for high level language skills among the workforce in Hong Kong, especially as it moves from a manufacturing to a service industry base. ECR6 highlighted a number of areas for action with regard to language standards. Specifically, the report recommended:

> The concept of "benchmark" qualifications for all language teachers should be explored by the Advisory Committee on Teacher Education and Qualifications (ACTEQ) with a view to making proposals to the Government as early as possible in 1996.
>
> Minimum language proficiency standards should be specified, which all teachers (not just teachers of language subjects) should meet before they obtain their initial professional qualification. The standards should be designed to ensure that new teachers are competent to teach through the chosen medium of instruction. (Education Commission, 1995, p.16)

The movement toward benchmark qualifications for all language teachers foretold the new HKSAR Government's quality education agenda—the desire for a fully trained language teaching profession in primary and secondary schools. The benchmark policy initiative would effect all teachers in Hong Kong, not only those who are language teachers of Chinese, English and Putonghua, but also teachers of other subjects who operate in either a Chinese or English language medium. The initiative, by its nature, will, once implemented, directly affect the lives and careers of thousands of people and ultimately the lives of children in Hong Kong schools. Therefore, to ensure quality and representativeness of stakeholders in the process, a great deal of interaction, discussion and consultation was subsequently undertaken with relevant bodies and individuals such as principals, teachers, and other members of the education profession. Other institutional bodies, members of Government, and

lay persons in the public, business and commercial sectors were also consulted.

The extensive trialing and piloting of the proposed language benchmarks continues and has been approached from an incremental and phased perspective. The process, which has taken course from late 1996 to the present, has included:

1. A Subject Committee composed of approximately 30 members from tertiary teacher education institutes, teachers and principals and teachers from local schools, as well as members from Education Department and other bodies involved in teacher education in Hong Kong was established. The brief of this committee was to set examination specifications and an examination syllabus.
2. For each of the five test papers, Moderation Committees were set up under the aegis of the Hong Kong Examinations Authority to produce sample material for distribution to teachers.
3. A representative random sample of approximately 400 teachers for English were invited to take part in a pilot assessment exercise so that actual levels of ability might be estimated, in order to compare actual levels of ability with desirable standards recommended by the Subject Committee.

The HKSAR Government's targets for the implementation of benchmarks are that:

- Initial benchmarks for teachers of English language in lower secondary schools should be finalised by mid 1999.
- Benchmarks as exit standards in the Teacher Education institutions are expected to be implemented by 2000—2001.
- All serving language teachers should be benchmarked by 2005, and all teachers who teach through the medium of English or Chinese should be benchmarked by 2008.

The proposed benchmark initiative, if successfully implemented, will have a profound effect on the teaching profession in Hong Kong. It remains, however, to be a very contentious issue.

Integration Reform

The policy shift from special school placement toward the integration of disabled students into mainstream classrooms began in 1986. However, despite recommendations concerning the re-skilling of regular teachers for supporting students with learning needs, minimal implementation followed. In response to concern from parents of disabled students, the ED recommended that a study be made of how integration might best be achieved. In addition, The Board of Education (1997) noted that regular primary classrooms contain significant numbers of students who are experiencing difficulty in learning and that this trend would continue in the future.

Whether the needs of these children will be fully met, and whether teachers are adequately trained to meet their needs, are issues that continue to be debated. Recommendations have been made that course providers in Special Education work toward improving the course content and structure of programmes designed for Special Education teachers and that Special Education be strengthened in initial teacher education programmes (Board of Education, 1996).

The 1997 Report on the Review of 9-year Compulsory Education specifically identified three major areas of concern that involve meeting the needs of students with special educational needs in regular classrooms. These deal with the range of individual differences, behavioural problems, and learning differences. Other indicators of the need for broader training in special education have emerged from seminars and workshops run by the Professional Teachers' Union. These meetings have given rise to the development of papers that have been submitted to the Education Department suggesting that regular class teachers must be adequately prepared to work effectively with low achieving students. Finally, Wilson (1997) raises the issue of gifted and talented students in Hong Kong. He suggests that catering for these students will help them achieve their potential, and benefit society.

Consistency, connectedness and cultural fit

Though the reforms briefly discussed are considered, on the whole, progressive, a number of interrelated issues can be raised in relation to their implementation and acceptance at an organisational level. We now analyse the policies in terms of their *consistency*, *connectedness* and *cultural fit*. These frames are defined below. The analysis will touch upon certain parts of the policies only.

> *Consistency* refers to how people interpret the thrust of the reforms and reform components or whether they in fact confuse educators through proposing apparently contradictory purposes. Questions asked include: Are the thrusts of the reforms consistent? That is, do they send contradictory meanings to those charged with implementing the reforms in their organisations?
> *Connectedness* refers to whether the reforms or reform components are linked in terms of what they are trying to achieve and how they are achieved. Questions can be asked about whether the huge array of quality reforms are connected to each other coherently at various levels: Are the thrusts of the reforms coherent? That is, are the reforms purposefully linked to each other?
>
> *Cultural fit* refers to whether the reforms and reform components are appropriate given the unique culture and context of Hong Kong and Hong Kong's educational institutions. The questions guiding this frame include: Are the thrusts of the reforms culturally appropriate? That is, are the reforms in their present forms appropriate for the Hong Kong culture and context?

Consistency in School Management

The reforms proposed in ECR7 do not present an overly consistent picture. This is reflected within and between a number of other reforms. For example, one form of inconsistency for educators in schools is between the simultaneous demand for internally driven improvement— agendas supposedly decided upon by the school to meet its unique needs—and externally driven demands for accountability. One example can be drawn from the ECR7 policy document. It states: "In proposing ways to improve the quality of school education, we consider some common standards and measures necessary. However, we are mindful to avoid uniformity which may overly restrict or restrain schools from developing their own characteristics" (p. 6). The tension between these dual aims becomes even more pronounced in other sections of the document. The example below illustrates pressures for diversity in Hong Kong schools arising from ECR7.

> School education in a modern society should be pluralistic. We should allow schools to pursue their own goals and improve performance in different domains with a variety of approaches. To involve teachers, parents and students in school management is conducive to the development of quality school education. This will not only help balanced development of students and gain the support of parents, but also enable the school to collate effectively views of teachers. (p 17)

In the same document are equally strong requirements for accountability and for conformity. In their pursuit of quality education, the ED proposes the adoption of a "whole-school approach" to inspections, which calls on an external panel of "experts" to evaluate the performance of schools. In order to build a quality culture in schools, a number of measures must be taken. They include:

- setting clear and commonly accepted goals for school education and having these goals clearly understood by all players in the school system;
- translating the goals into achievable, observable and measurable quality indicators;

- developing indicators for assessing school aims and using these indicators as the basis for school plans and external assessment. (pp. 7-8)

The issue then is not one of whether quality assurance programs are necessary, but that schools are often confused by inconsistent system pressures calling for both individual action and direction and imposed accountability. An unintended outcome of regulatory mechanisms, such as quality assurance, may be a tendency toward risk avoidance and orthodoxy in many schools which, in turn, can detract for other facets of the reform.

Consistency in Teacher Education

Internal and external pressures have fuelled the rapid and dynamic pace of teacher education reform. During the 1990's there were significant changes in the Directorate of the HKIEd resulting in the almost totally restructuring of the organisation. Similarly, the change in Government of Hong Kong brought with it a fresh emphasis on improving education, in particular the hastened call for an all-graduate teaching force.

The result has been an inconsistency in the way HKIEd staff behave and respond to reforms based on ideological differences about the nature of graduate-level academic study. Within the HKIEd a tension existed, more notably during the initial development of degree-level courses between what can be loosely described as academic rationalists and social-constructivist educators. Academic rationalists placed emphasis on ownership of subject content, focus the teaching content on the development of subject knowledge and more summative modes of assessment. Academic rigour and the desire for external accountability were seen to drive these lines of thinking. However, social constructivist educators placed greater emphasis on the integration of subject knowledge, pedagogical knowledge and teaching methods. The modes of assessment used reflected similar integration and a greater emphasis on process than product. The tensions were amplified by a lack of direction and inconsistent feedback through reports from Government about the preferred qualities of Hong Kong teachers and, to some extent, by the background experiences of staff.

Consistency in English Language Benchmarking

It has been mentioned above that the benchmark initiative deals with three languages—English, Chinese and Putonghua. Standards should therefore be consistent across the three languages. There has been a considerable difference in approaches to the benchmarks for the three languages in terms of philosophy and well as in the approach to marking. For example, with reference to marking, it needs to be considered whether the approach should be from the positive viewpoint of "can do" skills, as opposed to penalising a teacher for errors and failing someone after a certain number of errors have been made.

One issue that has aroused great controversy in the local media focuses on who should be benchmarked. The initial thrust of the benchmarking exercise focused on establishing benchmarks for lower secondary school teachers of English language, for Chinese as a medium of instruction in primary schools and Putonghua as a foreign language in secondary schools. If the Government's claim that teacher standards in language ability form a cornerstone in the upgrading of education, it is crucial that the exercise not stop at this initial cohort of teachers but continue to examine teachers and teacher educators across all sectors of education.

It has been agreed by many sectors of education that benchmarks should be introduced for teachers in pre-service training. What is less clear is the extent to which the policy will be implemented for in-service teachers. As might be expected, there is considerable opposition from serving teachers (with marked pressure from the Professional Teachers' Union) who state that serving teachers have already been certified and therefore do not need to be "re-certified."

A further case concerns exemptions in terms of whether – or indeed *should*—any teacher(s) be exempted in terms of qualifications, background or age. This is a very contentious issue, as exemptions need to be examined on a

BEST COPY AVAILABLE

case-by-case basis.

Raising standards requires a substantial financial commitment. On this basis, it must be stated that the HKSAR Government is being consistent in its approach to the upgrading of education. It realises that it cannot be done on the cheap. Recurrent resources have been set aside (some US$100 million for the period 2000 to 2008), so that language courses are available for every teacher in Hong Kong (there are approximately 50,000 teachers across the different educational strata in Hong Kong). It is expected that these teachers will want to enrol on such courses.

Consistency in Integration in Special Education

Arguably one of the most glaring inconsistencies in integration is the practice of integrating disabled students into regular schools by placing them in special classrooms within the schools. This is at odds with a recent equal opportunity ordinance aimed at eliminating discrimination against the disabled (Disability Discrimination Ordinance, 1998). There are further problems of inconsistency between policy and its interpretation. For example, *inclusion* has been called "integration," "mainstreaming" and "normalisation," and schools have interpreted each of these terms differently. Another inconsistency stems from a mis-transferrance from small-scale research findings to larger scale implementation.

Connectedness in School Management

Many of the reforms in Hong Kong have been driven by different educational, political, economic and social agendas. Some policies, such as the Target Oriented Curriculum (TOC) and the SMI were introduced during British rule as a means of democratising education. Others were introduced to smooth the change of sovereignty and yet other to address political calls for an increase in standards. Often, these reforms have been simply stacked on top each other with little consideration of how they support or relate to each other.

As an example, consider ECR7 and the Target Oriented Curriculum—the major school curriculum reform vehicle. TOC is directed at teachers in the classroom while ERC7 largely provides administrative, organisational and structural strategies for school reform. ECR7's effects are felt mostly at the whole-school and department levels rather than at individual teacher and classroom level. If school performance is most directly affected by quality teaching, learning and curricula, then ECR7, with its focus on management and governance, stops short of penetrating to the classroom-teacher level. It then becomes an act of faith to believe that SBM will necessarily transform the variables, which directly impact on school performance—namely, the cognition and behaviours of teachers and students in classrooms. ECR7 uses the core concept of school culture but offers little on how to build such cultures to promote quality teaching, learning and curricula.

TOC, on the other hand, aims to influence student learning at the classroom level and neglects the organisational level. TOC is not even mentioned in the ECR7 document. Therefore, the question is whether policy makers have considered the linkages—how the reforms support each other —between these two key areas? The answer appears to be "no." Both reforms are perceived as discrete entities, the former seen as the business of principals and senior teachers, the latter, the concern of classroom teachers. Both reforms need to be considered as an integral whole and all stakeholders need an appreciation and understanding of how they can best enhance student learning and school performance (Dimmock & Walker, 1998a).

Connectedness in Teacher Education

One cited reason behind the teacher education reform initiatives was the perceived need for teachers to cope with an increased range of curriculum reforms. However, teachers and teacher educators have struggled with these reform policy initiatives because of a lack of connectedness between them. For example, the relationship between TOC, integration, and benchmarking, at a macro and micro level has not been made clear. Reform guidelines lack detail or stated expectations,

and therefore individuals within the education community including teacher educators are forced to second guess the exact nature of the reform and how it may or may not connect with other reforms.

Within the HKIEd this has led in some cases to significant differences in understandings about the reform intent and in respect to responsibility for developing reform related materials. The result has been confusion and conflict about the effect of reform implementation at both the tertiary and school levels.

Connectedness in Benchmarking

We have discussed the issue of improving education through the perspective of upgrading teacher professionalism. While language is important, it is only one aspect of an able teacher, however. Holistically, one aspect of connectedness can be perceived from the declaration (HKSAR Chief Executive's Policy Address, 1997) that the teaching profession will move to an all-graduate profession, and that, from 2004, all teachers in secondary schools will need to hold a Postgraduate Diploma in Education (PGDE) in order to be able to work in schools – which is not currently the case. However, in terms of benchmarking, there is a perceived lack of connectedness between the design, development and test specifications of the benchmarks for the three languages (English, Chinese and Putonghua).

Connectedness in Integration

Current integration reform finds itself in competition with several other reforms simultaneously foisted onto schools. For example, under SBM, schools can make decisions about meeting their own needs and priorities. While this suggests that integration might be more readily achieved, the reality of the situation is that due to the vicissitudes of school examination results, when given a choice, schools will give priority to reforms which result in improved examination results—at the expense of integration. Many schools fail to perceive the connectedness between integration and other reforms. While this can be partly blamed on the unwillingness of schools to include students with special education and learning needs, ED has an obvious duty to connect with schools through communication and develop firmer bonds to counter this problem.

Culture in School Management

The final issue relates to the cultural applicability of educational reforms in Hong Kong. Reforms such as ECR7 are driven very much by global educational trends. For example, ECR7 is reflective of School-Based Management policies emanating from Western English-speaking countries. Given Hong Kong's status as a "colony" until very recently, the importation of the educational reform agenda is perhaps not surprising. Nonetheless, the phenomenon of exporting reforms from societies and importing them into others whose characteristics, values and conditions are different raises concerns about their cultural appropriateness.

While Hong Kong people display many characteristics of "Westernisation," the underpinning culture is very much Hong Kong Chinese. Among the questions this poses in regard to educational reform are the following:—to what extent are British, American and Australian policy blueprints appropriate to meet the educational needs of Hong Kong? For non-western societies, are there more appropriate alternatives to SBM and to curriculum reforms driven by student-centred approaches and learning outcomes? If there are not, then what, if any, adaptations to imported Western policies are needed? This is particularly relevant at the point of school implementation. These issues do not appear to have been seriously considered by policy makers but certainly must be dealt with continually at the school level (Dimmock & Walker, 1998b).

Culture in Teacher Education

The flow-on effect of educational reform in Hong Kong during the 1990's has resulted in significant changes to the preparation of teachers. The decision to create

the HKIEd has placed teacher education under the microscope, and increased attention on the quality of teacher educators. Many staff at the HKIEd feel they have been forced to join a university-type culture in which their experience, qualifications and professional practices are not valued. Staff are required to attain higher degrees, including doctorates, undertake research, publish in internationally recognised journals, undertake teaching attachments in local schools, and update the depth and breadth of their subject knowledge, teaching content and assessment practices. These changes are not out of the ordinary for many university-based teacher educators. However, for many staff, their origins and experience lay in sub-degree granting institutions, where the emphasis and expectations were somewhat different.

The shift to a university culture and associated work practices has resulted in significant tension within the institution. The emphasis on greater public accountability, staff appraisal, promotion and substantiation based increasingly on an individual's ability to conform to the shift in work culture, has resulted in the loss of experienced staff.

Culture and Benchmarking

The perspective of culture may be viewed from two angles. First, from the perspective of what might be termed "respect," the introduction of benchmarking will inevitably mean that teachers may risk a possible loss of standing. Having to sit an external test such as the benchmark test to prove their worth may mean a possible loss of face, certainly if they were to fail. Second, in many older, more established and traditional schools, a teacher is often regarded as a "sage." While it is acceptable for teachers to foist tests on their students and to make their students aware of their shortcomings, the possibility of being afforded the same treatment is creating some concern.

This also links to the perspective of an "exam culture." Hong Kong is a very exam-oriented society, where teachers frequently apply various benchmarks to their students' performance. However, when teachers themselves are subjected to a benchmark test in front of a live class, this puts a different face to the benchmark assessment. Teachers are apprehensive about the spread of the benchmark culture to include an assessment of own language ability.

Culture and Integration

As with other policy initiatives, integration reform has, in general, come from a Western perspective. Within schools, there are a number of potential cultural impediments. First, most schools are driven by the need to achieve highly in public examinations. Any threat to such achievement may result in open resistance to integration. Second, there is also a tendency for teachers to gear their teaching to the average achievers and ignore those who experience difficulty in learning.

Both these aspects strike at the heart of integration. There is little evidence of the Hong Kong Education Commission's 21st century blueprint push toward "...help (for) all its students whatever their ability..." The Hong Kong school culture is further characterised by curriculum rigidity. The need to teach to the examination is pervasive. Sometimes such rigidity is manifested by excessive adherence to the curriculum, or an outdated style of teaching. Disabled students need flexibility in what and how things are done. Cultures have differing attitudes toward disablement, and in some instances those who are different, may not be highly valued. It is only by education and supported exposure to disabled students that schools and personnel become less resistant to change. There is comfort in the status quo, usually set by the dominant culture, in this case, so called "normal people." The cultural status quo is maintained by the omission of disabled students from regular schools, and by their grouping into categorical special schools.

Conclusion

Issues of consistency, coherence and culture have led many within the educational community to become cynical about the "real" effects of educational reforms. Despite the noble purpose of many of the reforms, such cynicism, if left

unchecked, has the potential to further damage the efficacy and influence of the reforms at the level where they are intended to make a difference—at a school and classroom level. It is to be hoped that due consideration of the factors involved in reform implementation will lead to more positive and effective changes in the quality of education in Hong Kong.

As with most contexts, Hong Kong policy makers are continually making reforms. This is evident in Hong Kong, as Education Commission Report, No. 8 (ECR8) (Education Commission, 1999) is released with the publication of this paper. ECR8 proposes wide-ranging reforms to the Hong Kong educational system at kindergarten, elementary, secondary and tertiary levels of the educational system, and moots reforms which will serve to accentuate the issues of *consistency, coherence* and *culture discussed in this article.*

References

Dimmock, C., & Walker, A. (1998a). Comparative Educational Administration: Developing a Cross-Cultural Conceptual Framework. *Educational Administration Quarterly, 34*(4), 558-95.

Dimmock, C., & Walker, A. (1998b) Transforming Hong Kong Schools: Trends and emerging issues. *Journal of Educational Administration, 36*(5), 476-509.

Disabilities Discrimination Ordinance. (1998). In *Law of Hong Kong,* Chapter 487. Hong Kong: Government Printer.

Education and Manpower Branch and Education Department. (1991). *The School Management Initiative – Setting the Framework for Quality in Hong Kong Schools.* Hong Kong: Government Printer.

Education Commission. (1992). Education Commission Report No. 5. (ECR5). Hong Kong: Government Printer.

Education Commission. (1995). Education Commission Report No. 6. (ECR6). Hong Kong: Government Printer.

Education Commission. (1997). Education Commission Report No. 7. (ECR7). Hong Kong: Government Printer.

Education Commission. (2000). Education Commission Report No. 8. (ECR8). Hong Kong: Government Printer.

Education Commission. (1999). Education Blueprint for the 21st Century: Review of Academic System: Aims of Education: Consultation Document. Hong Kong: Government Printer.

HKSAR Chief Executive's Policy Address. (1997). Hong Kong: Government Printer.

The Board of Education. (1997). Report of the Sub- committee on Special Education. Hong Kong: Government Printer.

The Board of Education. (1997). Report on Review of 9- year Compulsory Education. Hong Kong: Government Printer.

Wilson, H. P. (1997). A Comparison of Policies and Implementation Strategies for the Education of Gifted and Talented Children in Selected Pacific Rim Countries. In J. Chan, R. Li & J. Spinks (Eds.), *Maximizing Potential: Lengthening and Strengthening our Stride.* Hong Kong: Social Sciences Research Centre, University of Hong Kong. Eleventh World Conference for Gifted and Talented Children, July-August 1995, University of Hong Kong, Hong Kong.

About the Authors

Chris Dowson
Senior Lecturer, Department of Special Needs
Hong Kong Institute of Education, Hong Kong.
Email: dowsonc@ied.edu.hk

Chris Dowson specializes in inclusion studies and communication. His research interests focus on hearing impairment, second language learning and quality indicators in education. He teaches courses on classroom practice and effective methods.

Peter Bodycott
Principal Lecturer, Department of English,
Hong Kong Institute of Education, Hong Kong

Peter Bodycott has taught in schools and higher education in Australia, Singapore and Hong Kong. His most recent teaching, research interest and publications focus on pre- service teacher thinking, second language learning and teaching, and the role of narrative in leadership and teacher education.

Allan Walker
Associate Professor
Department of Educational Administration and Policy
The Chinese University of Hong Kong

Allan Walker specializes in educational leadership and policy. His major research interests center on the influence of societal culture on educational administration and leadership, principal assessment and strategic planning in schools.

David Coniam
Associate Professor
Department of Curriculum and Instruction

The Chinese University of Hong Kong, Hong Kong
David Coniam is a teacher educator, working with ESL teachers in Hong Kong secondary schools. His main publication and research interests are in language testing, computational linguistics and language teaching methodology.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

BEST COPY AVAILABLE

This article has been retrieved 436 times since May 30, 2000

prior vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 8 Number 25 May 30, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Mathematics Achievement by Immigrant Children: A Comparison of Five English-speaking Countries

Gary G. Huang
Synectics for Management Decisions, Inc.
Arlington VA (U.S.A.)

Abstract

In this study, I examined academic achievement of immigrant children in the United States, Canada, England, Australia, and New Zealand. Analyzing data from the Third International Mathematics and Science Study (TIMSS), I gauged the performance gaps relating to the generation of immigration and the home language background. I found immigrant children's math and science achievement to be lower than the others only in England, the U.S., and Canada. Non-English language background was found in each country to relate to poor math and science learning and this disadvantage was stronger among native-born children—presumably children of indigenous groups—than among immigrant children. I also examined the school variation in math performance gaps, using hierarchical linear modeling (HLM) to each country's data. The patterns in which language- and generation-related math achievement gaps varied between schools are different in the five countries.

The public school system as an institution plays a critical role educating immigrant children and facilitating their participation in the larger society. This system in the U.S., succeeded in integrating European immigrants, is now facing a serious challenge as newcomers of non-European heritage have become the primary source of immigration over the decades. This shift in origins of the immigrants is a most striking development in U.S. immigration history (Fix, Passel, Enchautegui, & Zimmermann, 1994). Asians and Hispanics are the fastest growing groups among foreign-born population in the U.S., rising from 1.5 percent each in the early 1990s to 25 percent and 43 percent, respectively, in 1990 (Bureau of Census, 1993). Asian and Hispanic children, respectively, represent 3.5 percent and 14 percent of the U.S. elementary and secondary student enrollment in 1992, more than doubled from the 1.2 and 6.4 percents in 1976 (NCES, 1995).

Many developed nations share this challenge. The trend of globalization has brought rising waves of foreign labors, refugees, and immigrants into affluent countries. Today, the U.S., Canada, Australia, New Zealand, France, Germany, Britain, and other European countries are receiving newcomers from different regions of the world. The public schools in these countries confront the daunting task to educate children of immigrants.

Given the gravity of the issue, ironically, educators know little about the schooling of immigrant children. Little research has systematically dealt with the issue. It is unclear as to how the new generations of immigrants do in the school system and what their great diversity has to do with their schooling. It is even more uncertain about how schools are acting to help immigrant children learn math and science, subjects that are critical for competing in today's technology-oriented labor market. No baseline comparison is available regarding education of this group in the U.S. and other nations.

The lack of knowledge about immigrant children's education and general well being concerns educators and policymakers. The Federal Interagency Forum on Child and Family Statistics has published annual reports on children (Federal Interagency Forum on Child and Family Statistics, 1998). But the reports contain little information specifically about children of immigrant background. A recent study of immigrant children released by the National Research Council and the National Institute of Medicine points out that there is virtually no public dissemination of information on even the most basic indicators of the conditions of children in immigrant families (Hernandez & Carney, 1998). In a policy study report, the National Commission on Immigration Reform also calls for increased attention to and resources for immigrant children's schooling (see Schnaiberg, 1997). My study was intended to remedy this shortage of knowledge by comparing math and science performance of immigrant children in five English-speaking countries.

Literature Review and Research Questions

The available research on immigrant children's school performance is inconclusive even regarding the basic conditions of performance. Some studies suggest that the children of immigrants do better in school than the rest of American children; their performance is above averages (Rumbaut, 1996; also see Viadero, 1998, Lapin, 1998). In social adaptation, physical and mental health, foreign-born immigrant children were also seen to fare at least equally well as other children in the U.S. (Hernandez & Charney, 1998). On the other hand, there is evidence that immigrant children, especially Hispanics and others with impoverished background, suffer poor academic achievement and lower educational attainment (e.g., McPartland, 1998; Vernez & Abrahamse, 1996). A foremost concern for research is to provide clear description of this population's schooling with solid baseline indicator of performance.

Aggregated comparisons may mask crucial variation within the immigrant population. For example, while Hispanic adolescents of all generations have grade point averages and math test scores that are lower than those of white adolescents in U.S.-born families (NCES, 1998a), academic achievement of immigrant students appears to *decline by generations* (Hernandez & Charney, 1998). The social, economic, and cultural factors that either protect or disadvantage immigrant children are not well understood. Thus, baseline indicators should also summarize

performance differences by important subcategories of the immigrant children, such as generation of immigration, sex, native language, and socioeconomic status.

A small number of recent studies of immigrant children's academic achievement provide some insights for understanding the variation among immigrant children's academic achievement. For example, Hao and Bonstead-Bruns (1998) used the concept social capital to explain immigrant children's academic performance. This concept, though useful in understanding the behavioral and cultural attributes of immigrant groups affecting academic learning, is less relevant to study of the functioning of institutions, such as public schools. It is not clear from such research as to how schools could reduce the detriment caused by meager social capital for an immigrant child. Theories and research are needed to sort out institutional factors that account for the wide variation and the changing pattern of this population's academic performance. As a preliminary study intended to address some of these concerns, I examine the following issues in the analysis.

Generation Difference

The generation of immigration distinguishes a number of demographic characteristics among children from immigrant families. Compared with children in U.S.-born families, first-generation immigrant children (the foreign-born) are more likely to experience high poverty; to have a large family with both parents; and their parents are more likely to have attained little education yet to participate in labor force (Hernandez & Charney, 1988). Second-generation children (those born in the U.S. to at least one foreign- born parent) tend to experience substantially less risk than do first-generation children, but are likely to lose psychological resilience that the first generation often demonstrates. Such cross-generation distinctions imply different risks and strength for immigrant children's schooling.

The analysis first addresses the question about the performance gap relating to the generation of immigration in different countries. The second question is to what extent this gap differs across schools in each country. To answer this question, the analysis explores the variation of the generation gap across schools in each country. With the international test results available from the data, it should be particularly interesting to see how the school- level variation of the gap differs across countries. The resulting baseline indicator may reveal the extent to which the *overall* school setting relates to the variation of the gap—in contrast to the extent to which individual factors account for the variation. Future study may elucidate school roles in reducing the generation gap by examining specific school factors relating to the variation of the performance gap.

Language Barrier

Limited English proficiency handicaps immigrant children's learning on key subject areas such as mathematics and science. Language barriers are often more detrimental for children of low socioeconomic background. Living in socially and linguistically isolated communities, poor immigrant children can hardly improve their new language skills and the language barriers persist over the school years. On the other hand, bilingual proficiency, defined as the mastery of both the mother tongue and a new language, is found to be a strength for immigrant children's cognitive growth (e.g., Bumberger & Larson, 1998; Hao & Portes, 1998).

I first estimate the size of the math and science performance gaps related to non-English language background in each country. I then examine the variation of the gaps between schools in each country. While these baseline indicators are descriptive, they imply the extent to which the overall school context is associated with the variation of the gap-- relative to the individual level variation. The analysis may provide a ground for further study of specific school functions in reducing language-caused performance gap for immigrant children.

School Variation of Performance Gap

Does school has something to do with the performance gaps? It is conceivable that the average performance and the performance gap between immigrant children

and the other children may vary across schools. Schools with different demographic composition, resources, and curricular and instructional programs theoretically could achieve different levels of excellence and equity. Relevant to policymaking, gauging such school-level variation is crucial for further assessing institutional role in achieving educational equity. Understanding the school-level variation in performance gaps and school features relating to such variation can help school improve equity. In this preliminary analysis I only examine the school-level variance in math achievement gap relating to the generation of immigration and language backgrounds.

Data Source

TIMSS is the most comprehensive and rigorous international education comparison ever (NCES, 1998b). I extracted TIMSS Population 1 (students of grades 3-4 or ages 8-9) data of five English-speaking countries including the U.S., Canada, England, Australia, and New Zealand, with unweighted samples size of, respectively, 10,670, 14,639, 5,584, 10,433, and 4,670. Conducted in 1995, TIMSS researchers tested the mathematics and science knowledge of more than half a million students in 41 countries at three grade levels—primary, middle, and end of secondary school. TIMSS ensured that the participating students in each country were representative of its population. It generated information on the background and math and science achievement tests for children of the participating countries. While tests on math and science were administered to students, survey data were collected from teachers, schools, as well as students. The resulting information encompasses student demographic background and math learning experience; teachers' background and instruction; and school facilities, program provisions, and demographic attributes. Information for identifying foreign-born children is available, including the nation of birth for both the parents and the child.

The TIMSS nationally representative sample designs generated data for the population of each target age group (or grade level) in a country. The sample for a given age group in a country was selected in a two-level stratified design. In this design, a school sample representative to the national population of schools was drawn first, and within each selected school, typically one classroom at the target grade level was selected for the test and survey. While certain minority groups were oversampled, sample weights were provided to compensate the bias resulting from the oversampling. Unit nonresponse bias was corrected by sample weights as well.

The tests were designed through collaboration among experts from the participating countries. Recognizing vast differences in social and educational context, the tests were meant to measure students' general math and science knowledge and skills at the given age/grade. The results were widely accepted as valuable for cross-national comparison, given the caution of contextual differences among the participating nations (Forgione, 1998). Four items were used to identify students' immigrant background. They presented information about the child's birthplace (foreign- or native-born in one of the five countries), the number of years living in the current country, and the foreign-born status of the child's mother and father. I defined a child as a first-generation immigrant if the child was foreign- born regardless of the birthplace of the parents, and a second-generation immigrant if the child was born in the current country to one or both foreign-born parent; and the rest were considered as non-immigrants. With a data item about student home languages, I categorized students as a non-native language speaker if he or she reported that a language other than the TIMSS test language (English) was "often" or "always" spoken at home.

Analytical Methods

The analysis included two components. To generate baseline indicators of the overall performance patterns, I ran a series of descriptive analysis. To estimate school-level variance of performance gaps, I conducted two-level hierarchical linear modeling.

Descriptive Analysis

Descriptive analysis entailed comparing means of the test scores for the groups of interest. As specified earlier, baseline indicators of math and science performance gaps between immigrant and non-immigrant children will be estimated in a comparison of means with significance tests (all at the $p<.05$ level if not otherwise specified). All the remaining indicators will be generated by breaking down the test data by two categorical variables, immigrant status and non-English language background, with significance tests.

I ran the procedure with data for each country. The five plausible values for estimating performance on mathematics were used. The estimates from the five runs were then averaged as the final estimates in the comparisons (see TIMSS User's Guide for rationale for this special approach, International Study Center, 1988). Student-level sample weight (TOTWGT) was used to correct bias from unequal sampling of some student groups and unit nonresponse. I used jackknife procedures to correct the design effects caused by the stratified clustering sample design (rather than simple random design). See Chapters 5 and 7 of the User's Guide (International Study Center, 1998) for rationale of using sample weights and special procedures for correcting design effects.

HLM Procedure

To assess school-level variance of the performance gap relating to immigrant status, I used hierarchical linear modeling (HLM) technique (Bryk & Raudenbush, 1992). HLM was appropriate for this part of analysis because in the TIMSS design students as level-1 units were nested in schools (level 2) and HLM enabled me to separate the variance by two levels and to formally estimate the portion of variance taking place at school level.

In an unconditional (one-way ANOVA) with random effect model, I estimated variance separately at the student and school levels. This model answered the question as to whether schools differed from each other in average math performance. It provided basic estimates for making decision if it was necessary to further model the variance at the two levels. The unconditional models were:

At student level (level 1), $Y_{ij} = \beta_{0j} + r_{ij}$ *and*

at school level (level 2), $\beta_{0j} = \mu_{00} + u_{0j}$.

As the school level variance was sufficiently large (10 percent or more of the total variance, measured with the intraclass correlation coefficient) for each country, I specified random coefficient models to estimate school-level variance of the math achievement mean and achievement gaps associated to home language and the first- and second generation immigrant backgrounds (all student-level predictor variables were centered around the school mean). At level 1, the equation had the overall achievement mean, the average achievement differences relating to the non-English language and immigrant status, and the random error,

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{LANGUAGE}) + \beta_{2j}(\text{FIRST\_G}) + \beta_{3j}(\text{SECOND\_G}) + r_{ij}.$$

At level-2, the equations included no school variables but only the school average math score (the intercept) and the estimates of the variance around the average measures of the three gaps (the slope):

$$\beta_{0j} = \mu_{00} + u_{0j} \text{ and}$$

$$\beta_{qj} = \mu_{q0} + u_{qj} \text{ where } q = 1, 2, 3.$$

In case the gaps did not vary statistically significantly at the school level, the random effect u_{qj} *was removed from the equation and the effect was estimated only as fixed.*

I used the software package HLM (version 4.03) for the analysis, running the Plausible Value procedure available from the package (Bryk, Raudenbush, &

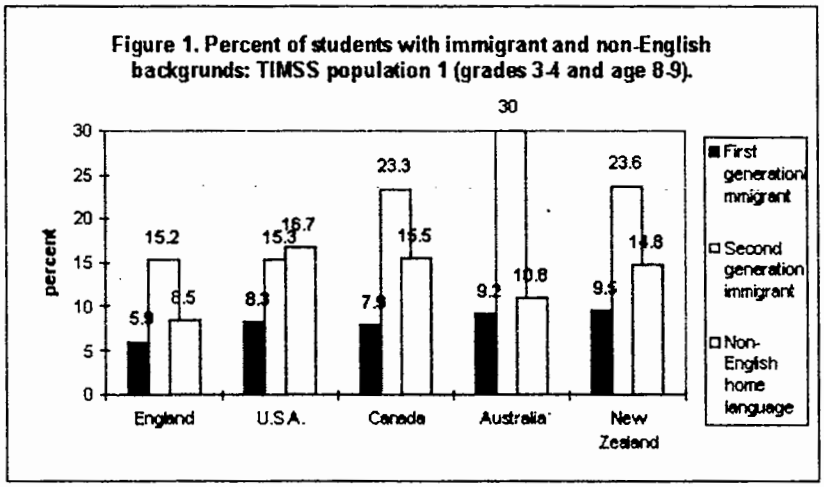
Congdon, 1996). This procedure included the five plausible values as the outcome variable and automatically averaged the resulting estimates after the runs. Normalized student level weight and school level weight were used in the procedures for generating the estimates to the student population in each country.

Findings

Students of Immigrant and Non-English Backgrounds

Each of the five countries' elementary student populations contained a substantial portion of students with immigrant and non- English backgrounds (see Figure 1). Australia and New Zealand had the highest rates of immigrant students of both the first and second generations, followed by Canada and the U.S. Strikingly, the second generation immigrant children comprised almost one third of Australia's population of third and fourth graders. The U.S. had a relatively high proportion of children of non-English background (16.7 percent), though this group was fairly large in Canada and New Zealand as well.

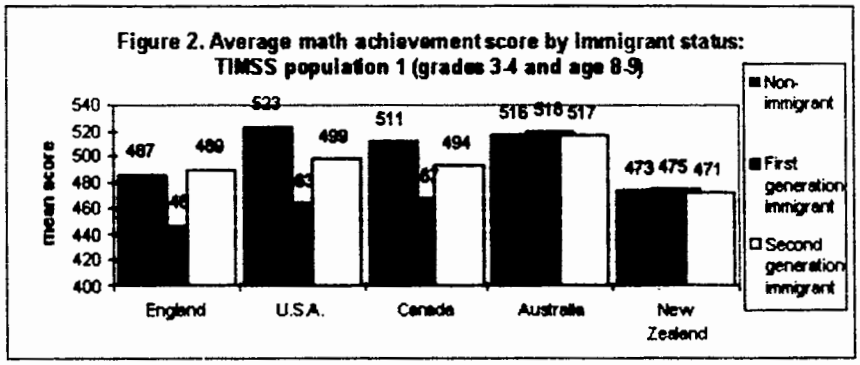
Figure 1. Percent of students with immigrant and non-English backgrunds: TIMSS population 1 (grades 3-4 and age 8-9).



Performance Gaps Associated with Immigrant Status

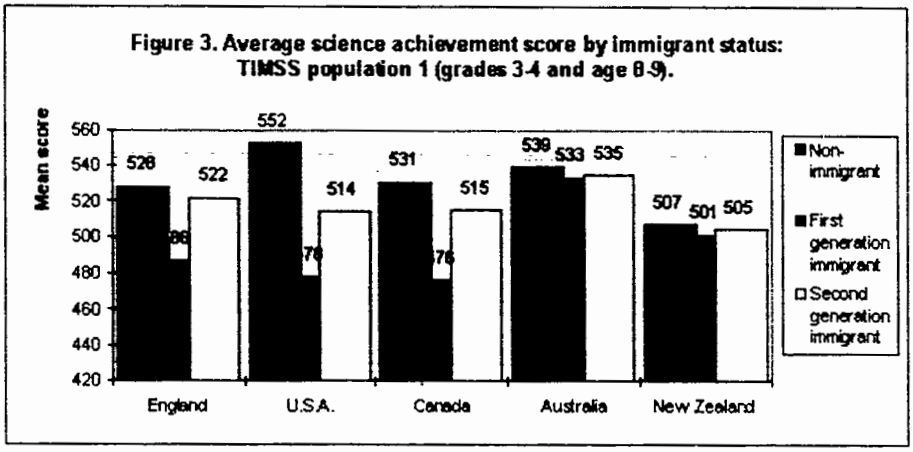
The math achievement gaps to the disadvantage of immigrant students took place only in England, the U.S., and Canada, not in Australia and New Zealand. This pattern is particularly evident in the gap between non-immigrant and the first generation immigrant children (Figure 2). In England, the gap in math score was 41 point, in the U.S., 60, and in Canada, 44, all statistically significant; whereas in Australia and New Zealand, the gap was not observed. In the U.S. and Canada, the non-immigrant children scored higher than the second-generation immigrant children; but in England, this difference was not statistically significant.

Figure 2. Average math achievement score by immigrant status: TIMSS population 1 (grades 3-4 and age 8-9)



The patterns of performance gaps associated with immigrant status were similar in science (Figure 3). In short, immigrant students lagged behind in math and science learning in England, the U.S., and Canada, but they did not in Australia and

New Zealand.



Immigrant and Language Background

Non-English home language is clearly a disadvantage to students' math and science learning regardless of immigrant status. In each group (non-immigrant and first and second generations of immigrant children), those whose home language was not English averaged substantially lower score than the rest of the students in math (Tables 1). Further, the language disadvantage was more acute among native-born children than among immigrant children. Consistent in each country, the second-generation immigrant children with non-English home languages did better in math than the non-immigrants with non-English home languages. I speculate that the latter was likely to be the indigenous groups or the groups that experienced persistent social and linguistic isolation, e.g., the American Indians and Hispanics in the U.S. Unfortunately, TIMSS contains no data to allow me confirm this assumption. With exception of the U.S. and Canada, this pattern holds between the first generation immigrant children and non-immigrants as well, though to a lesser extent.

Table 1
Average Math Achievement Scores by Immigrant Status and Home Language: TIMSS Population 1 (grades 3 or 4 and age 9 or 10) in the Five Countries

	England	U.S.A.	Canada	Australia	New Zealand
Non-immigrant	486.6	522.9	511.1	516.3	472.6
English	489.6	527.9	514.4	518.2	479.8
Non-English	419.2	472.2	470.5	436.2	407.7
First-generation Immigrant	446.2	462.9	466.5	517.9	474.7
English	452.4	475.5	476.2	534.8	491.9
Non-English	426.8	449.3	457.8	499.8	445.4
Second-generation immigrant	489.3	498.5	493.5	516.5	471.4
English	496.8	505.4	499.1	521.9	482.0
Non-English	464.9	486.5	477.9	489.7	427.4

Two-level Analysis

School level variance was substantial and statistically significant in all the five countries (Table 2). As indicated by the intraclass correlation coefficients, school-level variance proportional to the total variance around the given country's average math achievement ranged from 9 percent (Canada) to 26 percent (New Zealand). This finding suggests that to a considerable extent, students' math scores in each of the five countries tended to cluster around their school average scores. The reliability of the achievement measure was around 0.80, with exception of Canada, where the estimate was only 0.49. These baseline statistics justified further two-level modeling to examine the performance gaps relating to the language and immigrant

status.

Table 2
Two-level unconditional models: Baseline estimates from TIMSS Population 1 math achievement (plausible values average) in the five nations

Parameters	England	U.S.A.	Canada	Australia	New Zealand
Average school mean g_{00}	483.51	505.18	501.88	522.53	470.85
Reliability of the dependent variable	0.88	0.82	0.49	0.75	0.86
Intraclass correlation	0.19	0.18	0.09	0.16	0.26
School-level variance u_{0j}	1,700.63	1,755.81	1,349.94	1,864.95	2,402.65

Note: The school-level variance for each country was significant at $p< 0.001$ level.

Table 3 presents the estimates from the two-level random coefficient models. The first panel shows the fixed effects. The overall mean of each country (the intercept μ_{00}) provides a reference for interpreting the other estimates. First, non-English home languages were indeed a detrimental factor to children's math learning across the five nations. The large and negative coefficients consistently indicate that children with non-English home language background achieved lower than the overall mean in each country. The language barrier to math learning seems especially solid to students in England and the U.S.

The immigration status was a disadvantage only in some countries. Clearly, there was a negative relationship between the first generation of immigrants and the math achievement in England, the U.S., and Canada. But the relationship was reversed in Australia, where the first generation immigrant children achieved higher than the national average (a positive 13.4 at $p<.01$ level). There seems no relationship between the generation of immigration and achievement among New Zealanders as the two coefficients were small (2.42 and –5.46) and not statistically significant. The gap between the second generation of immigrants and the national average in general was narrower than that between their first generation counterpart and the national average. The second-generation children in England appeared to do slightly better than the national average (a higher score of 8.42 at $p<.05$).

The estimates for random effects revealed how the above statistics varied at the school level. The language-related achievement gap varied among schools only in England; in other countries, this gap was rather stable across schools. The math achievement gap related to the first generation immigrant status did not vary across schools in any of the five countries. This finding implies that the problem of this group (or its strength in Australia) in math learning was regular across schools. Finally, the gap associated with the second generation of immigration in the U.S. varied substantially across schools, indicating that schools probably might have some thing to do with this group's performance. This gap also varied across schools in New Zealand, despite that the fixed estimate for the effect was nil (not statistically significant). This irony probably hints that the second-generation immigrant children performed quite differently in New Zealand pending on school environment, although the *average* difference at student level was not observed.

Table 3
Two-level random coefficient models: Estimates for TIMSS Population 1 math achievement

Parameter	England	U.S.A.	Canada	Australia	New Zealand
Fixed effects: Student-Level Effects (Level-1 models)					
Intercept (overall mean achievement) μ_{00}	482.47***	504.95***	501.86***	522.46***	470.85***
Non-English language difference μ_{10}	-38.26***	-39.15***	-14.09***	-20.68***	-30.06***
First generation immigrant difference μ_{20}	-21,87***	-26.95***	-44.75***	13.64**	2.42
Second generation immigrant difference μ_{30}	8.42*	-4.25***	-12.21***	-7.61*	-5.46
Random effects: School-level variance (Level-2 models)					
School mean achievement, u_{0j}	1710.05***	1783.27***	1352.19***	1868.44***	2087.59***
Non-English language difference, u_{1j}	674.44*	—	—	—	—
First generation immigrant difference, u_{2j}	—	—	—	—	—
Second generation immigrant difference, u_{2j}	—	1076.18***	—	—	421.99**
Student level variance (Level-1 random effect), r_{ij}	7121.75	7387.03	13760.00	10017.03	5988.68

* p<.05; ** p<.01, *** p<.001
The symbol "—" indicates that the random variance was too small to model and thus the associated variable was specified only as a fixed effect in the model.
Note: All student-level predictor variables were centered on school means.

Summary

This analysis only touched on the surface of the immigrant children's academic learning in the five developed countries. It described the status of the group's math and science performance and help to settle the issue as to whether immigrant children achieve the same level as do non- immigrant children. In a cross-national comparison based on fairly comprehensive and reliable test information, the analysis indicated that in the U.S., England, and Canada, immigrant children--especially those known as the first generation of immigrants--did lag behind in math and science achievement. Further, non-English home languages, typically spoken by children of immigrants and indigenous people, were strongly and negatively related to lower math and science performance.

Considerable effort is needed to untangle the complicated issues surrounding the newcomers' schooling. For example, immigrants' socioeconomic status, family environments, gender role, and health conditions, could critically influence these children's math and science learning. Moreover, in academic subjects such as reading, writing, and social studies, where the language is either a pivotal tool of learning or simply the subject of study, we know even less about immigrant children's learning experience. Immigrant children's schooling and performance in those subject areas call for extended research.

The analysis also hints at the overall potential effect that schools might have in

reducing the performance gaps associated with the immigrant and non-English backgrounds. The first-generation immigrant children's disadvantage (in the U.S., England, and Canada) and the strength (in Australia) in math performance seem consistent across schools in a given country. Does this finding suggest that schools can make little difference regarding immigrant children's learning? Maybe. However, it may also imply the overwhelming effect of immigrant socio-cultural conditions on their schooling and, possibly, the public education systems' uniform indifference to the group's needs.

To an extent differentiated by the countries, performance gaps associated with the second-generation immigrants and non-English home language varied among schools. This finding implies that schools could possibly make some difference in narrowing the gaps. Learning about specific school factors that may work to close the gaps requires further research. School factors such as socio-demographic attributes, resource allocation, special programs, staff training, and curriculum and instruction methods are subject to study if we are to understand the learning processes of the increasingly large group of immigrant children in public schools.

References

Booth, A., Crouter, A. C., & Landale, N. (Eds.). (1997). *Immigration and the Family: Research and Policy on U.S. Immigrants.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods.* Newbury Park, CA: Sage.

Bryk, A. S., Raudenbush, S.W., & Congdon, R. T. (1996). *HLM: Hierarchical linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs.* Chicago, IL: Scientific Software International.

Duran, B. J.; Dugan, T. & Weffer, R. E. (1998). Language minority students in high school: the role of language in learning biology concepts. *Science Education, 82,* 3, 311-41.

Federal Interagency Forum on Child and Family Statistics. (1998). *America's Children: Key National Indicators of Well-being.* Website: http://childstats.gov/ac1998/ac98.htm.

Fix, M., Passel, J. S., Enchautegui, M. E. & Zimmermann, W. (1994). *Immigration and Immigrants: Setting the Record Straight.* Washington DC: Urban Institute.

Forgione, P.D. (1998). *What We've Learned From TIMSS About Science Education in the United States.* Washington, DC: National Center for Education Statistics. Website http://nces.ed.gov/Pressrelease/science/index.html.

Hao, L. & Bonstead-Bruns, M. (1998). Parent-child differences in educational expectations and the academic achievement of immigrant and native students. *Sociology of Education, 71,* 3, 175-198.

Hao, L. & Portes, A. (1998). *E Pluribus Unum:* Bilingualism and loss of language in the second generation. *Sociology of Education, 71,* 4, 269-294.

Hernandez , D. J. & Charney, E. (Eds.). (1998). *The Health and Well-being of Children in Immigrant Families.* Washington, D.C.: National Academy Press

Huang, G. G. & Weng, S. (1998). Minority post-secondary education attendance, high school desegregation and student characteristics. *Race, Ethnicity and Education, 1,* 2, 241-265.

International Study Center. (1998). *User's Guide for the Third International Mathematics and Science Study (TIMSS) and U.S. Augmented Data Files.* Boston,

MA: Boston College.

Lapin, L. (1994). Study: Immigrants' Kids Learn Better. *The Sacramento Bee*, February 23, 1994, p.1.

McPartland, J. (1998). *Project #7126: The Adaptation of Immigrant Children in the American Educational System Center for Research on the Education of Disadvantaged Students (CDS)*. Johns Hopkins University website http://www.csos.jhu.edu/default1.htm.

NCES (1998a). *The Condition of Education 1998*. Washington, DC: the author.

NCES (1998b). *Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context*. Washington, DC: the author.

NCES (1995). *The Condition of Education 1995* / Supplemental Table 40-2 Enrollment in public elementary and secondary schools, by race/ethnicity: 1976, 1984, 1986, 1988, 1990, and 1992 on website: http://nces.ed.gov/pubs/ce/c9540d02.html.

Oakes, J. (1990). Opportunities, Achievement. and Choice: Women and Minority Students in Science and Mathematics. *Review of Research in Education, 16*, 2, 153-166.

Rumbaut, R. G. (1996). The New Californians: Assessing the Educational Progress of Children of Immigrants. *CPS Brief*, 8, 3, 1-14.

Rumberger, R. W. & Larson, K. A. (1998). Toward explaining differences in educational achievement among Mexican American language-minority students. *Sociology of Education, 71*, 1, 69-93.

Shavarini, M. K. (1996). California's Immigrant Children, book review. *Harvard Educational Review, 66*,3, 668-674.

Schnaiberg, L. (1997). Panel urges greater focus on immigrant children's needs. *Education Week 17*, 29, p. 18

U.S. Bureau of Census. (1993). *We the American...Foreign Born*. Washington DC: the author.

U.S. Congress. (1980). *Science and Engineering Equal Opportunities Act*, Section 32(B), Part B of P.L. 96-516. Washington, DC: the author.

Valdles, G. (1998). The World outside and inside schools: language and immigrant children. *Educational Researcher, 27*, 6, 4-18.

Van, J. (1994). Immigrant Children Get Better Grades, Study Finds. *The Chicago Tribune*, February 23, 1994, p. 6.

Vernez, G. & Abrahamse, A. (1996). *How Immigrants Fare in U.S. Education*. Santa Monica, CA: RAND Center for Research on Immigration Policy.

Viadero, D. (1997). Immigrant children succeed despite barriers, report says. *Education Week*, 17, 29, p. 14, Appendix

Wainer, H. & Steinberg, L. S. (1992). Sex Differences in Performance on the Mathematics Section of the Scholastic Aptitude Test: A Bidirectional Validity Study. *Harvard Educational Review, 62*, 3, 323-336.

About the Author

G. Huang, Ph.D.

Senior Research Analyst
Synectics for Management Decisions, Inc.
North Moore St.
Arlington, VA 22209

Phone: 703.807.2324 703.528.2857
Email: garyh@smdi.com

Gary Huang is a research sociologist working in education and public policy related areas. His research interests include minority children's socialization and schooling, institutional effects on learning, rural education, and cross-cultural research and communication. He works mainly with large-scale survey data analyses but has a broad interest in other approaches to research. Supported by a grant from U.S. Department of Education, he is currently working on a cross-dataset synthetic analysis looking at the district-level resource and student-level academic achievement relationships.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

BEST COPY AVAILABLE

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz Garcia (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 254 times since June 5, 2000

prior vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 8 Number 26 June 5, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

From Manpower Supply to Economic Revival: Governance and Financing of Chinese Higher Education

Chengzhi Wang
University of Illinois at Urbana-Champaign

Abstract

With an introduction to the overall underdevelopment of higher education in China compared with the American counterpart, this article briefly examines the main trends of over two decades of development of the governance and financing systems of China's higher education sector. This article analyzes the resource allocation from governments and revenue generation in institutions under the reform policies of administrative decentralization and financing diversification. The new "Great Leap Forward" in higher education in 1999 and beyond, i.e., the radical and, to a certain extent, desperate mass higher education policy and practice of expanding enrollments in order to spur domestic consumption, is critically analyzed. By examining the ongoing institutional merging and "co-building" and the most recent enrollment expansion, the writer points out the economic significance for higher education of overcoming diseconomies of scale and inefficiencies. However, the long-range outcomes of the seemingly exciting investment in and consumption of mass higher education are difficult to predict.

Introduction

The significant issues such as reform, privatization, access, efficiency, equality, and equity are closely related to Chinese higher education administration and financing systems that are experiencing radical changes and restructuring[2]. (Note 2) In this article, I try to make a brief macro analysis of the case of Chinese higher education in the reform era from 1978 until the present primarily from the perspectives of governance and financing. From meeting modernization manpower requirements and producing technically qualified and politically correct human resources for about two decades, higher education in China now orients itself to stimulating investment and consumption, primarily on the demand side, in order to help the state revive the slumping economy.

First, I introduce Chinese higher education by comparing it with the well-known practice (e.g., long history, large scale, and high-level development) of American higher education. Second, I examine the main policy shifts of a more than two-decade development and general governance and financing operations in higher education. Third, I analyze the resource allocations of governments and revenue generation of institutions under reform policies of administrative decentralization and financing diversification. Fourth, I critically introduce and analyze the recent appearance of radical policy and practice to expand enrollment. Through stimulating the nationwide family investment and consumption of higher education, the state decision-makers hope that the move of mass higher education will help reinvigorate domestic consumption and help regain the state's sustained economic growth. In conclusion, by reflecting on institutional merging and "co-building" and the most recent radical enrollment expansion, I emphasize the economic ramifications of overcoming diseconomies of scale and inefficiencies of higher education for the development of Chinese economy. Meanwhile, I point out the results of the ongoing radical policies and practices of mass higher education remain very difficult to predict.

Overall Underdevelopment

For about two decades since the late 1970s, higher education in China has been experiencing tremendous changes and reforms. The reforms such as policy shifts toward decentralization of administration and diversification of financing have resulted in a great development in a number of fronts in the higher education sector. The rapid expansion in enrollments, reported to have increased to about 10 percent (Plafker, 1999) at the end of the century, was hailed as transition toward mass higher education (Hayhoe, 1993)[3]. (Note 3) However, compared with the general practice in the American higher education system[4], (Note 4) the first impression of the Chinese higher education system appears, among others, small in scale, short in history, and immature in development.

There were only 1,000 public regular colleges and universities in China, with a total enrollment of less than four million before 1999, which is the start of what I call the new "Great Leap Forward" in higher education when the enrollment ratio reached 10 percent. According to most recent Chinese official statistics, the number of these public institutions with an enrollment of 5,000 or more is less than one-seventh (CSSB, 1996, pp. 112-113). The average enrollment increased from 2,927 in 1996 to 3,112 in 1997 (CEY Editorial Board, 1998). Obviously, there exist diseconomies of scale in the higher education sector.

In terms of history, the first university (now Peking University) in the modern sense was established in 1898. After that, sociopolitical instability and turbulence in China in the first half of twentieth century largely precluded serious development of higher education[5]. (Note 5) After the founding of the People's Republic in 1949, the higher education sector, though it soon gained great development under strong influence of the Soviet model, was nearly abolished during the most radical years of Cultural Revolution (1966-1976) (Cleverly, 1985; Lofstedt, 1980). After 1978, the American

model of higher education was the one copied in China (Pepper, 1996).

Still in a stage of immature development, the higher education system in China is now more likely to be hyperpoliticized and ideologized even in the reform era. The typical examples are the nationwide compulsory three-month- to-one-year military education for students in colleges and universities in the years after the 1989 student movement and the alleged school-organized student demonstration after the NATO bombing of the Chinese Embassy in Belgrade in 1999. In addition, still struggling to grow out of the political control and command plan, higher education institutions are not well prepared for either the opportunities or the challenges of the free market. Besides, most institutions do not have clearly defined missions, performance-based management, or financing mechanisms. Few institutions have long-range institutional development goals. Internal and external inefficiencies and resource waste are still prevalent. Furthermore, after the policy of tuition and fees was applied in all public regular institution in 1996, effective and adequate financial aids from governments are generally unavailable, nor is the perfect market available where students and parents of poor families can obtain loans to invest in higher education. It is very difficult for students from poor families to obtain equal higher educational opportunities.

Compared with the fully developed American counterpart, higher education in China, to a certain extent, is still fumbling toward institutional autonomy, academic independence, and professional development. Chinese higher education institutions are making efforts to overcome inefficiencies, inequities, and underdevelopment (World Bank, 1997) through, for example, obtaining World Bank loans and following its recommendations. The new "Great Leap Forward" in the enrollment expansion in 1999 is the radical move that the policy decision-makers deem as a new way to develop higher education and, more importantly, to help revive the nation's economy (Note 6).

Development Trends

It is known that social and private benefits and monetary and non-monetary returns help drive the development of higher education (McMahon, 1974; Leslie & Brinkman, 1994). In addition, politicization of education has a special role in Chinese educational development, which is marked by hyper-politicization, politicization, and de-politicization at different periods of time (Sautman, 1991). Social and private benefits and monetary and non-monetary returns are also the driving forces for higher education development in China. In the reform of the 1980s, however, the state's manpower requirements for modernization and the pressure for international parity were among the immediate driving forces to expand, reform, and develop higher education.

Since the new state development policies of reform and "opening to outside world" were implemented in 1978, the Chinese government has placed top priority on education, in particular on higher education in order to produce urgently needed skills and talents for economic reform and national modernization. Two major measures were taken in the higher education sector to achieve these goals: enrollment enlargement and institutional multiplication.

The period between 1978 and 1985 witnessed a rapid growth in the number of enrollments and institutions (Table 1). Most of the growth in the number of institutions occurred between 1982 and 1985. The total number of institutions grew from 715 in 1982 to 1,016 in 1985 (Cheng, 1993, pp. 201-214). In 1985, the central government promulgated the "Resolution on Education Reform," which became the Education Act in 1996, initiating sweeping reform in all education sectors including higher education. In 1993, to speed up the reform and transformation from a planned economy to a market economy, the central government enacted new policy guidelines, namely "Guidelines of Chinese Educational Reform and Development." These new legislation and policies advocated decentralization of institutional administration and management, and diversification of educational financing while the central and upper level governments maintained managerial oversight and policy

regulation (Lewin et al., 1994).

Reforms in the higher education sector after 1985 featured a rapid increase in enrollments and with a growing effort to participate in market economy, rationalize specializations, and restructure curriculum and instruction, among others. But the total number of institutions did not increase significantly. In addition, the higher education sector has since been evidencing Westernization and globalization. The American model of a higher education system is gradually replacing the Soviet model for Chinese colleges and universities (Pepper, 1996).

Table 1
Development in Institutions and Enrollments, 1977-2000

Year	Institutions	FTE Enrollments[a] (In millions)	Annual Increase (In thousands)
2000	<1,020	>4.90	>331
1999	<1,020	4.50	>=331
1998	1,020	3.41	58
1997	1,020	3.35	167
1996	1,032	3.18	115
1995	1,064	3.05	120
1994	1,080	2.93	290
1993	1,065	2.64	360
1992	1,053	2.28	150
1991	1,064	2.13	-30
1990[b]	1,075	2.16	-20
1989	1,075	2.18	0
1988	1,075	2.18	100
1987	1,063	2.08	90
1986	1,054	1.99	200
1985	1,016	1.79	340
1984	902	1.45	140
1983	805	1.31	130
1982	715	1.18	-120
1981	704	1.30	130
1980	675	1.17	130
1979	633	1.04	173
1978	598	0.86	242
1977[c]	404	0.63	

Note. From Asian Times (1999); CEY Editorial Board (1997; 1998); Ministry of Education (MOE) Department of Development and Planning (1998); China State Statistic Bureau, Education Statistics Yearbook of China, 1992-1995; World Bank (1997); Zhao (1995).

[a] FTE Enrollments in associate, bachelor and graduate degree programs. Inconsistent statistics may be found in different official Chinese sources.

[b] 1990 and 1991 enrollments shrank from previous years probably because of the negative enrollment policy in response to the 1989 nationwide student movements.

[c] The Higher Education Entrance Examination System, which was abolished for

several years during the Cultural Revolution (1966-1976), was reinstated in 1977.

As reported in Table 1, public regular higher education institutions increased to 1,080 in 1994. In 1995, the number of institutions decreased to 1,054. In 1996, 1997, and 1998, the numbers of regular public colleges and universities are 1,032, 1,020, and 1,020, respectively. According to Zhao (1995), the ongoing remarkable trend of institutional merging and amalgamation and establishment of cross-institutional consortia has resulted in a decrease in the total number of institutions. The merging trend in Chinese higher education is in sharp contrast with difficulties in institutional merging in the United States. In 1997, 162 colleges and universities merged into 74 institutions (CEY Editorial Board, 1998). Zhao (1998) explored institutional merging and amalgamation as a remarkable aspect of restructuring Chinese higher education, but could not adequately explore this phenomenon. The merging and amalgamation actually were accompanied and facilitated by policies that upgraded institution's rankings in the higher education hierarchy and increased their share of resources. In addition, institutional merger and amalgamation were the only option other than closure for institutions owned by several central ministry-level departments that were cut off during Premier Zhu Rongji's bold governmental restructuring and downsizing in 1998. The merging is still going on, and I believe it will further reduce the numbers of colleges and universities.

In 1999, the central government decided to increase enrollment by 44 percent over the previous year (Liaowang News Weekly, 1999, p. 33), making the enrollment incidence as high as 10 percent for the first time in Chinese history. It was hoped that this radical enrollment expansion would satisfy the longstanding high demand for college education by families and students. More importantly, after many other attempts to revive the national economy proved unsatisfactory, decision makers hoped that the expected large-scale consumption and investment in higher education by households would stimulate domestic economic development (Plafker, 1999). Enrollments will continue to increase by 300,000 or more each year beyond 1999 according to the education authorities (Asian Times, 1999). Thus, the average unit cost in higher education is expected to be lower with the production of a larger volume of graduates, services, and research. Economies of scale in Chinese higher education sector are being sought.

Governance and Financing Systems

Higher education institutions are vertically administered and financed by one of the three types of administrative authority: (a) The MOE (Ministry of Education, which was renamed the SEC, State Education Commission in 1985, and renamed MOE in 1998), (b) the non-education ministry-level departments in the central government, and (c) provinces and province-level municipalities. The institutions of MOE and the central ministry-level governments are funded with budgetary allocations from the Ministry of Finance through MOE. Generally, the financial allocations are based simply on head-count enrollments, plus irregular, special-purpose funding. The provincial institutions are funded by the department of finance in each province and province-level municipality through MOE's provincial branches, plus irregular "encouraging" funding from the central government.

In 1995, there were 36 national "keypoint" universities funded through the SEC, with enrollments accounting for 11 percent of the total (Table 2). The average size was about 6,680 students. There were 331 ministry-funded institutions with enrollment taking 34 percent of the total. The average size was only about 2,100 students. There were 687 provincial and municipal institutions with enrollments of 55 percent of the total. The average size was about 1,600 students. In 1997, the average enrollment size of the three types of higher education institution grew to 3,112. All of the colleges and universities (except for a few recent amalgamated ones such as the Zhejiang University and the Sichuan Union University) are similar to very small U.S. colleges, according to American higher education enrollment numbers. But because of their diseconomies of scale, excessive high unit costs,

ineffective organization structures, mismanagement, high student subsidies, and limited revenue sources (Hartnett, 1993), Chinese colleges and universities lack the economic efficiency, academic vitality, professional development, affirmative action, and democratic participation apparent in colleges and universities in the U.S.A.

Table 2
Number and Enrollment in Regular Colleges and Universities, 1995
(Enrollment in 1,000)

	# Institutions	Undergrad. enrollm.	Short-cycle enrollm.	Total enrollm.	Undergrad. enrollm.	Total enrollm.
SEC/MOE	36	223	47	269	15%	11%
Central Ministries	331	629	328	956	41%	34%
Provincial or municipal authorities	687	666	907	1573	44%	55%
Totals	1054	1518	1282	2799	100%	100%

Note. From China State Statistics Bureau (1996, pp. 112-123) and World Bank (1997), with the author's modification.

Of the total enrollments in these public regular institutions, 52 percent were enrolled in degree-earning undergraduate studies, 44 percent in short-cycle (associate degree) programs, and 4 percent in postgraduate studies in 1995. These institutions employed 1.04 million staff, of whom 38 percent were faculty, 44 percent were administrative and supportive staff, and 18 percent were employed in organizations and companies affiliated with the institutions. Of the total faculty and staff, only 2 percent had a doctoral degree, 19 percent a master's degree, 49 percent a bachelor's degree, and 30 percent held short- cycled diplomas or equivalent educational attainment (World Bank, 1997, p. xiii).

In 1997, the total enrollments in colleges and universities reached 3.35 million. The higher education sector employed 1.0315 million staff, of whom 405,000 people, about 40 percent were faculty, all others were administrative and supportive staff, and employees in organizations and companies affiliated with the institutions (CEY Editorial Board, 1998). The number of faculty is slowly increasing while the number of administrative staff is decreasing. Despite the fact that student numbers in both regular public and adult higher education institutions were included, the officially published student-faculty ratio increased from only 8.91:1 in 1995 to 9.81:1 in 1997 (CEY Editorial Board, 1997, 1998).

It should be pointed out that some central ministries, for instance the Ministry of Finance and the Ministry of Foreign Economic Relations and Trade, are more powerful and richly funded than other ministry-level departments. Some provinces and municipalities, in particular those in the east and south coastal regions, are much more economically developed than those in the hinterland. Consequently, there exist inequalities in allocation of financial resources among institutions from the three types of authority.

In recent years, in order to mobilize resources to better manage and finance institutions and improve institutions' internal and external efficiencies, MOE has encouraged gongjian ("co-building") colleges and universities in collaboration with provincial and municipal governments and/or industry. Collaborations between MOE and other ministries, between MOE and provinces, between universities and corporations, and among different institutions have been increasing greatly in hopes of achieving better

management and financing of colleges and universities. In 1997, 100 universities had officially announced their "co-building" partners ranging from provincial governments, and central level ministries to corporations. In all, 228 colleges and universities had signed official collaborative contracts with "cooperators" and "partners" including provincial governments, central ministry-level departments, and other institutions. For instance, a total of 129 employers and organizations participated in the "co- building" of, or in cooperation with, the Inner Mongolian University in north China (CEY Editorial Board, 1998, pp. 155-180)

Resources Allocation and Generation

China has experienced sustained economic growth for about two decades in the reform era since 1978, with an impressive average growth rate of about 9 percent per year in real terms. In recent years, economic growth has slowed because of multiple reasons. Given domestic economic growth and the perceived international parity, spending on education in China is a mixed picture. Great progress has been achieved but there is great room to improve.

Because a market economy gradually replaced the rigid centralized planning, and localities and employers could retain much of their earnings without including them for taxes, the growth of government revenues fell far behind that of GDP, increasing at an annual average of only 2.6% (World Bank, 1997). However, government expenditures increased at 3.3% per year higher than revenue, resulting in budget deficits almost every year. Public expenditure on education increased by an annual average of 10 percent between 1978 and 1994, far exceeding the growth rates of the total government revenues and expenditures. Though overall public spending decreased over the years, public spending on education in proportion to total government spending rose from 6.2 percent in 1978 to 17 percent in 1994 (World Bank, 1997), and stayed about 16 percent during 1995-1997. Yet, public spending as a percentage of GDP rose from 2.1 percent in 1978 up to 3.1 percent in 1989, fell to 2.2 percent in 1994, and gradually fell to 2.47 percent in 1996, and then rose to 2.54 percent in 1997 (MOE Department of Development and Planning, 1998). This level of spending is very low in comparison with the average of 2.8 percent of least-developed countries, 4.1 percent of developing countries, and 5.3 percent of developed countries (UNESCO, 1995, pp. 2- 28). Some researchers have criticized this low level of public spending from international parity (Tsang, 1994). Spending on education as a percentage of GDP would probably be slightly larger if the community's support for education at village and township levels were taken into account. It is very hard to calculate the nationwide local and community contribution and investment in education in both physical and financial resources.

The public allocation to higher education grew by an annual average of 9.7 percent between 1978 and 1994. Public spending on higher education rose from 20 percent of the total expenditure on education in 1978 up to 29 percent in 1984, then fell to about 17 percent between 1989 and 1992, and rose to 19 percent in 1994. The budgeted public allocation accounted for 95.9 percent in 1978, 86.9 percent in 1990 and 81.8 percent of the total revenues in the higher education sector in 1992 (Table 3). Given a very low enrollment ratio in higher education, pubic spending in higher education was high in comparison with its Asian neighbor countries. Asian countries and regions including Japan, Korea, Malaysia, and Taiwan spend only 11 to 17 percent of total public education expenditures on higher education (World Bank, 1997). Unlike Japan, the United States, and many other countries, China has not sufficiently utilized private resources to support public higher education. Though booming in 1990s, private higher education in China is still under strict governmental control and scrutiny. The reasons for this practice stem from the government's political and ideological considerations, the profit orientations and the low quality of education in private colleges and universities.

In 1990, public spending per student in higher education was 193 percent of GDP per capita. Public spending per student in secondary education was 15 percent, and in primary education was five percent. In 1994, public spending in higher education was 175 percent, still considerably higher than the average of 98 percent in East

Asia (World Bank, 1997, pp. 41-42). In other countries in East Asia and the United States with mass higher education, the large sizes of enrollments and efficient utilization of resources result in the economies of scale and reduced unit costs.

Table 3
Financing Sources: Public Allocation from Governments and Revenue Generation in Institutions

Sources:	1978	1988	1990	1992
1. Total Budgeted Allocation	95.9%	87.7%	86.9%	81.8%
Recurrent Expenditure	74.8	64.9	65.3	61.4
Capital Expenditure	21.1	22.9	21.6	20.4
2. Total Institution-Generated Revenues	4.1	12.3	13.1	18.2
Total of 2.1 and 2.2	4.1	10.5	11.4	13.6
2.1 Revenues from institution funded activities		10.3	10.7	12.8
From institution-affiliated enterprises		2.8	3.1	3.7
From commissioned training		2.1	1.9	2.3
From education services		0.9	1.1	1.1
From commissioned research and consulting		1.0	1.2	1.3
From other funded activities		2.7	3.0	3.7
2.2 Donations and Gifts		0.2	0.7	0.8
2.3 Student tuition and fees		1.8	2.9	4.6
Total	100%	100%	100%	100%

Note. From Chen Liangkun (1994) in (World Bank, 1997), p. 46, with the writer's modification. The published data for most recent years are not available.

Under the centralized command plan system before the reform started, higher education institutions were exclusively financed through governmental appropriation according to budgetary planning. The previous year's allocation was used as basis for the next year's allocation, with possible incremental adjustment according to the situations of the institution and the whole sector. Unused funds, if any, had to be returned to governments by institutions at the end of the year. The centralized, tightly controlled budgetary system did not provide incentives and initiatives for efficient utilization of funds and institutional efficiency improvements.

The higher education financing system has been restructured through educational reforms. The major financing restructures include the following. First, along with decentralization in administration and management, decentralization in financing has been achieved. The central government delegated financing responsibilities to provinces and central ministries to finance institutions. Second, institutional autonomy and a simple formula-based approach (i.e., head-count of enrollment) were introduced in funding institutions. The institutions are given autonomy in spending money, and the governance authorities exercise the supervisory functions to hold institutions accountable in addition to overseeing their political correctness. The institutions are not required to return the unused funds at the end of the budgetary year. Third, financing is diversified in order to mobilize resources. The institutions are encouraged to generate and mobilize resources in any possible way.

As for diversification in financing, generally the following principal sources of financial resources have been tapped and expanded: (a) Institution-affiliated economies such as enterprises and companies, which accounted for 3.7 percent or more of total higher education revenue in 1992 and beyond. It is the largest share of the generated revenues (Table 3). (b) Commissioned training for companies, which

accounted for 2.3 percent of total higher education revenue in 1992. (c) Research and consulting services, which accounted for 1.3 percent of total revenue in 1992. (d) Donations and gifts, which accounted for only 0.8 percent of total revenue in 1992 compared with zero in 1978. (e) Tuition and fees, which account for an increasingly large portion of revenues since 1996, though published official statistics are unavailable.

Again, there exist different types of inequalities. Inequalities exist between institutions in cosmopolitan areas and small cities, between market-oriented and traditional departments, between liberal arts institutions and institutions of engineering and business, and between key institutions of large alumni and new local institutions with little bases for attracting donations and gifts. In addition, the enthusiastic pursuit of revenues in many institutions has resulted in the phenomenon of "running schools, running business," and negatively affected learning, teaching, and research (Kwong, 1997).

Special mention should be made of tuition and fees. Before 1978, college students paid no fees and were assigned jobs upon graduation. The 1985 education reform allowed institutions to admit students outside state plan but sponsored by enterprises or self-financed. Institutions have charged a low level of fees to students under the state plan since 1989. In 1992, students in the state plan were charged an annual tuition fee of 300-600 RMB, or $36-72 USD, and room and board of 100-200 RMB, or $12-24 USD. There are regional and sub-sector disparities in fee levels. In 1994, the distinction in fee level among students under the state plan, enterprise-financed students and self-financed students was abolished. In 1995, the tuition fees for students in most institutions were about 1,300 RMB, or $157 USD per student per academic year. Some institutions could charge more but were ordered not to exceed 2,700 RMB, or $324 USD (World Bank, 1997). Students in teachers' institutions were exempted from tuition fees because of the chronic shortage of teachers. In 1996, the MOE required all public regular institutions to charge tuition and fees. The MOE fixed the price of tuition in regular programs at 1,200 RMB, or $145 USD per student per academic year, with 10 percent adjustment by local higher education authorities based on local economic conditions (CEY Editorial Board, 1998). According to visiting professors from five Chinese universities at the University of Illinois at Urbana- Champaign that I have interviewed, tuition at their universities was in the range of 2,700-3,100 RMB per student in 1999-2000 academic year, far exceeding the MOE regulated prices.

Tuition and fees were the very important components of private participation in investment in higher education. However, sufficient and diverse financial aid, in particular the financial mechanisms to adequately take care of students from poor families, were not available. The poor would be denied higher education opportunities because of their inability to pay the growing tuition and fees. Because of the imperfect market, it is very difficult for the poor to borrow money to invest in higher education.

A new student loan program was launched by the MOE, the Ministry of Finance, and the People's Bank of China with the endorsement of the State Council (Guangming Daily, 1999a). It was reported that in September 1999, the Commercial Bank of China would provide loans to college students with the subsidy of five percent interest from the government. My interviews with visiting professors from the five universities revealed that this program had not been implemented at their universities in early spring 2000. They responded that a few banks under the encouragement of local governments did try to make loans to students from poor families, but in very small amounts, usually several hundreds of RMB. What was worse, banks required borrowers to pay the loans before their graduation for fear that lenders could not reach borrowers after their graduation.

New "Great Leap Forward" in Higher Education

According to the Chronicle of Higher Education, in July 1999, MOE officials and

the State Development Planning Commission announced that China's public regular colleges and universities would be allowed to enroll a total of 1.53 million new students, or 331,000 more than originally planned. The move started in 1999 was another attempt by the Chinese government to find new ways to revive the slumping economy. As pointed out, the perceived economic significance of family consumption and investment in higher education by the central authorities would help facilitate the pursuit of economies of scale in the higher education sector. But policy-makers' expectations to help reboot economic growth are the direct driving force for higher education to radically expand enrollments.

Calculating that the typical Chinese student spends some 10,000 RMB, or about $1,200 USD each year on tuition, housing, and expenses, it was expected the move would generate a wave of domestic consumption worth an estimated $400 million USD to the Chinese economy (Plafker, 1999). In 1999, 450,000 more university freshman students than the previous year were admitted than originally planned. This constitutes a 44 percent increase over the new enrollment in 1998 (Liaowang News Weekly, 1999, p.33). In addition, recruitment to adult higher education institutions increased by 100,000 above the previous year. Some regarded the new enrollments in the whole higher education sector as the largest increment since 1949 (China Youth Daily, 1999). The proportion of high school graduates going on to post- secondary education grew from 1.4 per cent in 1978 to 9 per cent in 1997. The figure in 1999 was about 10 per cent, which the government hoped to gradually increase to 15 per cent by 2010 (Plafker, 1999).

The China Education Daily (1999b) reported: "Enrollment in higher education will further increase next year, MOE has announced that higher education institutions will recruit 3 million freshmen in the year of 2000, an increase of nearly 10 percent over the 2.8 million admitted in 1999." The numbers of new enrollments in 1999, including the new enrollments of regular public, adult and private higher education institutions, are probably larger than previously thought. Many cities and provinces made their own enrollment expansion plans. For instance, Shanghai has planned to enlarge access to higher education and to raise the gross enrollment to 40 percent of the age cohort (China Education Daily, 1999a), an unprecedented higher education enrollment ratio in Chinese history.

In my interviews, visiting professors from Chinese universities expressed unanimously that their universities enrolled more students than expected. Presidents of colleges and universities, professors, as well as students and parents, were excited about the news of enrollment expansion. But as higher enrollment quotas were assigned to each institution, presidents and professors knew there would be difficulties in absorbing the unexpected increase. One professor from a university in north China said that, to his knowledge, in the provincial enrollment meeting with the governor and education officials in late summer 1999, presidents had to agree to enroll the given quota before the conference could be dismissed.

An MOE official explained that the effect of the increase on the economy is three-fold. First, the enrollment of more students in universities creates a demand for more buildings and equipment, which, in turn, will stimulate the development of some relevant sectors of the economy, such as construction and service industries. Second, there is a shift of over 300,000 high school students to tertiary education institutions each year (in the expansion). This will relieve pressure on the employment sector (by over 300,000 positions) at least for the next three or four years. Third, household money savings will flow out of the banks as more university students pay their tuition fees (Asian Times 1999). Obviously, the expanding enrollments is intended to immediately stimulate consumption and reinvigorate domestic demand.

Many questions arise about the radical enrollment expansion. First and foremost, is there any significant empirical evidence to support the hypothesis that radical enrollment expansion will stimulate economic growth? After careful studies by Professor Wei Xin (1999) and his research group at Peking University, conservative

answers were provided. On the side of supply of higher education, regular higher education institutions do not have the potential for expansion to the degree that policy-makers assumed. Nevertheless, it is almost impossible for the private institutions to expand enrollment under the current strict control of state regulations and rules. On the side of demand, the ability of the general public to pay tuition and fees is questionable. The total number of household bank savings in China with a population of about 1.2 billion reached 5,300 billion RMB, or about $640 billion USD at the end of 1998. However, the money was not equally distributed among households. The richest 20 percent of the households owned over half of the total household income. The Gini coefficient in China increased from 0.288 in 1995 to 0.388 in 1998, and over 0.400 in 1999. What is more important, it is difficult to expand the capital infrastructure of colleges and universities. If one million more students are admitted each year and if the MOE institution infrastructure standards are followed, a total of 100-300 billion RMB will be needed to invest in infrastructure construction within the four-year cycle. Currently, it is almost impossible for the governments to make such a huge investment. If this financial burden is transmitted to students and families through rising tuition and fees, higher education then becomes even more unaffordable for the low-income majority.

Second, what about the quality of education after colleges and universities expand their enrollments, some even beyond their capacities? The visiting professors from China that I interviewed expressed their concerns by comparing their own tutoring experiences and the educational achievements of their students before and after the enrollment expansion. Education authorities also worry about the deteriorating quality of education. According to China Education Daily (1999c), the Department of Higher Education of the MOE has issued a directive to require colleges and universities to ensure the quality of teaching and learning after the expansion of enrollments in 1999. To improve teaching and learning is a challenge for all institutions. For instance, specialized colleges normally offer 2-3 year certificate courses. But with the expansion of higher education in 1999, many 2-3-year colleges that are allowed to offer certificate courses are also providing bachelors degree courses. Guangming Daily (1999b) warned that this trend would threaten the quality of education.

Third, what about employment after four years of education? The National Coordination Workshop for Employment of University Graduates 1999 stated that the employment situation was not satisfactory in some ways because of the aftermath of the Asian financial crises and downsizing of governments and state-owned enterprises. MOE urged the relevant government agencies to offer opportunities to new graduates and it also asked universities to encourage students to enter non-government organizations and self- employment enterprises (Southern Daily, May 23, 1999). After three or four years, when the graduates are ready for employment, can the unemployment pressure be relieved? Can the economy recover and labor markets be reinvigorated to take in the large number of college graduates? Without other cautious and compatible prevention measures, it is possible for Chinese university graduates to repeat the unemployment or underemployment experienced of higher education graduates in some developing countries such as Sri Lanka and India.

Conclusion

Large numbers of small institutions are one characteristic of the Chinese higher education system for over two decades. In addition, Chinese higher education has relatively low internal and external efficiencies. The low efficiencies are typically represented by the under-utilization of personnel and physical resources, and over-specialization and rigidity in instructional programs. Rationalization of specializations and units within the institution, joint production of neighboring institutions, institutional merger or consolidation, and increasing the size of institutions are the four ways for Chinese higher education to help overcome diseconomies of scale (Tsang & Min, 1993).

Fortunately, recent trends and practices evidence the following: curb the institutional

multiplication, encourage merger and amalgamation and "co-building," increase enrollments without growth of institution numbers, rationalize institutional programs and management, and other types of reform measures. These trends and practices are aimed at achieving economies of scale and efficiencies of higher education.

The new "Great Leap Forward" in higher education expansion in 1999 and beyond, on the demand side, satisfies families' strong desire for higher education for their children, and, indeed, stimulates household consumption of and investment in higher education in the short run. Yet, such a radical move also brings questions and concerns about its impact on student achievement and the quality of education, on graduates' employment, and on economic growth in the long run. Chinese political and educational authorities should look to both international experiences and domestic educational and socioeconomic realities in implementing the new "Great Leap Forward" policies, before it is too late.

Notes

1. I wish to acknowledge helpful comments from Professor King Alexander of the Department of Educational Organization and Leadership at the University of Illinois at Urbana- Champaign, who carefully read the first draft. I wish to thank the EPAA Editor and anonymous referees for their helpful advice and comments. In this article, I concentrate my discussion and analysis on mainstream higher education in China, i.e., regular public higher education. Adult higher education and private higher education are two other types of higher education. The former is part-time, aimed at upgrading educational attainment of workers, teachers, and other groups in the workforce who wish to seek higher education without interrupting their employment. The latter appeared after the education reform that was officially initiated in 1985. Though many applauded the appearance and quick expansion of private education, only 20 private colleges and universities had been accredited by the central educational authorities as of 1997 (Zhang, 1997). In 2000, there are only 37 non-governmental private colleges and universities that are authorized to issue associate degrees (China Youth Daily, 2000). The development of private higher education cannot maintain its momentum. The major reason, perhaps, is the lack of governmental subsidies, which leads to institutional autonomy and independence but, meanwhile, hinders the communication and cooperation between the policy decision-makers and the private institutions. Furthermore, the lack of governmental subsidies leads the private institutions to seek quick investment returns at the expense of satisfactory and healthy institutional growth.
2. For these issues, see, for example, K. Lewin, A. Little, H. Xu, and J. Zheng (1994), J. Henze (1984), pp.93- pp153, M. Tsang & W. Min (1992), and World Bank (1991; 1996; 1997).
3. Hayhoe (1993) predicted that the higher education enrollment rate in China would reach 10 percent at the end of the century. From what was reported by Plafker (1999), Hayhoe was correct in her prediction. Plafker reported that the total number of higher education institutions was 1,032 in 1999. Actually, that was the number of institutions in 1996. In 1999, the number must have been smaller because of increasing institutional mergers.
4. For the American higher education system and financing policy shifts, see, for example, M. Mumpher (1996) and P. M. Callan, and Finney, J. F. (1997).
5. It should be noted that mission colleges and universities, of which many were established by American missionaries, experienced most impressive progress and development between 1910-1937 (Deng, 1997, pp. 67-90). These mission institutions meanwhile also stimulated, directly or indirectly, the development of Chinese national colleges and universities before 1949.
6. For the "Great Leap Forward" in education, the hyperpoliticized, frenetic, radical, and unrealistic education expansion movement in 1958, see, for example, J. Kwong (1979).

References

Asian Times (1999). August 2-8, 1999.

Callan, P. M. & Finney, J. E. (Eds., 1997). *Public and private financing of higher education: Shaping public policy for the future*. Phoenix, AZ: Oryx Press.

CEY Editorial Board (1998). *China education yearbook 1998*. Beijing: People's Education Press.

Cheng, K. M. (1993). *Zhongguo dalu jiaoyu Shikuang* (The Educational Realities in Mainland China). Taibei: Taiwan Commercial Press.

China Education Daily (1999a). September 7.

China Education Daily (1999b). September 24

China Education Daily (1999c). December 10.

China Youth Daily (1999). June 25.

China Youth Daily (2000). March 8.

Cleverley, J. (1985). *The schooling of China*. Winchester, MA: George Allen & Unwin.

CSSB (China State Statistic Bureau), Education statistics yearbook of China, 1992-1995. Beijing: China Statistics Press.

CSSB (China State Statistics Bureau, 1996). The abstracts of China statistics yearbook 1996. Beijing: China Statistics Press.

Deng, P. (1997). *Private education in modern China*. Westport, CT: Praeger.

Guangming Daily (1999a). June 29.

Guangming Daily (1999b). October 21.

Hayhoe, R. (1993). An Asian multiversity? Comparative reflections on the transition to mass higher education, *Comparative Education Review, 39*(3), 17-32.

Hartnett, R. A. (1993). Higher education funding in open door China, in P. Altbach & D. B. Johnstone (Eds.), *The Funding of Higher Education: International Perspective* (127-149). New York: Garland.

Henze, J. (1984). Higher education: The tension between quality and equality, in R. Hayhoe (Ed.), *Contemporary Chinese education*. Kent, England: Croom Helm.

Kwong, J. (1979), The educational experiment of the Great Leap Forward, 1958-1959: Its inherent contradictions. *Comparative Education Review* (23, October), 443- 455.

Kwong, J. (1997), The New Educational Mandate in China: Running Schools Running Business, *International Journal of Educational Development, 16*(2), 185-194.

Leslie, L. L. & Brinkman, P. T. (1994). *The Economic value of higher education*. Phoenix, AZ: Oryx Press.

Lewin, K., et al. (1994). *Educational innovation in China: Tracing the impact of the 1985 education reform*. Harlow, UK: Longman.

Liaowang News Weekly (1999), August 16.

Lofstedt, J. (1980). *Chinese educational policy: Changes and contradictions 1949-1979*. Stockholm: Almqvist & Wiksell International.

McMahon, W. W. (1974). *Investment in Higher Education*. Lexington, MA: Lexington Books.

MOE Department of Development and Planning, *Educational Statistical Yearbook of China* 1997, 1998. Beijing: People's Education Press.

Mumpher, M. (1996). *Removing College Price Barriers: What Government Has Done and Why It Hasn't Worked*. Albany, NY: SUNY Press.

Plafker, T. (1999). China Increases University Enrollments, Hoping Student Spending Will Revive Economy, *The Chronicle of Higher Education*, September 3, 1999.

Sautman, B. (1991). Politicization, hyperpoliticization, and depoliticization of Chinese education, *Comparative Education Review*, November 1991, 669-689.

Pepper, S. (1996). *Radicalism and education reform in 20th- century China*. Cambridge: Cambridge University Press.

Plafker, T. (1999). China increases university enrollments, hoping student spending will revive economy, *The Chronicle of Higher Education*, September 3.

Southern Daily (1999). May 23.

Tsang, M. (1994). Costs of education in China: Issues of resource mobilization, equality, equity, and efficiency, *Education Economics, 2*(3), 287-312.

Tsang, M. C. & Min, W. (1993). Expansion, efficiency, and economies of scale of higher education in China. *Higher Education Policy, 5*(2), 61-66.

UNESCO (1995). *Statistical Yearbook*.

Wei, Xin (1999). *On contributions of scale expansion of higher education to short-run economic growth*. Available at: http://www.hedu.pku.edu.cn/kuozhao/kzhkt.htm .

World Bank (1991). *China: Provincial education planning and finance study*. Washington D. C.: the Writer.

World Bank (1996). *China: Management and finance of higher education*. Washington D. C.: the Writer.

World Bank (1997). *China higher education reform*. Washington D. C.: the Writer.

Zhang, D. C. (1997). *A brief introduction to China Minban higher education institutions*. Beijing: China Agricultural Science and Technology Press.

Zhao, F. (1998). A remarkable move of restructuring: Chinese higher education. *Education Policy Analysis Archives, 6*(5). Available at: http://epaa.asu.edu/epaa/v6n5.html.

About the Author

Chengzhi Wang
218 Coble Hall
801 South Wright Street

Champaign, IL 61820

Email: cwang2@uiuc.edu

Chengzhi Wang is a Ph.D. candidate in Comparative Education and Social Sciences in the Department of Educational Policy Studies and a research assistant with the Graduate College at the University of Illinois at Urbana-Champaign.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 555 times since June 16, 2000

Education Policy Analysis Archives

Volume 8 Number 27 June 16, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

The Challenges to Distance Education in an Academic Social Science Discipline: The Case of Political Science

Steffen Schmidt
Iowa State University

Mack C. Shelley
Iowa State University

Monty Van Wart
Texas Tech University

with

Jane Clayton
Iowa State University

Erin Schreck
The Eaton Company

Abstract

This article reports the results from a national survey directed to the department chairs of political science to assess the current and future state of distance learning in that discipline. The insights of this research are relevant to all social science fields and offer important insights to other academic disciplines as well. Key findings of the study include the low utilization of distance learning courses, a low degree of importance currently attributed to distance learning and modest expectations of future growth, ambivalent acceptance of a

future role for distance learning, the common use of Internet-related technologies, low levels of faculty knowledge and interest about distance learning, limited institutional support, and serious doubts about the appropriateness and quality of instruction at a distance. We propose a model of the size and scope of distance learning as a function of three factors: the capacity of distance learning technologies, market demand, and faculty and university interest in distance learning. The article concludes with suggestions of critical areas for future research in this dynamic, fluid post-secondary environment.

Introduction

On March 26, 1999, at 6:29 a.m., CNN ran an advertisement for UCLA's distance learning program. It was the first full-blown, national commercial inviting students from around the world to ignore their local, physically accessible college or university and to opt instead for accredited courses taken at a distance. This was an important symbolic event because it promoted the "third way" of delivering higher education with a seriousness that has not been seen before in the United States. The first way is to have students travel to a college or university and live in residence, no matter whether the distance they traverse is near or from the other side of the world. Generally such students are full-time. The second way is to provide classes for students who commute from local or not-so-local areas. Such students are more likely to be part-time. The third way is to provide education at a distance, which was pioneered in correspondence courses and later in public television classes (McIsaac & Gunawardena, 1996).

The third way has long been characterized by a tiny share of the student audience, thought to have less serious students, and subject to criticisms about inferior quality (Jaffee, 1997, Noble, online; see Rahm, Reed, & Rydell, 1999 for a good review of the challenges). In reviewing the literature on distance learning, one quickly discovers both hyperbole and deep skepticism (Schmidt, 1999). Advances in technologies, new economic forces, and a changing university environment certainly require a reexamination of many of the old assumptions about distance learning (Mingus, 1999). Joseph Hardin and John Ziebarth, at the National Center for Supercomputing Applications, publishing in *The Future of Networking Technologies for Learning,* suggest that "...very soon every teacher and student will need access to the information represented on the Web in order to be competitive in their work and in their lives" (Hardin & Ziebarth). Further, some experts (for example, the Pew Higher Education Roundtable) suggest that 30 to 50% of all post-secondary learning will take place through some form of distance learning.

Yet others suggest—including substantial numbers of faculty—that this is a passing fad suitable for only a narrow niche of courses, and that traditional settings will remain the overwhelming method of education (Clark, 1993). The most optimistic predictions of advocates who watched the rapid transfiguration of the communication world by the Internet are likely excessive in both quantity and speed of any market transformation. However, distance learning seems unlikely to be a mere instructional fad. Examples of the seriousness of the phenomenon are not difficult to find.

One of the most impressive manifestations of distance learning is the establishment of the new virtual universities. By far the most successful major distance education institution is the British Open University, which has granted 227,000 degrees (Blumenstyk, 1999) since 1971and has an excellent reputation despite Great Britain's conservative educational tradition. American experiences are still mixed. Although small, Jones International University has gained accreditation (Olsen, 1999a). Some of the virtual universities are up and running moderately well, such as the Southern Regional Electronic Campus. For most it is too early to tell, such as the Western Governor's University (WGU, the Colorado Community College Virtual University, Penn State's World Campus, and the United States Open University. For all the news and hyperbole of WGU and California Virtual University, they have underachieved initial expectations (Newcombe, 1999) and the California Virtual University had its plug pulled in 1999. Yet this is not stopping new, well-funded entrants such as Kentucky Commonwealth Virtual University

(Young, 1999) and Michigan Virtual University. These huge education syndicates indicate a willingness to devote the considerable resources needed to provide the substantial retooling in technology, systems, and personnel that is necessary for large-scale success.

In the summer of 1999 a new virtual university consortium named Cardean University (www.unext.com) was launched partly with financing from former junk bond king Michael Milken. It will offer complete graduate programs. What's important about this venture are the five prestigious universities who are part of the venture –the University of Chicago, Columbia University, Carnegie Mellon University, Stanford University, and the London School of Economics and Political Science. This project looks more promising than some given the high- octane nature of the participating institutions.

Perhaps as important is the adoption of distance learning technologies by prestigious universities (Newcombe, 1999). Stanford offers a full engineering degree and Duke offers a full MBA on-line (which integrates occasional live sessions as do many quality distance programs). Examples of fully on-line classes now exist at Oxford and Harvard. The question of broad-scale penetration of distance learning in higher education is less an issue now. Rather, the question now focuses on how much penetration, in what specific areas such as political science, and how it can be done most effectively.

Commercial examples, while different in nature, give evidence of the liabilities of adopting a wait-and-see attitude toward new technologies. Faculty have seen the college textbook market dramatically transformed by newcomers such as Amazon.com, VarsityBooks.com and, more recently, Bigwords.com. Traditional textbook wholesalers such as textbooks.com (Barnes and Noble), efollett.com, and ecampus.com (Wallace) have scrambled to get on-line (Kiernan, 1999). The effect of electronic commerce has been devastating for both university-owned and locally owned stores. The local university bookseller in Ames, Iowa, reported a 30% drop in sales as the result of a full-page ad that appeared in many targeted college student newspapers and through the use of handbills on campus. University- owned and locally-owned bookstores are beginning to combat this trend in different ways. One strategy is a buying consortium with a centralized on-line access point (Carr, 1999). Another strategy is for the university to turn book sales entirely over to an on-line provider such as VarsityBooks.com. The online provider then pays the institution a percentage of the sales and the bookstore ceases to sell textbooks (Olsen, 1999b). Although this commercial analogy should be applied to complex, degree-granting institutions of higher education with extreme caution, it is interesting to ponder whether there could be a similar critical-mass shift in higher education distance education as well. One important point of difference currently is that quality distance education programs are not less expensive in tuition than conventional programs, and frequently are more costly (Blumenstyk, 1999). This situation may shift in the next few years with technology advancements and increasing faculty experience.

Research Issues in Distance Learning: An Overview of This Article

Many issues have arisen regarding the proper role and effect of distance learning: the globalization of the competition for students among institutions of higher education, the pressures for cost-cutting and cost effectiveness in the new economy, the challenge to traditional institutions of higher education posed by virtual universities and by the growth of for-profit universities, concerns among faculty about job security and the implications for promotion and tenure as well as reward structures, concerns about the content quality of distance learning, and a series of technical issues such as intellectual copyrights, accreditation, transferability of credits across institutions, and the integrity of undergraduate and graduate programs of study. Some of these issues are being addressed at a general level in journals such as *The American Journal of Distance Education, Distance Education, ED Journal*, the *Journal of Classroom Technology*, *Kairos*, and *Training and Development*. Yet we would argue that these big and interesting questions can be understood best by examining where disciplines such as political science presently stand. This study offers an empirical assessment of the current scope of, as well as several of the major contributing factors to, the role played by distance learning in

higher education generally and more specifically in political science.

To help make sense of the contemporary changes occurring in distance learning, we begin by briefly proposing a theoretical construct for the factors affecting the growth of distance learning. This exploratory study provides an empirical baseline for some—but not all—of the array of factors relevant to a more exhaustive understanding of distance learning.

First, what is the scope of distance learning in political science curricula? The answers to several more specific questions of the scope of distance learning are addressed in our results. How frequently are distance learning classes offered? What percentage of credit hours are attributable to distance learning classes? What is the level at which distance learning is used? What are the perceptions of department chairs (thus indirectly of departments) on the importance and/or faddishness of distance learning?

Second, we address the types of technologies that have been implemented to deliver distance learning classes in political science. Are generational differences among faculty cohorts a major consideration in what methods have been and are being adopted? Do the faculty members participating in distance learning courses make full use of newly available Internet-based technologies? How many relevant distance learning technologies are used on average by actively engaged instructional faculty? What does the future hold in store for faculty abilities to adjust to rapidly evolving new technologies?

Third, what is the profile of political science faculty knowledge about, their interest in, and the incentives for providing distance learning? How much do faculty understand the new technologies, what interest do they have in learning more about it, and how much support is available for the opportunity to experiment with the new technologies? What are the characteristics of the faculty members who are engaged in distance learning? What is the nature of faculty perceptions about the quality of distance learning? What is the appropriateness of distance learning to the political science arena? How do such methods compare to traditional methods? Finally, in the estimation of faculty, what is the overall effect of distance learning likely to be on students, departments, universities, and ultimately, themselves?

After reporting and interpreting the findings, this article suggests critical areas for future research in this dynamic environment.

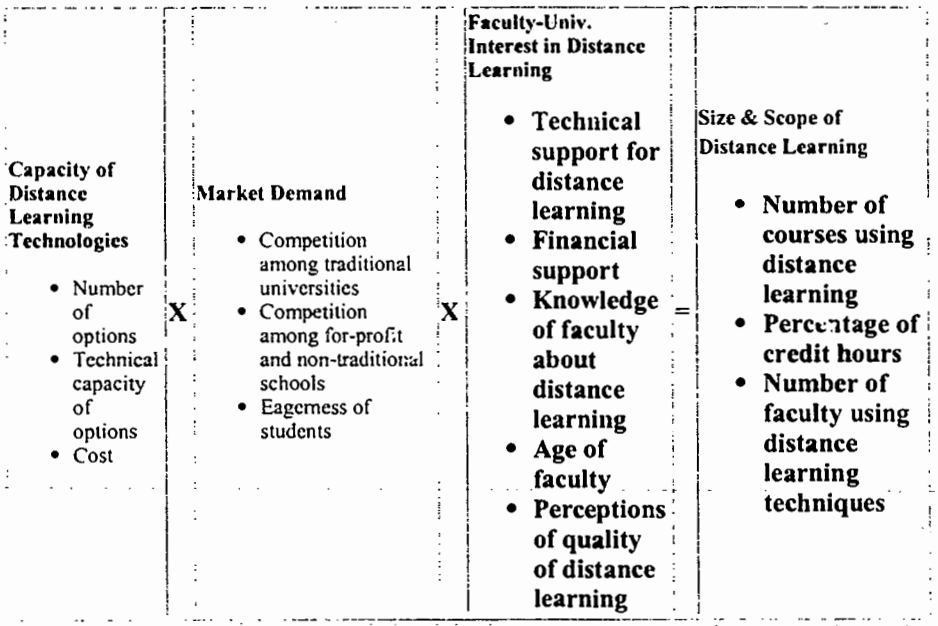
Major Factors Affecting the Growth of Distance Learning

The size and scope of distance learning is affected by three major domains (for an excellent overview of these and other issues in the higher education context, see Boaz et al., 1999). First, it is affected by the *capacity of the distance learning technologies*. If the capacity is relatively weak, the size and scope will be more limited. The sheer number of distance learning options is important. A greater number of options means that distance learning provides a greater array of opportunities and also allows for a greater degree of synergy among those options. For example, Web-based classes normally are enhanced significantly by using email for individual student- instructor conferences and regular mail for textbooks and proprietary materials that cannot be scanned and sent electronically. Another important factor is the technical capacity of each of the options. Clearly the rapid expansion of Internet-related technologies will have a considerable effect on the long-term growth capacity of distance learning. A related factor is the cost of different technologies. Falling or increasing costs dramatically affect the willingness of individuals and institutions to experiment with and to institutionalize distance learning options.

A second important domain is *market demand*. How eager are students for distance learning options? Which students, and how many students, are interested in distance learning exclusively, and which students are interested in distance learning for selective purposes? Another important aspect is the competition among the universities themselves. If universities fail to provide many options, and those options are limited in scope and quality, then distance learning will remain a small part of the market. However, even if only a few universities provide strong national and regional options, they can stimulate great competition because of their ability to penetrate distant markets at little or no additional cost.

A third domain is the *level of faculty/department/university interest* (Brigham, 1992). The level of technical support will affect the scope of distance learning. So, too, will the incentives used to encourage departments and individual faculty members. An indication of the attitudinal barriers and institutional constraints confronting successful implementation of distance learning is provided by the results of a 1998 survey of professors by the American Association for History and Computing (on-line, 1998, Trinkle, 1999). The evaluation by 65% of the respondents was that their institution's technology policies were misguided or insufficient. Of course, the knowledge of faculty about distance learning options also is critical. We believe that the generational age of faculty members also will have an effect, since older faculty members typically are less apt to adopt new technologies and to change their teaching styles radically, as distance learning often requires. Finally, the perceptions of faculty members (and their institutional units) about the quality of distance learning are crucial as well. For example, if large or important groups of faculty feel that distance learning is fundamentally inferior and if they thereby largely ignore such options altogether, then distance learning is likely to have a slow, tough path even if technical capacity (such as bandwidth) grows dramatically. See Figure 1 for a graphic representation of these relations.

Figure 1: Factors Determining the Size and Scope of Distance Learning



Research Methods and Results

The Survey Instrument

In the fall of 1998 a national survey instrument with 21 questions was designed and field-tested to explore the extent and perceptions of distance learning in political science departments in colleges and universities throughout the United States. Following appropriate adjustments, the survey was mailed to 812 political science departments representing both undergraduate and graduate education programs in the United States. A total of 296 useable questionnaires were returned, for an overall response rate of 36%; the functional response rate for certain questions was less because of their nonapplicability to portions of the respondents. The questionnaires were sent to chairs of departments since it was felt that they would have the best overview from which to answer the questions posed. We speculate that responders would be slightly more active in distance learning on average than nonresponders. Thus, it seems likely that to the degree that there is any respondent distortion in our findings, it would exaggerate the results, leading us to report in this study that there was slightly more activity in distance learning than there is in fact.

Respondent Characteristics

Although only three-quarters of the respondents completed the requested demographic data, the characteristics of the respondents seem to reflect the breadth of the field of political science, with the bulk of the respondents coming from institutions with enrollments under 10,000 and from departments having 10 or fewer faculty members. See Table 1 for a breakdown of respondents by size of student body and political science faculty.

Table 1
Characteristics of Universities and Colleges Surveyed

University		Department	
Student Body Size	**%**	**Faculty Size**	**%**
Under 5,000	44.9	2-6	43.3
5,000-10,000	20.7	7-10	20.7
10,000-15,000	12.9	11-15	14.2
15,000-20,000	10.7	16-25	16.9
Over 20,000	10.7	over 25	4.7

Findings

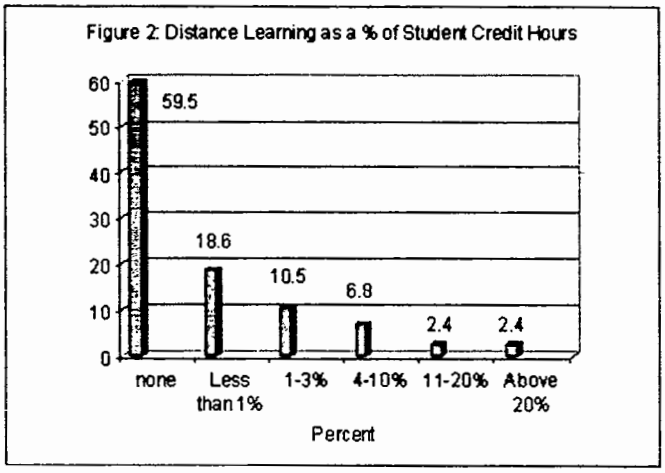
Size and Scope of Distance Learning

Perhaps the single most important set of data was captured in Table 2, which summarizes responses to the question: "Does your department use distance learning technology for any of its courses?" Note that the broad wording allowed some respondents to include classes that were primarily face-to-face but that use supporting distance learning technologies. (Note 1) Nonetheless, a substantial 57.5% of the responding departments do *not* use distance learning technology for *any* of their courses. (Note 2) One-third reported using some distance learning in one to three classes. Approximately 10% reported the use of distance learning in 4 or more classes.

Table 2
Use of Distance Learning in Political Science

Degree of Usage	%
None	57.5
1-3 classes	32.0
4-8 classes	7.1
More than 10 classes	3.4

A related way of examining the scope of distance learning is to assess it as a proportion of the department's full credit-hour usage. When responding to the question "Approximately what percentage of your students' credit hours are distance learning this semester?" fewer than 5% of the reporting departments indicated that 10% or more of the department's total credit hours were generated by distance learning. Only 22.1% of departments reported the level of distance learning usage at 1% or more of student credit hours. See Figure 2 for the breakdown of distance learning usage by credit hours. Clearly the number of institutions that are completely uninvolved is very high among respondents, and it is likely that the nonresponding members of the surveyed population have an even lower proportion of distance learning utilization. Further, of those institutions that do utilize distance learning technologies, the number that make extensive use of them is very small.



Although the usage of distance learning may be relatively limited, in what part of the political science curriculum is that use most common—in undergraduate, graduate, or training courses? Respondents could choose multiple answers; thus the sum of percentages across all response categories may be greater than 100%. In the programs reporting the use of distance learning technology the bulk of such utilization is concentrated in undergraduate classes. At this level, utilization is split fairly evenly between lower- and upper-division undergraduate courses (in 58.4% and 66.4% of responding departments, respectively). Departments engaged in distance learning identified graduate classes 32.8% of time, and training programs were selected by only 6.4% of the responding departments.

Several questions surveyed the degree to which the department chairs thought that distance learning was an important component of their department's curricular offerings. These findings reflect not only the relatively low utilization rates, but also perceptions about a low level of importance attributed to distance learning at this time. Three-quarters of the respondents *strongly* disagreed that distance learning was a major component of their curricula, and only 8.8% moderately or strong agreed that it was. See Table 3 for results.(Note 3)

Table 3
Perceptions About Distance Learning as a Major Curriculum Component

Degree of Agreement	Responses to "Major Componentin Curriculum"	%
Strongly Disagree	1	74.4
	2	13.3
	3	3.5
	4	2.8
Strongly Agree	5	6.0

All of the questions thus far have evaluated the current scope and perceptions about the importance of distance learning in departments of political science. What about future use and importance? When asked if "distance learning will be used to some extent in every course in our department," the respondents were still relatively pessimistic. This statement was softened by the terminology "to some extent," which includes the Web-based technologies that are likely to become substantially more pervasive, but also was made more stringent by the term "every." The department chairs' perceptions of future growth of the use of distance learning were surprisingly

modest. The proportion strongly disagreeing with the statement of future use of distance learning was 62.7%, while only 13.7% agreed strongly or moderately. Table 4 reports these findings.

Table 4
Future Extent of Distance Learning in Political Science Courses

Degree of Agreement	Responses to "Future Extent"	%
Strongly Disagree	1	62.7
	2	15.1
	3	8.5
	4	4.9
Strongly Agree	5	8.8

Respondents also were asked if they thought "distance learning is largely a fad." This question was meant to elicit information about the future of distance learning again, only using different language. The responses, however, did not mirror the results for the preceding question. Only 21% of responding departments strongly or moderately agreed that distance learning was largely a fad. On the other hand, 44.3% strongly or moderately disagreed with the statement. In other words, although political science department chairs reported relatively low use of distance learning currently and were not much more optimistic about increased usage in their own departments in the future, they did not feel, as a group, that distance learning was transitory in the field. This would seem to indicate a perception (or perhaps resignation) that some departments or entities in the field would become major providers, but that most departments would be modest users of distance learning. See Table 5 for a summary of the results.

Table 5
Perceptions of Distance Learning Faddishness

Degree of Agreement	Responses to "Largely a Fad"	%
Strongly Disagree	1	20.0
	2	24.3
	3	34.6
	4	14.6
Strongly Agree	5	6.4

Type of Distance Learning Technologies Used

Another important question had to do with the type of distance learning technology that actually was used by political science faculty members. Ten choices were provided in a menu, with an eleventh option of "other." Respondents were asked to circle all technologies that applied in their respective departments. The percentages reported here are for distance learning users only; however, it must be remembered that distance learning users represent only 42.2% of the total population of respondents for this question. By far the most popular methods were Internet/World Wide Web delivery (58.4%) and e- mail interaction with remote students (54.4%). Other common methods employed were: multiperson computer interactions (32.8%); fiber optic, full-motion video, and two-way audio (32.0%); physically having the instructor at an off-campus venue (29.6%); correspondence by mail (25.6%); and telephone conferences (22.4%). Less common were public television class delivery, satellite

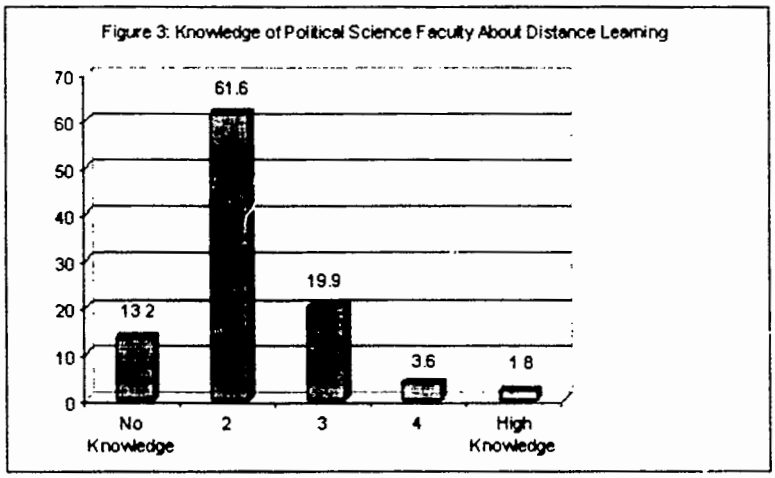
delivery, and other methods listed on the questionnaire or filled in voluntarily by the respondents. User respondents indicated the use of three distance learning technologies on average. See Table 6 for a comparison of the usage rates of the different methods. It is interesting to note that the most commonly used methods also are the newest; that is, they are all Internet- related technologies.

Table 6
Types of Distance Learning Technologies Used
(Multiple Responses Allowed)

Type of Distance Learning Technology	% of Distance Learning Users
Internet/World Wide Web delivery	58.4
E-mail interactions with remote students	54.4
Multiperson computer interactions (E.g., chat rooms, simulations, etc.)	32.8
Fiber optic full motion video and two-way audio	32.0
By physically having instructor at off-campus venue	29.6
Correspondence by mail	25.6
Telephone conference	22.4
Public Television class delivery	15.2
Satellite up/downlink	12.0
Satellite downlink only	6.4
Other	11.2

Faculty-Department-University Interest in Distance Learning

If faculty members are not knowledgeable about distance learning alternatives, they will not be able to use them. Respondents were asked, "How much knowledge about distance learning does the average member of your faculty have?" Seventy-five percent of the respondents said that the average faculty member has no or very little knowledge of distance learning on a 5-point Likert scale. Only 5% were quite knowledgeable. Another 20% were moderately knowledgeable about some aspects of distance learning. See Figure 3 for the results.



When asked about the level of interest in using distance learning techniques in the future, the response rates were similar to the question about levels of knowledge and the overall mean was identical. The specific question was, "How much interest in using distance learning techniques in the near future does the average faculty member in your department have?" A surprisingly large majority (68.1%) reported a definite lack of interest (a 4 or 5 on a 5-point scale) among faculty and active interest (a 1 or 2) was expressed by only 12.0%.

Only when a longer time frame is assumed are the respondents inclined to think that usage rates will increase substantially. In responding to the statement, "distance learning is a growing interest in our department," only 22.0% are inclined to agree either strongly or moderately. See Table 7 for a summary of the results from this question. An even more dramatic indication of the long-term pressure is the comparison of those who strongly agree that there will be a short-term upswing in interest with those who think there will be a long-term increase. While only 2.1% see a strong surge in short-term interest, 8.4% see a long-term interest. This four-fold increase may be due partially to familiarity, but it also likely is due to the integration of younger faculty members who are significantly more apt to be familiar and comfortable with distance learning. It also may be due to perceptions of technology improvements, access, and cost reductions.

Table 7
Growing Interest (longer term)

Degree of Agreement	Responses to "Growing Interest"	%
Strongly Disagree	1	28.2
	2	26.8
	3	23.0
	4	13.6
Strongly Agree	5	8.4

Although the average current and near-term level of interest was perceived to be very low, another aspect of distance learning diffusion is the presence of distance learning "pioneers" among the faculty. A pioneer is a person who is willing to take risks and try new and experimental technologies and to seek improvements in their application. Pioneers often are important in the widespread incorporation of distance learning technologies in an academic department because they act as both champions for the concept and role models of successful applications. The ability to identify a resident expert among the faculty is an indicator of a stronger distance learning prospect in the future. One interest in conducting this study was to establish a cohort of those who are perceived as pioneers or leaders in the area, for future study and support. When asked if there is "a person in your department who would be considered well informed or highly interested in distance learning?" and asked to identify that person, 47.1% responded affirmatively and provided a name.

What types of encouragement and support do faculty get to change old habits and invest the time and energy in new delivery techniques, some of which are inherently more labor-intensive and more demanding than traditional instruction? When asked "Are faculty pursuing distance learning with any assistance? (Circle all that apply)," 37.3% responded that they did not get any assistance whatsoever. Of those who did get assistance, 55.2% indicated some technical support, 23.3% indicated financial support, 28.7% indicated equipment support, and 5.4% indicated "other." These rates of response tend to indicate broad technical support from the department; interestingly enough, the reported rates of support were significantly greater than the reported rates of distance learning usage. However, when asked if the specific faculty members received "special incentives or compensation," 69.2% responded negatively even though recognition was one of the affirmative options. Thus, the response rate for specific faculty incentives (30.8% of all respondents) is significantly less than the reported rate of overall distance learning usage (42.5%).

Financial support for faculty was the most common means of encouragement and support, reported by 21.3% of all respondents to the survey (and by 75% of those responding affirmatively to this question). Of those who responded that special incentives or compensation were available to faculty members (less than one-third of the total respondent pool), the source of support was identified as the university by 63.3% of respondents, while 33.3% identified the college and 15.5% identified the department or other sources.

The Perceived Quality of Distance Learning

What are the perceptions among faculty chairs regarding the quality potential of distance learning? Overall, those perceptions are not good. When asked to agree or disagree with the question, "distance learning is generally not an appropriate way of teaching political science," nearly three-quarters of all respondents agreed with the statement. Nearly half of those strongly agreed (a 4 or 5) and the other half were in general agreement (a 3). Only 7.9% strongly disagreed with the proposition that distance learning was a generally inappropriate way to teach political science. See Table 8 for results.

Table 8
Appropriateness of
Distance Learning in Political Science

Degree of Agreement	Responses to "Distance Learning Not Appropriate"	%
Strongly Disagree	1	16.1
	2	21.1
	3	37.6
	4	17.2
Strongly Agree	5	7.9

Are faculty chairs more favorable when asked about distance learning at its best? When asked to agree or disagree with the question, "distance learning can be as good or better than conventional teaching," only 20.6% agreed strongly (a 4 or 5 on a 5-point scale), and another 33.1% moderately agreed. However, 46.2% felt that distance learning was incapable of ever being as good as conventional teaching, *even when distance learning was at its best*. See Table 9 for results. These two questions, taken together, indicate widespread and profound reservations about distance learning as a quality medium for educational delivery in political science. This finding goes a long way toward explaining the relatively small scope and role of, and the very modest interest in, distance learning.

Table 9
Distance Learning as Good or Better
Than Conventional Teaching

Degree of Agreement	Responses to "As Good or Better"	%
Strongly Disagree	1	18.1
	2	28.1
	3	33.1
	4	14.6
Strongly Agree	5	6.0

A series of four questions in the survey inquired about the effects of distance learning on the quality of education regarding students, faculty, department programs, and colleges or universities. The perceptions of faculty chairs in three of these areas—on the educational process for students, faculty, and departmental programs—follow a similar pattern and have identical mean response levels. Approximately 40% of responding department chairs are neutral about the effects of distance learning on the quality of education, indicating they believe that distance learning will neither improve education nor diminish it. Approximately an equal number feel that the educational process will be diminished. In these three cases, then, those who strongly feel it will diminish the educational process outnumber those who strongly feel it will enhance it by a 2-to-1 margin. The respondents are significantly more positive, on average, when the question relates to the educational effects on the college or university; however, those who strongly feel that the effects will be negative still outnumber those who strongly feel that the effects will be positive. See Table 10 for the responses to these four questions.

Table 10
Positive Effects of Distance Learning on Various Constituencies

Degree of Agreement	Response Options	Positive Effect on Students	Positive Effect on Faculty	Positive Effect on Departments	Positive Effect on Universities
Strongly Disagree	1	10.2%	10.2%	12.6%	10.5%
	2	28.0	30.7	27.7	24.2
	3	43.7	39.4	40.3	37.1
	4	14.6	17.7	15.8	21.9
Strongly Agree	5	3.5	2.0	3.6	6.3

Discussion

It was proposed here that the size and scope of distance learning are affected by three major factors. This relationship could be represented by the following formula:

Size and scope of distance learning =
(capacity of distance learning technologies)
X (market demand)
X (faculty/university interest in distance learning).

This study has examined intensively only the dependent variable in this model—the size and scope of distance learning—and one of the three elements of successful distance learning. Department chairs are well situated to provide information and opinions about the size and scope of distance learning, as well as the level of interest in distance learning among their faculties, departments, and universities. However, we did not investigate either the capacity of distance learning technologies or the nature of market demand because academic department chairs may not be particularly well situated to provide more than impressionistic data in this area. Nonetheless, the data supplied through this study provide an important baseline and the means to design some hypotheses about those areas that have not been studied directly.

First, the size and scope of distance learning in political science are small from any perspective. For such low size and scope, according to our model, all the contributing factors must be relatively small. Furthermore, the size and scope of distance learning in political science are projected to stay small for some time. In our survey, the only item indicating that department chairs may see possible long-term growth in this area of the field is the question related to faddishness. That is, most

chairs do not see distance learning as a fad, even though little or no short- term growth may be projected.

Certainly the level of interest in distance learning demonstrated by the chairs of political science departments was low overall. The average level of knowledge was quite low, the extent of near-term interest was very small, over half of the departments failed to have an identifiable pioneer, and specific support and financial incentives were not the norm. Also, faculty chairs as a group were very skeptical of the quality of distance learning, with significant blocks of them harshly critical of distance learning, even at its best. These data are important because they indicate that if future growth is likely to occur in distance learning in the field of political science, it is unlikely to come from institutions and faculty *as educators*. Institutional push from within is unlikely to be the chief promoter of distance learning.

Technical capacity was not studied directly here. However, one question—the type of distance learning technologies employed—did provide indirect information. Numerous methods are already in use. It remains to be seen whether many of these methods are going to play a small role, as methods of distance learning have done in the past, or whether they are a beachhead and provide a launching point for substantial future expansion. The Internet does provide genuinely new and affordable distance learning options, although the software and expertise are still limited across the higher education landscape. Because the Internet already has reconfigured other enormous industries such as mail and telephone, and because it is beginning to make gigantic inroads in commerce itself (book sales were the example used earlier in this article), it does seem that higher education is wise not to assume that new technologies are merely a fad. Nonetheless, issues of quality and faculty inertia must be overcome by continued growth in user-friendly technological improvements if significant increases in distance learning are to be seen in the short-term or medium-term.

Neither was market demand examined directly in this article. However, some indirect evidence on that point is provided by the results of certain questions in the national survey of political science department chairs. There were no suggestions in these data that distance learning competition is significantly affecting political science departments at this point, and only 10 institutions (3.4% of the sample) indicated that they offered 10 or more distance learning classes. Although it would seem likely that market demand will increase, it is impossible to predict with any accuracy how quickly demand will increase and to what degree. The data presented here suggest that most political science chairs are not gearing up for greater demand in the near- term. Yet at a broader level some established institutions seem to be gearing up nationally with significant incentive and program enhancements, and the new virtual universities are still ramping up. Although it has been found that over 90% of all universities with enrollments over 10,000 and 85% with enrollments over 3,000 have some distance learning classes (McGlynn, 1999), individual departments are far less consistent and supportive. It is simply too soon to tell just what this will mean for higher education generally, and for political science specifically.

Future Research

Although it is customary for researchers to call for more study in their area of interest, that is more than a pro forma recommendation in this case, given the exploratory and incomplete nature of the research to date on distance learning in political science. We believe that there are at least three critical areas to examine in more detail. First, it is important to provide a baseline on two of the contributing factors. Of the elements of the model that we propose, which identifies three elements that in combination lead to the growth of distance learning, we were able to study in depth only the result (current size and scope) and one contributing factor (faculty/university interest) because of the nature of the audience surveyed. Two elements (the capacity of distance learning technologies and market demand) are not studied here directly. Such study requires an examination of the specific technical capacities of distance learning related to political science courses, perhaps through case studies, and an examination of demand factors, perhaps by investigating the leading competitors, surveying various types of students, and scrutinizing related

disciplines.

Second, one aspect of the faculty/university interest factor that desperately needs further exploration is the highly negative perception about the quality of distance learning. Are there any relevant examples of high-quality distance learning in each of the different distance learning domains (two-way interactive video, Web-based, correspondence, etc.)? If so, what are the factors that lead to the high level of quality? What structural problems need to be overcome or minimized? What are the structural opportunities on which to capitalize? What are the common problems encountered in implementing distance learning, and how can communication be encouraged to share knowledge about what would be necessary to overcome them? Clearly, political science chairs, as a group, perceive that there are problems with distance learning. The most immediate utilitarian question is: What can be done to minimize the legitimate concerns about distance learning? Following from the answer to that question is the other essential query: What can be done to change the perceptions about distance learning that construct barriers to its successful implementation? These questions need to be addressed with the goal of achieving practical programmatic assessment, perhaps along the lines suggested by Banta, Lund, Black, and Oblander (1996) and the American Association for Higher Education (1992).

Third, it is important to track the baseline data longitudinally. We intend to repeat this survey after two years to see what changes have occurred with our targeted audience, political science department chairs.

Conclusion

In many respects, the results of this survey provide sobering reminders of the difficulties and complications associated with the adoption and diffusion of new instructional technologies (see, e.g., Rogers, 1995). Political science faculty (and their departments), as with many academic disciplines, seem to lag rather far behind in the adoption of innovative distance learning technologies. Incentives for faculty members to participate in distance learning are at best sporadic and uncertain. Levels of interest and participation in distance learning cannot be expected to increase appreciably until there are clear and sustained benefits for faculty members to take part in what often is a major drain on their time and intellectual energy. Publication requirements for promotion, tenure, merit increases, and honorific recognition may not coincide with outlets available for publishing the results of scholarly studies on distance learning. Also, the time and energy commitment required to get innovative distance learning courses off the ground may detract greatly from what it takes to be a fully functional academic professional in a discipline like political science. It would be of great interest to know if other disciplines evidence similar characteristics of career opportunity structures.

Addressing the perceived quality of distance learning courses is essential in any effort to get faculty members to commit themselves to the evolving instructional possibilities associated with instruction at a distance. It is imperative that distance learning not be seen as a poor stepchild within the broader departmental curriculum, nor that it be seen as providing watered-down versions of on-campus offerings. To achieve the objective of integrating distance learning within departments of political science in particular—and within any other academic department—issues of course quality and curricular integrity cannot be ignored. As with any innovation (Rogers, 1995), several stages of progression toward widespread adoption of distance learning will be followed, with varying degrees of success. There is likely to be a high level of resistance in the academic context arising from a combination of individual and institutional impediments that raise barriers to adoption.

James J. Kaput of the Department of Mathematics at the University of Massachusetts-Dartmouth and Jeremy Roschelle at the University of California, Berkeley indicate in regard to implementing digital education initiatives that there exists in traditional education "... an entrenched layer-cake, formalist-oriented curriculum that prevents most students from seriously engaging with important ideas. This curriculum is held in place by powerful interlocking forces and deeply institutionalized habits that allow space for innovation and growth only at the margins" (Kaput & Roshelle, on-line).

A powerful demonstration effect may be achieved by disseminating exemplary case studies of how to do distance learning right and by evaluating how best to link distance learning with the more successful aspects of higher education curricular innovations such as learning communities. Overall, an emphasis on holistic approaches to higher education, rather than on the development of specific course-based competencies, would seem to be a necessary prerequisite for enhancing perceptions of the quality of distance learning (Leip, 1999). How to achieve that holism is not obvious, but a reasonable starting point might be to establish specific recognition (for example, faculty teaching excellence awards) of outstanding performance in distance learning and thereby provide institutionally- supported targets toward which all can aspire. More general reward structures that enhance the opportunities for promotion, tenure, and advancement certainly need to take into account the special requirements imposed by a commitment to distance learning. Failing that, it is difficult to see how disciplines such as political science can be expected to join other fields of study in expanding and maintaining a commitment to distance learning. The proposed guidelines for Information Technology in Political Science drafted by an ad hoc committee of the Computers and Multimedia section of APSA is a good start in this direction. (On the Web at http://www.public.iastate.edu/~sws/).

Ferdi Serim has put the dilemma we face nicely,

> The symbiosis between education reform and the integration of technology into learning is profound: technology requires the rich learning environments envisioned by reformers; reform demands the power of technology to put people at the center of their own learning. Systemic adoption of reform will take a critical mass of educators, who must await the realization of the promises of technology to transcend isolation and join in collaborative professional growth.

We who are concerned about the future and direction of education face a scalability problem: reform requires these educators to rise to the level of performance typically encountered in master teachers. This realization can invoke a sensation of paralysis. The resulting inertia mirrors the way that fear of technology prevents many of our peers from having the experiences which would enable them to embrace, then direct, the potentials that technology-savvy educators rhapsodize about." (Serim)

In the end, we agree with Dennis Trinkle (1999, p. A60) that "the reality of distance learning is complex, and we must give it the measured consideration it demands." With Trinkle, we believe that distance education is a means to an end; hence the end must be measured by student learning outcomes and by institutional and programmatic academic integrity.

Notes

(c) 2000, Schmidt, et. al.

The authors wish to thank the Iowa State University College of Liberal Art and Sciences and SAS Consulting (http://www.*doctorpolitics.com*/ for generous support in conducting this survey.

1. Narrower wording might have stated: "Does your department have classes that are primarily distance learning based?"
2. Reported response percentages for individual questions are based on those responding; nonresponses for individual questions are excluded.
3. An alternate question asked for the same type of information but used the opposite perspective: "Distance learning is a marginal part of teaching in our department." The results were nearly identical and therefore are not reported here.

References

American Association for Higher Education. (1992). *Principles of good practice for*

assuring student learning. Washington, D.C.: American association for higher education.

American Association for History and Computing. (1998). Found at http://mcel.pacificu.edu/JAHC/JAHCII1/ARTICL ESII1/Trinkle/trinkleindex.html

Banta, T. W., Lund, J. p., Black, K. E. & Oblander, F. W. (1996). Assessment in practice: Putting principles to work on college campuses. San Francisco, CA: Jossey-Bass.

Blumenstyk, G. (1999), Distance learning at the Open University: The British institution's success has inspired imitators in the United States. *The Chronicle of Higher Education,* July 23, pp. A35ff (Information Technology section).

Boaz, M. et al. (1999), *A handbook for instructors*. Mission Viejo, CA: League for Innovation in the Community College.

Brigham, D.E. (1992). Factors affecting the development of distance education courses. *Distance Education - An International Journal,* Volume 13, Number 2, on line at: http://www.usq.edu.au/dec/DECJourn/v13n292/b righam.htm

Carr, S. 1999. Independent bookstores plot counter-offensive against chains, on-line sellers. *The Chronicle of Higher Education,* July 29 (online daily news, Headlines news section).

Clark, T. (1993). Attitudes of higher education faculty toward distance education: A national survey. *The American Journal of Distance Education, 7*, 2, pp. 19- 33.

Guernsey, L. (1999, September 2). Click here for the ivory tower. *The New York Times*, p. D1.

Hardin, J. & Ziebarth, J. Digital technology and its impact on education, [National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign] *The future of networking technologies for learning.* U.S. Department of Education's Office of Educational Technology, found at: http://www.ed.gov/Technol gy/Futures/hardin. html

Jaffee, D. (1997). Asynchronous learning: Technology and pedagogical strategy in a distance learning course. *Teaching Sociology*, *25* (October), 262-277.

Kaput, J. & Roshelle, J. Connecting the connectivity and the component revolutions to deep curriculum reform. *The Future of Networking Technologies for Learning.* U.S. Department of Education's Office of Educational Technology, found at: http://www.ed.gov/Technology/Futures/kaput.h tml

Kiernan, V. (1999). On-line bookstores use marketing and scholarship donations to attract students. *The Chronicle of Higher Education,* Friday, August 27 (online daily news, Information Technology section).

McGlynn, A. (1999, July 15). Earn a degree, no commute. *Des Moines Register*, pp. 1M, 4M.

McIsaac, M. S. , & Gunawardena, C. N. (1996). Distance education. In *The handbook of research for educational communications and technology*, edited by D. J. Jonassen. New York, N. Y.: Simon & Schuster Macmillan, 403-437.

Newcombe, T. (1999). Virtual universities: Revolutionizing education or just digital diploma mills? *Government Technology, 12*, 10, 20ff.

Noble, D. Digital diploma mills. Found at: communication.ucsd.edu/dl/ddm1.html.

Olsen, F. (1999a). Virtual institutions challenge accreditors to devise new ways of measuring quality: As technology makes inroads in higher education, critics warn of approving 'experiments.' *The Chronicle of Higher Education,* August 6, pp. A48ff (Information Technology section).

Olsen, F. (1999b). On-line bookseller signs deals to handle textbook orders for 2 college stores. *The Chronicle of Higher Education,* July 23, p. A38.(Information Technology section).

Rahm, D., Reed, J. J., & Rydl, T. L. (1999). Internet- mediated learning in public affairs programs: Issues and implications. *Journal of Public Affairs Education, 5* (3), 213-223.

Rogers, E. (1995). *Diffusion of innovations* (4th ed.). New York, N.Y.: The Free Press.

Schmidt, S. (1999). Distance education 2010: A virtual space odyssey, a paper prepared for the conference *Teaching with technology: rethinking traditions.*

Serim, F. Building virtual communities for professional development in *The future of networking technologies for learning* U.S. Department of Education's Office of Educational Technology, found at: http://www.ed.gov/Technology/Futures/serim.html

Trinkle, A. (1999, August 6). Distance education: a means to an end, no more, no less. *The Chronicle of Higher Education*, p. A60.

Young, J. R. (1999). Kentucky's virtual university aims to help students and institutio *The Chronicle of Higher Education,* Friday, August 13 (online daily news, Information Technology section).

About the Authors

Steffen W. Schmidt
University Professor of Political Science
Iowa State University

Email: sws@iastate.edu

Steffen W. Schmidt is co-author of *American Government and Politics Today*; 2000-2001 (Wadsworth), *Issues in Iowa Politics*, (Iowa State University Press) with Lee Ann Osbun, *Friends, Followers, and Factions*, (University. of California Berkeley Press) and numerous scholarly articles. He has written extensively on information technology (IT) and chaired an ad hoc committee of the APSA Computers and Multimedia Section that recently recommended a policy statement on IT to the APSA. He is the Political Science Editor of StudentAdvantage.com.(http://www.studentadvantage.com)

Mack C. Shelley
Professor of Political Science and Statistics
Coordinator of Research in the Research Institute for Studies in Education at Iowa State University.

Professor Shelley's publications include *American Public Policy: The Contemporary Agenda* (Houghton Mifflin) with Steven G. Koven and Bert E. Swanson, and the forthcoming *Redefining Family Policy: Implications for the 21st Century* (Iowa State University Press) with Joyce M. Mercier and Steven Garasky. He is co-editor of the *Policy Studies Journal.* He also is a co-author of *American Government and Politics Today, 2000-2001* (Wadsworth)

Monty Van Wart

Director of the Center for Public Service
Texas Tech University.

Dr. Van Wart's most recent book is *Changing Public Sector Values* (Garland Press). His research on education includes *Training and Development in the Public Sector*, (Jossey-Bass). He teaches distance education classes every semester.

Jane Clayton
Materials Sciences and Engineering
Iowa State University.

Jane Clayton is conducting ground breaking work in the area of assessment of learning in Engineering. She was Project Team Coordinator for Analysis of the survey data for this study. She is the co-editor with Steffen Schmidt and Mack Shelley of the book *Readings in American Government*, Third Edition, 2000, Wadsworth Publishing.

Erin Schreck
The Eaton Company
Cleveland, Ohio. A graduate in Public Administration from Iowa State University, Erin Schreck conducted the pre-tests of the questionnaire used in this study and was the coordinator of data entry and coding.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatistica
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 230 times since June 20, 2000

prior vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 8 Number 28 June 20, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Government Policy on Teacher Evaluation in Greece: Revolutionary Change or Repetition of the Past?

Michail D. Chrysos
McGill University

Abstract

After nearly two decades of freedom from evaluation, teachers in Greece became the focus of a new evaluation system. In 1998, reformers sought to raise the level of student performance by the regulation of teacher performance through a top-down evaluation system administered by the Greek Ministry of Education and Religous Affairs. The probable effects of this evaluation system on teachers' professional roles and development are analyzed.

Political and Historical Framework

Greece represents a sound example of Cuban's (1995) argument that educational reforms return again and again. This occurs, he argued, because "reforms have failed to remove the problems they intended to solve". For over one hundred years, Greece has been characterized by abortive, short-lived educational reforms, which have never been implemented for more than a few years, and then were abandoned by the Ministry of Education and Religious Affairs (MERA) for having failed to bridge rhetoric, design and reality (Persianis, 1998).

Following the restoration of democracy in 1974, and the entry of Greece into the European Union in 1981, Andreas Papandreou's Socialist Government came to power. His agenda included the designing of new reform proposals that would accelerate the democratization as well as modernization of the Greek educational system. As a member of the E.U., Greece places emphasis on reaching West European standards and innovation. Greek schools, a highly centralized system under the jurisdiction of the MERA, has been following French and German teaching methods "... regurgitation of foreign pedagogical thought" (Curtis, 1994; Persianis, 1998). The country is divided into fifteen administrative regions for education, each of which is subdivided into 240 districts (Peripheria), and is headed

by evaluators-inspectors who monitor the application of the curriculum. The educational programs are directed by provincial and local authorities (Director and Employer of school Offices, one in each province) under the managerial general policy guidelines of the MERA. The latter is composed of all kinds of offices and institutions (Pedagogical Institute) that function according to central authority regulations, which motivate, lead, and sponsor any policies and draft laws, increasing the bureaucratization of schooling at all levels.

It is obviously difficult for those within educational bureaucracies to offer critical policy analyses. In Europe, educational control is governmental (France) or quasi- governmental (Great Britain), and it has been observed that educational policy is located within the administrations of liberal or conservative parties. In Greece, even minor changes depend on decisions made by the MERA, which reinforces the top-down manipulation of policy decisions.

It would not be misleading to say that there is no consensus on policy among the major political parties, especially as it relates to the New Democracy and Panhellenic Socialist Movement. Each party strives to promote its own ideological principles and interests rather than to develop on-going goals through mass political organizations or interest groups. The centralized nature of the administrative structure of the Greek Educational System has been challenged through various attempts at "political manipulation" by the governing party elite and the different interests groups (Gouvias, 1998). Moreover, each Minister claims to leave his stamp on any educational reform and ensure his lasting reputation in the history of Greek education. An instance of this appeared in June 1996, when the new Minister of Education, G.Arsenis (also a socialist) launched the reform for "Ethniko Apolyterio" (National Leaving Certificate). He promised to develop school curriculum, to provide in-service training for teachers, to reestablish a whole hierarchy of evaluators whose mandate would be to monitor and solve problems for the sake of teachers' improvement. The new reform was enacted by the passage of legislation, and instituted a politically motivated program of Teacher Evaluation. Unfortunately, the reform was announced "suddenly" without previous warning in the summer season (vacation for schools), a typical strategy the Greek state uses to secure legitimacy and reduce resistance.

Issues such as appointments, duties, inspection, evaluation and so forth, have always been worked out in drafts of legislation. The Minister with the cooperation of legislators and executives from the MERA wrote a reform bill, took it to the Parliament, and asked his colleagues to make it law, in a manner that Wilson (1996) ironically calls "ministerial responsibility." Greek Ministers actions reflect the attitude of centralized bureaucracies, which attempt to "secure" their positions by law before negotiating among practitioners and taxpayers. Instead, policy agendas must be socially negotiated in a "National debate of education" among all factions--the government, policy-makers, and practitioners, whicht in a broad sense facilitate communication in solving problems cooperatively (OECD, 1995).

In the new era of educational reforms, no area has received more emphasis than the quality of instruction and those employed to deliver it. Duke (1995) indicated that " the key to educational improvement lies ... in upgrading the quality of teachers"; central to improving the quality of teachers is the teacher evaluation-inspection-supervision process. The issue then becomes how to refine and change the content of the traditional top-down flow of policy.

In 1981, the Socialist Government passed Law 1340/82, which abolished the influence of inspectors. Since then, teachers and school organizations have been free of inspection. That Law of inspection remained in existence until recently, though with no substantial role in enhancing teaching quality. All these years, teachers were being appointed but were never formally evaluated. In this policy vacuum, teachers had the unique opportunity to take advantage of their newly found liberties and promote the professionalism of teaching; unfortunately, they did not avail themselves of this opportunity.

On the other hand, the model of a more flexible evaluation was a great challenge for Greece, which could not suddenly allow the whole educational system be in a vacuum without internal restrictions and rules. Reformers sought to raise the level of students' performance by the regulation of teacher performance. According to the Government Gazette 27/02/98 and the application of Law n.2525/98, the new

evaluation policy underlies the top-down evaluation of all participants from researchers, policy makers, evaluators, principals down to teachers. The results of these evaluations are to go directly to the Central Offices of MERA. Before analyzing some noteworthy issues as regards that evaluation, it is essential to discuss briefly the role of the government in policy making.

The role of the government

In Greece, the government is the principal source of funding. It sponsors any kind of policy research through the Pedagogical Institute. Its agencies are appointed and not elected, and are accountable to the public through the MERA. That situation creates the situation of a "crisis of confidence" (OECD, 1995), because any kind of policy making has the reputation of being fragmented and politicized, and as a result there is no trust among the stakeholders, either in higher levels of the hierarchy or at the base of school organizations. The social scientists perceive evaluation and authority as interconnected (Stone, 1988) in a centralized authoritative educational system, where there are levels of superiors (evaluators) and subordinates (evalua ees).

The former exercise authority based on the power of law and political skill rather than on interpersonal relations, whereas the latter show compliance with the control system. It is difficult for a single center to control the complex modern educational system. It is for this reason that the centralized system has been criticized for lack of imagination and its "top-heavy" structure in making decisions (OECD, 1995). The needs of the government and of the practitioners cannot both be met.

When one political party leaves office, it is replaced by another, which has different views and priorities. Furthermore, "clientelism" pervades Greek education--the belief that the criteria for appointment of teachers, evaluators and other employers or employees are usually political following the well-known "rousfeti" (personal favors by politicians to clients). Stone (1988) correctly argued that policy making tends to be essentially political and involves a struggle over ideas, implying that the development of policy has not followed a linear, rational model, but a model of differentiation. In this model, experts and policy makers generate and bring knowledge into theories, which, later on, teachers use and practice. Political parties with strong and consistent ideology - as in Greece - have stopped holding consultative meetings with teachers unions; they are convinced that they know what to do without consulting teachers.

Why the restoration of evaluation is so important

The current policies represent the first time that the MERA has paid so much attention to evaluating-supervising instruction, teaching and especially teacher appropriateness for school productivity. It is noteworthy that with the present policy everybody is being evaluated--from principals to employers of educational offices, directors, and inspectors-consultants. It is a top-down, multi- dimensional hierarchical form of evaluation. However, teachers are the focal group who are being evaluated and self-evaluated from multiple directions from higher levels (See Figure 1).

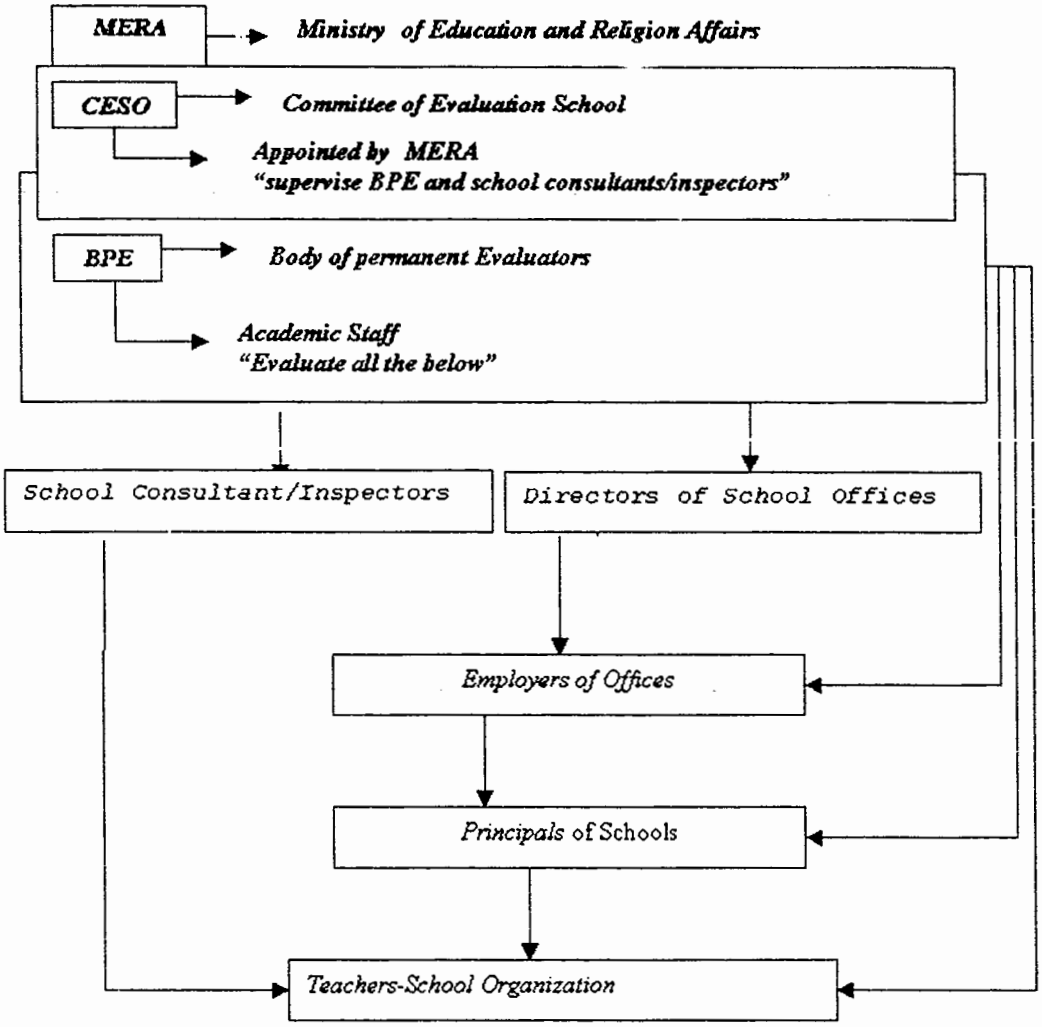


Figure 1. The Evaluation Pyramid

The only exception occurs at the top of the pyramid of evaluation, BPE (Body of Permanent Evaluators), whose members will not be evaluated but are elected by the MERA through public competition. The enabling legislation underlying this policy does not mention the qualifications of the personnel who will occupy this level of the evaluation system. At the highest level there is the Committee of Evaluation of School Organizations (CESO) which "supervises, controls and coordinates the functions of BPE and school consultants" (Law, 2525/98, article 5, FEK 188A' & Contemporary Education, 1997).

Evaluation is a significant tool in controlling what is going on in schools and it seeks to promote the self-development of teachers and the quality of their instruction. The type of evaluation that the new law in Greece proposes is twofold. It includes both a formative evaluation element, which is based on the "art of teaching" (Barber & Klein, 1984, pp.96-97) and emphasizes teacher performance and process of instruction, and a summative evaluation element, which is grounded on both processes and products of instruction. In fact, evaluation should empower teachers to use teaching methods that will benefit students' learning. It is not suggested that teacher evaluation be implemented in isolation, but rather in combination with other school improvement initiatives. However, the question that arises is whether the criteria of evaluation reflect international, national, regional and local needs of education. The general issues of the new policy remain the same across the country, but seemingly they are flexible to adjust to the local needs.

In Greece, the main contributors to evaluation theory and methodology have been academics and educational researchers - like those in BPE and

CESO—working under the directive guidelines of the "political center." The same happens in a variety of countries such as the United States of America, where evaluations are conducted by specialized external evaluators (Wilcox, 1989). They produce standard questionnaires that any level of employees dealing with quantitative outcomes must complete, instead of conferring or advising teachers. In this respect, the new reform appears to be a "non reform," inasmuch as it repeats and re-establishes anachronistic procedures, mainly those that move the government to the position of the employer, and the teachers to the position of employees in an atmosphere lacking mutual trust and collaboration.

On a positive note, the new system is the first time that teachers have the chance to evaluate themselves, though I am not convinced to what extent it will be a positive experience nor how powerful will be the final reports sent to the higher levels of official evaluation system. Undoubtedly, self-evaluation represents an innovation, since it affects the local community and the teachers of each school, who will have, first, their own rules in the policy of self-evaluation, to solve their own problems, and secondly, a reasonable degree of autonomy (Law D2/1938/26-02-98).

Who are the evaluators and what is their role?

One of the most noteworthy features of the new hierarchical policy of evaluation is the creation of two types of evaluator, the Internal (principals, directors, employers, inspectors, and consultants) and the External (BPE, CESO). In Britain and the USA, internal evaluation employs people who are not members of the evaluated institutions, rather they are specialists with the mandate to check on the use of public funds and insure that information be forwarded to the central government. They are experienced professionals who make formal and informal visits to school organizations to interpret (statistically) those organizations. Are these findings trustworthy, however?

It is worth mentioning that, paradoxically, in 1927, more than seventy years ago, the Greek National Committee of Education consisted of seven expert policy makers and four teachers. In other words, this Committee was an internal autonomous institution, which was not being controlled by the Ministry of Education (Contemporary Education, 1997). Since then, teachers have been marginalized and relegated to their traditional roles in the classroom.

The term "school consultant" was imported to Greece in 1982, to replace other euphemistic terms like Inspector, Supervisor, and Cooperator (Kotsikis, 1993). Clearly, evaluators--internal or external--possess different levels of expertise and experience. Under these circumstances, it is impossible for practitioners to join with researchers, policy makers, and academics in the cooperative quest for usable knowledge. The view that all are involved happily in a mutually beneficial exercise is a romantic fantasy.

Evaluators—mainly external—operate in different frames of thinking, use a different language and respond to different incentive systems. Teachers accuse their "superiors" of being unaware of what is going on in classes, because evaluators/supervisors often focus on complexity. Both are committed to the improvement of schooling. Evaluators or any person holding an administrative or supervisory position must have an appropriate attitude, knowledge, and skills (Bellon, 1984, p.219). The new policy for Teacher Evaluation is likely to favor evaluators-consultants who have had lengthy instruction and possess administrative skills rather than the interpersonal, communicative potential to know how to transmit ideas, and how to build positive working relationships based on trust and honesty. It is difficult to reduce interpersonal conflict without trust and respect (Bellon, 1984).

Another important dimension of evaluating is grounded on the kind of leadership that the evaluator conveys. Even though the role of the evaluators is controlled by law and by the bureaucractic hierarchy, evaluators need the charisma and the inspiration to perform like real leaders. They must be able personally to demonstrate transformational leadership behavior, i.e., to be able to discover and probe uncertainties, to stimulate the motivation and maturity to increase autonomy and sense of duty (Silins, 1994). Nevertheless, although external evaluators ought to be capable of translating their theoretical messages into practical applications, their

authority is inevitably insufficient to convince teachers to review their professional skills and to tackle new ideas of responsibility.

Another aspect of the new reform is the abolition of tenure for teachers in public schools. Among those current teachers who are certified, there are unqualified ones who do not pursue educational improvement. Although teachers reacted to that regulation, I have no doubt that the intention of the government is to appoint well-qualified teachers. You can imagine the tremendous importance evaluation takes on. Is that changed image seen as threatening, or does it lead to teacher improvement? It depends on the nature of the supervisor who will help teachers become more competent.

In the traditional type of supervision, teachers accepted passively the evaluator's opinions without complaints, and the supervisor decided which teaching methods the teachers should modify. The contemporary notion of "clinical" supervision breaks down the former distance between teachers and supervisors, while teachers themselves are allowed to decide what aspects of their teaching are to be observed and improved, empowering their self-supervisory skills (Reavis, 1977) and creating a "mutual support system called colleagueship" (Schonberger, 1983). It remains, however, the task for school consultants and BPE experts to bridge the stereotypic gap between themselves and teachers. The procedure of completing a questionnaire does little to bridge this gap.

When a teacher and an evaluator both think of complex stages of conceptual development, they employ a greater repertoire of instructional techniques; and consequently, the risk of supervisors misunderstanding teaching performance is drastically decreased. Teachers demonstrate improved teaching behaviors when supervision focuses on a specific behavior (clinical supervision) through active participation. On the other hand, evaluators (internal or external) have spent little time in classrooms watching teaching. Teachers treat evaluators formally and hinder them from understanding the reality of school life; they believe that evaluators lack adequate time and training to undertake effective evaluations.

Relationships between supervisors and teachers should be characterized by trust, credibility, and support for productive interaction; a trusting relationship implying confidentiality (Pfeiffer, 1982). Teacher needs and the evaluator's skills are important considerations in the successful implementation of a new program. Teachers consider their classrooms as private domains and no one invades this territory; but in an environment of sensitivity and respect, when the internal and, most importantly, the external evaluators perceive that they have met and conferred with professional teachers, the traditional hierarchy of formal authority becomes more functional and evaluation represents a more collaborative process.

Considering evaluation as a sort of "investigation," teachers can subjectively determine their standards, wants, desires and present their findings, and offer feedback for themselves and others. Because they are concerned about students and tasks of teaching as well, they reflect upon leadership and accountability for instructional and personal growth. Evaluators bring their own objective lists with questions that represent a standardized measurement of teacher performance based on complex sets of explicit and implicit standards and on teaching theory rather than on realities of classroom life. The evaluator must be a person who uses "judgment as a tool, works to make sense of a particular school" (Wilson, 1995). In opposition, even though supervisors should be the instrument of decentralization, they should draw the connective chain from local to central authority, and give the final statement regarding the quality of a school.

The new Greek evaluation system delegates authority to principals who, instead of advising and organizing instruction, can now control and set realistic expectations in achieving teaching objectives. Assuredly, that change is based on the lack of internal evaluators who could be engaged full-time in observing school life. In the meantime, it seems that the government intends to supplement the abolition of tenure by removing incompetent teachers from classrooms. In regards to the new policy, principals assume the authority to reject or alter teachers' goals, while evaluating teacher development within a general infrastructure.

My argument is twofold. On the one hand, principals consider their buildings as private territories regardless of teachers' opinion about their effectiveness to act as educational leaders. The question remains: Do principals have the appropriate skills to evaluate and supervise the teaching staff, even though they primarily identify themselves with a political party? Furthermore, how could the validity of evaluation be secured where there are personal disparities and differences? (Contemporary Education, 1997, pp.150). Finally, how can the whole school organization function in harmony and be productive within such an environment? Multiple evaluation presupposes more data and more opportunities to corroborate findings.

Evaluation and Evaluators from Teachers' Perspective

Teachers are of the opinion that "evaluation does not represent an external consideration of school reality but an internal one by people who are involved" (OLME, 1997; Duke, 1995). They want to set their own priorities on what knowledge would be most useful to their enterprises, and to strengthen a new professionalism, since teachers themselves would contribute with their own criteria in evaluation process (with emphasis on Self-Evaluation). Yet teachers require participating and planning for their individual students at the level of the Center of Decision (MERA). Such ambitions might change teachers' behavior and sense of accountability, and most important change their image and opinion of evaluation coming from higher to lower levels.

Teachers need to have confidence in the impartiality and competence of evaluators because the latter "are reluctant to use objective measures since they tend to face teachers as inadequate" (Barber & Klein, 1984). Usually teachers feel overloaded with both teaching responsibilities and episodes of evaluation-supervision that bring about frustration, conflict and pressure which in turn increase teacher stress and burnout. If evaluators adopt a new, more collegial, class-centered style rather than their office/authority-based manner in assisting teachers to define their instructional intent, autonomy will not be undermined and stress will be reduced (Goens & Knciejezyk, 1981).

Moreover, teachers complain about the lack of cooperation with senior administrative offices, which assume a different approach to education. They provide practical considerations about how evaluation may be carried out, and stress the importance for evaluators of living the reality of schooling every day. On the other hand, the academic and research communities start out with policies that represent politically attractive solutions. Duke (1995) claimed that "the initial impetus for changes has tended to come from political and theoretical-based rather then professional school-based demands and needs" (p.155). Politicians are so out of touch with the reality of schools that sometimes they do not even know if their policies are bad or if their goals are too abstract (Wilson, 1995). People whose work is crucial for the improvement of teaching and learning increasingly become

disengaged from the hard work of improving schools because others outside their workplace decide what the policies are going to be.

By all accounts, teachers, individually and through their associations (unions) resist policies they do not understand. When a new idea is introduced, resistance is the common reaction. Teachers are familiar and comfortable with prior procedures, because they know what to do. The unknown, unfamiliar can be frightening, since it will be analytically investigated and reviewed. The more complex and uncertain the policy-legitimate implications are, the more likely teachers will need information and insights into what evaluation is doing and what it achieves. Conflicts can be identified and discussed, while superiors and subordinates will have a wider range of options from which to choose and will become wiser from the effort of choosing (OLME, 1997).

Conclusion

Will Greece continue to appoint official evaluators based on political interests rather than on the past performance or qualifications of candidates? Will the inspector-supervisor-consultant-principal become an independent professional (school person) or will he remain a governmental technocrat? Whatever the outcome, it is imperative that the public know what schools are doing, and judge whether they are doing it well. It is important for schools to be monitored, to reveal bad practitioners, bad practice, and bad teachers. Furthermore, all the interest groups must show an increased sense of accountability, and work in a collaborative environment with explicit standards.

Government agencies are usually free from blame, while the achievement of a policy is placed primarily on the backs of practitioners-teachers. The evaluation process must be divided not in form, as occurs now, but in essence. Local policies should promote and facilitate the diffusion of innovations and initiatives from all people who are involved with education. We can no longer rely on bureaucratic mechanisms, on regulations in law that hinder change or on complex standards that force narrow definitions of effectiveness. Schools will change when we change our thinking about them (Wilson, 1995).

References

Cuban, L. (1990). Reforming again, again and again. *Educational Researcher*, 19(1), 3-13.

Curtis, G., E. (1994). *Greece, a country study*.

Dornbusch, S., M. & Scott, W., R. (1975). Evaluation and the exercise of authority (1st Ed.). San Francisco: Jossey-Bass Publishers (pp.134-150).

Duke, D., L. (1995). Teacher evaluation policy: From accountability to professional development. Albany: State University of New York Press.

Goens, A. & Kuciejczyk, J. (1981). Supervisors, do they induce or reduce teacher stress? *NASSP Bulletin*, 65(12), 24-27.

Gouvias, D. (1998). Comparative issues of selection in Europe: The case of Greece. *Education Policy Analysis Archives*, 6(4), (entire issue). Available online at http://epaa.asu.edu/epaa/v6n4.html.

Government Gazette (27/02/98): *Evaluation of instruction and teachers· The new reform* [in Greek], Legislative Degree 2525/98, 1340/82.

Kotsikis, E. (1993). *Administration and management in Education.* Athens: Ellyn [in Greek].

Merit pay and Evaluation (1984). [*Merit pay and teacher evaluation* by Barber, L., W. &.Klein, K., pp.93-97] and [*Evaluators of instruction* by Bellon, J., pp.210-231]. Bloomington, Ind.: Phi Delta Kappa, Center on Evaluation, Development and Research.

OECD, (1995). *Decision making in 14 education systems.*

Periodical of *Contemporary Education* (1997), p.10-13, 151-157. [In Greek].

Periodical of OLME (*National Union of teachers in Greece)*, (June-August 1997). [In Greek].

Persianis, P. (1998). "Compensatory legitimation" in Greek educational policy: an explanation for the abortive educational reforms in Greece in comparison with those *in* France. *Comparative education*, 34(1), 71- 84.

Pfeiffer, I., L. (1982). *Supervision of teachers: A guide to improving instruction.* Phoenix, AZ: Oryx Press.

Reavis, C., A. (1977). A test of the clinical supervision model. *Journal of educational research,* 70(6), 311-315.

Schonberger, V., L. (1983). The effective supervision, coordination and improvement *of* the instructional activities of prof. Colleagues. *Education,* 103(2), 129-131.

Silins, H., C. (1994). Leadership characteristics and school improvement. *Australian journal of education,* 38(3), 266-281.

Stone, D. (1988). *Policy paradox and political reason.* N.Y.: Harper Collins Publishers.

Wilcox, B. (1989). Inspection and its contribution to practical evaluation. *Educational research,* 31(3), 163-175.

Wilson, T., A. (1996). *Reaching for a better standard: English school inspection and the dilemma of accountability for American public schools.* New York: Teachers College Press.

About the Author

Michail D. Chrysos
Department of Educational Studies (Curriculum and Instruction)
Faculty of Education
McGill University

Graduate Studies Office
Duggan House 204
3724 McTavish Street
Montreal, QC, H3A-1Y2

E-mail: mchrysos*(a)* hotmail.com

In 1995, I received my BA in Ancient and Modern Greek Literature from the Department of Greek Philology of the University of Athens. In September 1998, after two years of teaching experience in a Preparatory Secondary School in Greece, I was accepted for the degree of Master in Education in Curriculum and Instruction in the Faculty of Education at McGill University. My main academic interests lie in the area of contemporary issues of curriculum development and improvement, qualitative research methods in education, and the relationship between literacy and

learning.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina-- Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois - UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 730 times since June 24, 2000

Education Policy Analysis Archives

Volume 8 Number 29 June 24, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Whither Advanced Placement?

William Lichten
Yale University

Abstract

This is a review of the Advanced Placement (AP) Program. In disagreement with claims of the College Board, there is firm evidence that the average test performance level has dropped. The College Board's scale and claims for AP qualification disagree seriously with college standards. A majority of tests taken do not qualify. It appears that "advanced placement" is coming closer to "placement." This article recommends that the College Board's policy of concentrating on numbers of participants should be changed to an emphasis on student performance and program quality.

Introduction

In 1953 the College Board began the Advanced Placement (AP) program, to challenge a small, elite group of able students. AP students took a college course in high school and an external exam to qualify for admission to advanced undergraduate work. The strength of AP was its eschewing fads for a solid collaboration between high school teachers and college professors, with an emphasis on subject content. An important feature was the evaluation of a high school student's work by outside examiners who were college faculty.

Since that time the program has taken on a life of its own and has spread widely throughout American high schools. The number of participants has more than doubled every decade. Today, more than half of American high schools and a third of four year college-bound seniors participate in this burgeoning program. More than a million AP exams, five hundred times the original number, are taken each year.

Whereas overall assessments of American public schools range from highly critical (National Committee on Excellence in Education, 1983, Ravitch, 1985, Finn, 1991) to favorable, even optimistic (Carson et al., 1993, Bracey, 1991-1998), all sides give AP their approval. This shows itself in a growing number of legislatures and state boards which support AP (twenty-three states in 1998, including D.C., College Board, 1998) in a variety of ways. The heart of the AP program is its examination, which is given at the end of the academic year, usually to high school seniors or juniors. Unlike norm referenced examinations, such as SAT and ACT, which are scored in percentiles or equivalent, AP gives criterion referenced examinations, which are pass or fail. The criterion in AP is whether or not the colleges will accept the student for advanced placement. Thus, any critical evaluation of the success of the AP program must hinge on the degree to which the program succeeds in overcoming this hurdle.

The College Board widely quotes its grade scale:

Table 1
Present College Board Interpretation of AP Scores
(approximate grade equivalents in parentheses)

5: extremely well qualified (A)
4: well qualified (B)
3: qualified (C)
2: possibly qualified
1: no recommendation

The College Board (1999a) claimed that,

> Almost two-thirds of the students achieved grades of 3 or above on AP's 5-point scale-sufficiently high to qualify for credit and/or enrollment in advanced courses at virtually all four-year collages and universities, including the most selective.

It is an open secret (Hyser, 1999) that both this claim and scale (Table 1) disagree with college standards. This disparity is a sign of remarkably poor communication between the colleges and the College Board. This paper discusses in detail the seriously misleading conclusions that follow from Table 1.

The Colleges and Advanced Placement

The success of the program is judged by measurable exam *performance*, as opposed to intangible benefits, which are difficult to evaluate objectively (Lichten and Wainer, 2000). The *raison d'être* of the program is qualification for advanced placement by the colleges and universities. To determine college practice, the author uses an enlarged version of the sample of Morgan and Ramist, 1998, but twice as large to include the lower end and make for more representativeness. (See Table 2. The sample may be slightly lenient, since it under-represents small colleges, which sometimes have stricter AP admission policies.)

These colleges and universities divide (by average AP scores) into three classes: "highly selective" (mean AP grade greater than or equal to 3.4, average SAT scores approximately greater than or equal to 610,), "selective" (AP 2.6-3.4, SAT *ca.* 500-610), and "non-selective" (AP ≤2.6, SAT *ca.* <500). (See Table 2. Sources for SAT or equivalent ACT scores are College Board (1999b) and Princeton Review (1998). AP data is obtained from the Educational Testing Service (ETS).) (Note 1) Then, with 5% dropped (typically colleges with only one AP candidate), the number of exams is 218,359 in highly selective, 519,521 in selective and 67,386 in non-selective schools.

The data in Table 2 differ for each of the three types of colleges. Highly selective schools require a "4" or more, with about three out of five exams qualifying to receive advanced placement. About half of the selective schools take "4's" and half take "3's", with about half of the exams qualifying. Non-selective schools usually accept a "3", but only one out of three exams qualify. Overall, scores of 5s and 4s qualify, 55% of 3s pass, and essentially all 1s and 2s fail, for an average pass rate of 49%. These results obviously disagree with College Board claims (Table 1 and subsequent text), and confirm Hyser (1999).

English Literature seems to have slipped farther than other subjects. Some colleges, not all highly selective, will not even accept a "5" for AP credit. The shift from a "3" to a "4" in selective colleges occurs more often for English Literature than for other subjects (Table 2).

Table 2
Data on AP for a Representative Sample of Colleges

College or University	Ave. Score		% ≥3	Number of...		Pass Score	Comments
	SAT	AP		...exams	...candidates		
Non-Selective							
Albany St U.	430	1.3	5.7	87	53	3	
Prairie View A&M U	420	1.54	11.1	81	48	3	
TN State U	460	1.71	16.2	271	166	3	
NC Agricultural Tech St	460	1.92	22.1	299	170	3	
Morgan State U	475	1.95	24.1	162	102	3	
Eastern KY U	455	2.07	28.7	366	190	3	
State U W GA	461	2.13	31.6	275	154	3	
Spelman College	537	2.22	33.2	561	311	3, 4	sci. & Engl. 4
U Southern MS	515	2.29	36.1	418	219	2	
Western KY U	495	2.36	40.3	514	258	3	
U West FL	535	2.36	44.2	240	115	3	
U NC Wilmington	454	2.37	41.8	977	525	3	

U TX Pan Am	NA	2.5	39.7	1282	559	3	
U South FL	545	2.52	45.9	1993	894	3	
U CA Riverside	511	2.55	47.5	4130	1576	3	
Appalachian St U	540	2.59	51.6	1732	802	3	
Selective							
George Mason U	515	2.63	49.4	1328	653	4	
FL State U.	576	2.69	53.2	4836	2030	3	
Auburn U	569	2.74	55.2	1707	836	4	
James Madison	585	2.74	57.5	4016	1631	4	
U. CA Irvine	520	2.77	56.3	8247	2708	4	
Clemson U	557	2.88	60.5	3963	1649	3	
U.CA Davis	565	2.94	62.4	7141	2658	3	
MI State U	540	2.95	62.7	4157	2086	3, 4	English et al 4
Cornell College	600	3.01	62.1	182	70	3	
U. GA Athens	599	3.02	66.1	6029	2493	3, 4	
U. Texas	601	3.08	67.8	14838	5063	3	
PA State U.	593	3.09	67.8	6362	2753	4	
UNC Chapel Hill	610	3.2	71.1	9386	2990	4	English 5
U UT	565	3.28	74.9	3835	1496	3	
Boston College	630	3.28	76	4213	1311	3, 4	
Tulane U	645	3.33	76.7	3002	973	4	
Brigham Young	610	3.35	77.9	10392	3960	3	
Highly Selective							
U. IL Urbana	610	3.42	78.3	10389	3596	4	
College of WM & Mary	655	3.59	83.3	3452	928	4	
U. Virginia	643	3.61	88.3	9488	2351	4	
Carnegie Mellon	641	3.79	87	3310	852	4	English 5
Cornell U.	660	3.81	88 4	9826	2315	4	
Duke U.	685	3.91	89.5	6615	1467	4	
Stanford U.	703	4.13	92.1	8390	1749	4	
Yale U.	730	4.25	94.4	5169	998	4	Engl 5 (or 4 & 760 SATV, II)

Of all exams that result in advanced placement credit, 32% came from students applying to highly selective colleges, 63% from selective colleges and only 5% from non-selective colleges. Overall college attendance divides approximately into 18% of students at highly selective colleges, 36% at selective institutions and 46% at non-selective schools (based on composite SAT score percentiles furnished by the College Board).

Extreme cases are Yale and the predominantly minority Albany (GA) State U. Applicants forwarding AP exams to Yale's admissions office take an average number of 5.2 AP exams. Three quarters of these 5169 exams (about 3900) from 998 candidates meet Yale's "4" requirement. At Albany State, with a freshman class of 660, 53 AP candidates take 87 exams, of which five are acceptable at a score of 3

or higher. The contrast between these two schools points up the successes and failures of the program.

The College Board Scale

To test the College Board scale (Table 1), assume for the sake of argument that all "qualified" and say half of "possibly qualified" persons merit AP. Then, if one applies Table 1 to the current figures for 1's. 2's, 3's, 4's and 5's (116, 240, 286, 207, and 142 thousands, with 0% ,50%, 100%, 100%, 100% passing, resp.), about (120+286+207+142)=755 out of the total of 991 (thousands) or 76% would qualify. Yet less than half of the sample qualified. The College Board scale overestimates the fraction of successful examinations by over a quarter of the total, by no means a trivial amount. In 1999, this would amount to approximately 300,000 examinations incorrectly predicted by the College Board's scale (Table 1). These examinations produce a revenue to C.E.E.B. of over $20 million and cause an obvious conflict of interest. Table 3 shows a scale, in agreement with Hyser (1999), which drops down by a full step on a five point range, i.e. such that half of the exams with a "3" qualify:

Table 3
A New Scale That Represents AP Data More Accurately Than the Old Scale of Table 1
(letter grades author's estimates)

5: well qualified (A)
4: qualified (A-, B+)
3: possibly qualified (B or C)
2,1: no recommendation

Under the same assumptions as before, (143+207+142) out of 991 (thousands), or 49%, would qualify. The latter figure agrees quite well with the data. Thus, the old scale (Table 1) is quite misleading and the new scale (Table 3) is a good fit. (Note 2) Note that *a majority of the AP examinations are not passing.* Since about one out three students taking the AP courses never take the examinations, the overall examination pass rate is only about one for every three course enrollments. (Note 3)

The College Board and the Colleges Disagree

The major disagreement between the two grade scales (Tables 1 and 3) shows a yawning gap in communication between CEEB and the colleges. Because the scoring criteria for A.P. are not public information, one can only guess at the causes for the discrepancy between the College Board's claims (Table 1) and the facts of college admissions. CEEB denies that such a discrepancy could reflect any change in quality:

> ...each exam grade indicates the same level of college-level learning from year to year and state to state. AP provides a true national standard of achievement that is constant over time. We make every effort to protect it from grade inflation. (College Board, 1996).

This claim, coupled with the allegedly consistent success rate, is a chimera for a several reasons. One major cause is apparent to this former college teacher upon inspection of Table 1: the very grade inflation that CEEB assures us does not exist.

The Lira failed to avoid inflation by pegging itself to the Euro, when the value of the Euro dropped. Likewise, AP scores have been pegged to college grades the same way since the beginning of the program in the 1950's, as shown in Table 1. As a person whose teaching career spanned this interval, the author remembers well the changes in grade scales. In the 1950's the average grade in introductory courses at

Yale lay midway between a satisfactory "C" and a good "B" (or at 80 on a numerical scale). Today a "C" is unsatisfactory and a "B" is satisfactory in reality; an average grade is midway between B+ and A- (at 90 on the same scale). Grades have gone up similarly in other colleges since AP's birth in 1956. Thus it should be evident from Table 1 that the AP scale would likewise shift by an entire grade, as it has. That the College Board misses this inference is a sign of the lack of contact between it and the colleges.

The constancy of the average pass rate at about 2/3, measured by fraction of scores greater than or equal to 3, is also illusory for a subtle reason, related to Simpson's paradox. Actually, in most AP tests, the fraction of examinations scoring at 3 or higher is decreasing over the years as the pool of test takers expands and takes in students of lower ability. The overall result appears to be constant because of shifts in test takers towards easier exams.

The number of U.S. History exams at 3 or more has declined to 51% in 1999. On the other hand, in English Literature the percentage of tests with scores of 3 or higher has held up to 68% in 1999. This result reinforces other evidence (see Table 2, comments) for declining grading standards in English Literature *vis a vis* other subjects. However, for both exams only about 40% of test-takers truly qualify for colleges AP. How could the quality of AP exam papers slide downward so badly? An explanation given by three authors, one of whom (Jones) is the present head of the AP program, is that

> ... over long time intervals test scores are not necessarily comparable, as the entire scale may gradually shift. Changing demographics of the test-taking population must also be considered.... (Pfeiffenberger, Zolandz and Jones, 1991)

Since the number of tests has increased five hundred-fold during the past 45 years, one should not be surprised at such a drift.

Another sign of the CEEB-college gap is the lack of qualified graders. To keep AP's *raison d'être*, one would want at least a majority to be college faculty who teach the subject matter of the AP examinations. Yet, of 556 graders in the 1999 AP U.S. History exam, 316 came from high schools and 60 from community colleges, unaccredited and other non-college sources, or colleges which failed to list their average SAT scores (typically very low-level institutions). Only a minority of 180 came from accredited four year colleges. Likewise, of 619 graders of the English Literature examination for 1996, only a minority of 269 were 4-year college teachers. (Note 4) An unfortunate outcome of this loss of contact is that the AP program seems to have lost its major source of quality, its close collaboration with the colleges.

Mandates

A serious source of disagreement between College Board and higher education faculty is the increasing number of legal restrictions. The colleges view these as micromanagement by unqualified lay persons which endangers the high quality of American higher education. In the words of two former University presidents:

> An important reason why American higher education has become pre-eminent in the world is the greater willingness of the government to respect the autonomy of colleges and universities and to refrain from imposing its own judgements on what Justice Felix Frankfurter once described as "the four essential freedoms of a university- to determine for itself on academic grounds who may teach, what may be taught, how it should be taught, and who may be admitted to study." (Bowen and Bok, 1998)

The College Board takes the opposite point of view and welcomes this type of government intervention as an aid to program (and revenue) growth:

> Because of the leadership shown by the legislators and educators in these

> states, the growth in their students' participation in the Program has been truly remarkable.(College Board, 1995)

Examples of State Mandates

Extra credit for AP courses. The state Regents have overridden a vote of the University of California, Berkeley faculty and have mandated that admissions staff give a full grade point extra credit for AP courses (Sahagun and Weiss, 1999). Extra credit towards admissions (in the University of California and others) also is based on *enrollment* in courses with the label "AP," not necessarily on satisfactory exam *performance*. Since the overall examination pass rate is only about one for every three course enrollments, mandating preferential admission to *enrolled* students is questionable.

Paying of examination fees. In the view of college faculty graders, the practice of some states' paying all examination fees indiscriminately encourages unqualified persons (even those who have not taken the AP course) to take a flyer and overloads the system with inferior examinations. As an extreme example, graders tell of examination papers that are totally blank, except for a message saying that the student took AP because of external pressure from parents or school. Since nothing was lost because the fee was prepaid, the student took the path of least resistance and handed in the blank exam.

Requiring that AP courses be given in all high schools. College faculty and deans cast a jaundiced eye on mandatory high school participation, which they view as dragging in schools that are unqualified to handle AP. As pointed out by the author and H. Wainer (2000), there are schools that fail even to produce a single "3" on any AP exams. In corroboration, Table 4 shows that states that pay student fees and require all high schools to offer AP tend to be at the bottom of the list.

Mandating acceptance of AP examinations with a "3" or higher. The College Board's qualification estimates (Table 1), backed by mandates in a growing number of states, would require acceptance into advanced courses of candidates with a score of "3". This would be unacceptable to colleges that no longer honor a "3". If these mandates were accepted, it would rob the colleges of the discretion to place students on the basis of all relevant information, not just a single, obsolete, numerical grade. That AP success could be a self-fulfilling prophecy follows from this scenario:

1. AP is seen as a successful, growing program.
2. The State wishes to improve its educational system.
3. College Board assures AP quality and the value of a "3."
4. On this cue, the State mandates college credit for a "3."
5. Colleges comply; the great majority of examinees get AP credit.
6. Enrollment in AP courses soars.

This scenario is a closed loop that includes the College Board and the State government. Out of the loop are the college faculties. Despite the CEEB's enthusiastic support of these mandates and its growing success in gaining state support, it is safe to predict that the colleges will resist. In the words of Bowen and Bok (1998),

> "... it is very difficult to stop people from finding a path toward a goal in which they firmly believe..." and efforts to impose solutions on the colleges are "likely to bring forth ingenious efforts...that can have a wide variety of other consequences, not all of them benign."

University faculty can use a variety of measures to circumvent state mandates on AP. Private universities of course are not bound by governmental rules. State universities have a harder time and do not always succeed, as is shown by UC Berkeley's well-known loss of diversity since affirmative action was voted down. However, state universities preserve quality by granting only elective credit to AP

 BEST COPY AVAILABLE

scores of "3." Another strategy, as discussed later in this article, is to place AP students in standard beginning classes, rather than in remedial courses. Nevertheless, the pressure from mandates is on college faculty either to go along and lower quality or to misreport their AP policy. In either case, Table 2 would be incorrect.

Table 4
Advanced Placement Scores by States

State	Number of Tests per 100 grads	Performance		Mandates*
		%=>3	%=>4	
DC	83.7	73.4	49.5	
Missouri	13.6	74.6	44.3	
Connecticut	47.0	72.1	43.8	
Massachusetts	46.7	72.0	43.4	
New Jersey	42.3	70.6	42.7	
Illinois	33.3	72.3	42.7	
Hawaii	34.9	67.2	41.6	
Maryland	48.0	71.5	41.5	
Delaware	40.5	71.2	41.4	
New Hampshire	32.4	70.4	41.3	
California	55.9	65.7	37.5	
Rhode Island	29.8	69.4	37.4	
North Dakota	9.3	72.1	37.2	
Tennessee	24.2	64.7	36.5	
Washington	23.6	68.4	36.5	
Wisconsin	29.6	68.3	36.4	
Iowa	14.2	70.0	36.3	
Montana	17.1	66.9	36.3	
Pennsylvania	27.1	65.7	36.0	
Virginia	56.7	65.6	36.0	C
Louisiana	10.8	63.8	35.3	
Colorado	36.5	66.3	35.2	P
United States	**36.6**	**64.1**	**35.2**	
Utah	63.5	67.6	35.1	
New York	62.4	64.1	35.0	
Oregon	19.9	67.1	34.9	
Ohio	24.5	65.5	34.9	
Wyoming	8.1	63.7	34.8	
Maine	26.1	67.4	34.4	
Kansas	13.7	64.6	34.3	
Michigan	26.8	65.3	34.0	
Vermont	31.6	64.5	33.9	
Idaho	16.2	67.1	33.5	
Arizona	27.5	63.0	33.1	
Georgia	34.0	60.3	32.6	P
Alaska	39.2	63.6	31.3	
North Carolina	42.6	59.9	30.9	
Texas	38.0	57.8	30.8	
Nebraska	12.1	62.7	29.9	
Florida	54.5	56.2	29.5	P

Minnesota	28.6	58.6	29.1	P
New Mexico	21.9	56.1	29.1	
Oklahoma	19.7	58.8	28.9	
South Carolina	44.5	55.1	28.5	C, P
Alabama	21.0	57.3	28.3	
Nevada	31.7	56.0	26.2	
West Virginia	15.7	55.2	24.3	
Kentucky	23.5	50.7	24.2	P
South Dakota	16.5	55.5	24.0	
Arkansas	15.3	52.0	23.9	
Indiana	21.6	50.2	23.4	C, P
Mississippi	14.2	45.5	19.9	

*Mandates: P= State pays fees for all AP examinees
C= All schools required to give AP courses

How AP Actually Performs

The College Board's literature has emphasized the positive aspects of the increase in *numbers* of test takers, but has paid less attention to actual *performance* of AP students (College Board, 1994, 1995, 1996, 1998). Consider some data (obtained from ETS) on actual choices made by students in calculus in 14 colleges. One finds the following distribution in Table 5.

Table 5
Actual Placement of Calculus Students in 14 Colleges

AP Score in Calc AB	Percentage taking first calculus course at level shown				
	No Course	Remedial	1st Calc	2nd Calc	3rd Calc
No AP exam	29%	45%	21%	3%	1%
3	24%	17%	37%	22%	2%

Note that the majority of incoming students without an AP background either took no math or enrolled in a remedial course. Also, only a small fraction (22%) of students with a score of 3 ("qualified" in Table 1) actually took an advanced course, although the majority (61%) placed out of the remedial course. This shows that, for scores of "3" and lower, the AP Calculus AB examination is no longer acting as an advanced placement, but more as a placement examination. (Students with a score of "1" or "2" usually are placed in the remedial course. Students with a score of 4 or 5 are likely to take an advanced course.) If one considers the overall performance of all AP students who finished Calculus AB and estimates that *ca.* 2/3 actually took the exam, *only a quarter or less achieved advanced placement* in this sample.

Especially in the foreign languages, colleges often use AP exams interchangeably with other criteria, such as SAT I and SAT II scores and even high school credits, to make placement decisions.

AP and Minorities

AP results for minority students are disturbing. (Note 5) The author finds College Board statements on this topic misleading (College Board, 1996; Coley and Casserly, 1992). For example, CEEB cites the movie "Stand and Deliver" on Escalante's success in teaching AP calculus to Hispanic children. However neither Escalante nor his emulators have succeeded in repeating his success with minority students. (Lichten and Wainer, 2000; Mathews 1988, 1997, 1998; Woo, 1998). Furthermore, most of his students took the AP Calculus AB exam, much of which is high school level material.

In the College Board's words (1996),

> Woodrow Wilson High School (Washington, D.C.) provides an excellent example of a predominantly minority urban high school with a well established Advanced Placement program that serves a substantial proportion of its students.

In actuality, in 1998, out of a total of 383 AP examinations, 85 were taken by African Americans, of which 18 received a "3" or higher (estimated 6 or 7 for "4" or higher). In its press releases, CEEB often quotes the increased number of minority students taking AP exams, but says nothing about their success rate. Consider the facts on minority AP performance. If a passing grade were 3, 35% of African-American AP examinations would qualify. A shift to a "4" would lower this to 14%, or one out of seven exams. These results are consistent with PSAT-AP ability-performance relation (Camara, 1997; Lichten and Wainer, 2000). Minority students typically score about one standard deviation (15 I.Q., 6 ACT, or 100 SAT points) below average, which translates into an AP pass rate of about half of that for majority pupils.

In urban school districts, such as Detroit, students in selective high schools perform well on AP exams. On the other hand, the much larger number of pupils at unselective schools do extremely poorly in the AP program. In some, not a single AP candidate passes the exam (Lichten and Wainer, 2000).

In the late 1990's more than 2 million persons graduated each year from high school, of which about 1 million (40%) went to four year colleges. About 400,000 took AP exams (18%). About 200,000 (9%) scored at "3" or higher and approximately 100,000 (4%) scored at "4" or higher. For African-Americans, the corresponding figures were about 250,000 graduates, 75 thousand (30%) to four year colleges, 15,000 (6%) AP exams, 5,000 (2%) passed at "3" or higher (less than 1% at "4" or higher). AP success occurs for a small fraction of high school graduates; for minority students, the fraction is extremely small.

In lawsuits on behalf of African-American, Hispanic and Filipino- American students, six civil rights organizations have charged the University of CA with discriminatory admissions policies. The suits cite the practice of giving extra credit for AP courses to college applicants and the lower availability to minority students of AP courses. (Berthelsen, 1999; Nieves, 1999; Rosenfeld, 1999; Rios, 1999; Sahagun and Weiss, 1999; Daniel *et al. vs.* State of CA *et al.*, 1999). UC claims to take into account inequality of opportunity for honors/AP students, but state mandates prohibit such discretion (Sahagun and Weiss, 1999). Clearly, admission policies that favor AP participants work against minority pupils. Affirmative action, in which lower test scores for minorities do not exclude them from admissions to selective colleges, is of proven benefit (Bok and Bowen, 1998).

Other Low Performing Groups on AP

Not just minorities are disadvantaged on the AP examinations. Table 4 shows large differences in AP performance among the states. Poor, rural states usually show low AP scores; wealthy, urban states generally do well. Thus, Washington, D.C. is at the top of the table (Note 6); IN does poorly. Preference on college admission to students in AP classes means students from low performing states and schools will be handicapped.

Common Sense and AP

There are few lasting success stories in American Education (Tyack and Cuban, 1995). As effective educational programs spread, the imitations often become less true to the original. A law of diminishing returns sets in as the originally well-qualified (often self-selected), well- informed and highly motivated group of teachers and pupils becomes flooded by the deluge of badly qualified, ill- informed and poorly motivated followers. The program becomes less selective and quality declines.

AP is no exception to the rule. Consider the largest AP program, English

Literature. From Haag's (1985) data, the average PSAT-verbal score of test takers in 1982 was an estimated 62 (recentered scale), far above average. By 1997, from Camara's (1997) data, the average had declined 9.5 points to 52.5, which is close to average (approximately 50 for the PSAT), an exceptional loss of selectivity . (The 50% success point for AP English Literature on the PSAT is 45, well *below* average.) To claim that quality could be maintained in the face of such dilution of the examination taker pool would be incredible. (Other programs, such as U.S. History, have been more selective.)

College introductory courses match the level of average students. Below average students take remedial courses. Only the small minority of above average high school students capable of doing college level work are suited to the AP program. As the AP program expands, it reaches students who are not yet ready to do college-level work. The data confirm common sense: only a minority of students are capable of doing college-level work in advance. Otherwise, standard introductory college courses would be unnecessary.

In confirmation, a survey of K-16 (school and college) students by the Education Trust (1999) showed the high school-college gap. Three quarters of U.S. high school graduates enter some kind of college, but many arrive unprepared. Nearly half take a remedial course, one third fail to make it into the sophomore class, and less than half graduate from college. With few exceptions, national and state standardized tests fail to cover the abilities needed in college. In the Trust's words, it "doesn't make any sense" that the fastest growing courses in high schools are college level (AP) and the biggest growth in college courses has been high school level, remedial courses. (Note 7)

In summary, *the major slide in the qualification scale,* the heart of AP, *results from lower average student ability.*

Whither AP?

The College Board endorses continuing the expansion rate of AP for the next decade (College Board, 2000). What would be the outcome of this policy? Classical economics says that the decision to increase production hinges on the *marginal rate of return*. Additional production increases profits up to the point of diminishing returns, after which losses outweigh gains. There are also intangible limits on expansion. If a farmer plants to the point that the grain becomes poor in quality, or the land is damaged by erosion, the damage to his/her reputation or land may not show in dollars and cents, but it could be important in the long run.

Likewise, expansion of the AP program reaches diminishing returns, as the marginal yield of pupils qualifying drops (Table 6, last column). In lieu of hard data (CEEB does not keep records of actual number of qualified examinations), this table is based partially on Table 2 (for the year 2000) and information the author could glean from various sources. Table 6 is based on a conservative projection of present trends, such that all selective colleges will no longer accept a "3". Actually, some colleges now require a "5" for AP in some subjects and some give no AP credit for English Literature.

Table 6
Estimated Diminishing Returns in the AP Program

Year	Number of Exams	% qualifying	Qualifying Examinations		Increase in Total Number	% of added number qualifying
			Number	Increase		
1960	15,000	75%	10,000	—	—	—
1970	70,000	75%	50,000	40,000	55,000	75%
1980	150,000	69%	100,000	50,000	80,000	63%
1990	500,000	60%	300,000	200,000	350,000	57%
2000	1,400,000	48%	650,000	350,000	900,000	39%
2010	2,300,000	35%	800,000	150,000	900,000	17%

Table 6 shows how further increases add relatively few qualified examinations. On the other hand, the costs mount in terms of examination fees, training teachers, smaller class sizes, lowered quality of graders and loss of respect for AP. The net benefits diminish to the point that continued expansion of the program does more harm than good. In the opinion of the author, that point was passed long ago.

Conclusions

A fundamental flaw in the AP program follows from the failure to distinguish between *criterion* and *norm referenced* programs. Norm referenced programs, such as SAT or ACT, put students in rank order for convenient sorting. The larger the number of parsons taking such a test, the better are the norms.

On the other hand, the colleges' AP criterion is inflexible. As long as AP served a small, elite population chosen from selective schools, increasing the program size had little or no effect on the pass rate or on quality. Now that the level of test takers has dropped below the criterion, the failure rate has increased sharply, and program quality has suffered.

To reestablish quality, major reforms to AP are needed. These include an honest grade scale which is aligned with college standards, removing unwise mandates, and better selection of faculty and students into courses, examinations and grading. (Note 8)

Notes

The author is indebted to Neil Dorans, Drew Gitomer, Penelope Laurans, Maxine Lurie, Jonathan Lurie, L. Scott Miller, Rick Morgan, Len Ramist, Howard Wainer, and Warren Willingham for helpful discussions, suggestions, criticisms and information. This paper was partially researched while the author was a visitor at the Educational Testing Service, 1998-1999. This paper is not approved by, and does not express the views of the Educational Testing Service nor of any of its employees.

1. If colleges were arranged by SAT, rather than by AP scores, the grouping would be slightly different. For example, Spelman College would be listed as selective.
2. This result (actually 26.5%) is robust. For example, if "possibly qualified" meant one quarter of the students passed, the resultant shift would be 27.1%.
3. W. Currie, quoted in Rothschild (1999), estimated about 55% of students enrolled in AP courses take the examinations. The more conservative figure of 2/3 used here changes the fraction of AP enrollees passing the tests to about a third.
4. Community college faculty do not have direct contact with the AP program and the content of AP-level college courses. They and high school teachers usually do not have the advanced education and research experience of college and university faculty.
5. Data for African-American scores in AP tests are from 1998 figures from ETS. The present paper does not consider Asian-Americans as "minority."
6. Washington, DC has a higher per capita income than any state. Also the

overwhelming majority of AP tests there are taken by students (majority as well as minority) from private schools.

7. A similar inversion occurs between AP English Literature and SAT II English. The former has average PSAT scores of 52.5 (roughly comparable to senior SAT scores of 540); the latter has average SAT scores of 568 (College Board, 1997). Students taking the AP exam have lower verbal ability than those who take the high school exam.
8. The College Board (2000) recently announced plans to put ten AP courses in every high school in the country by the year 2010 and expand the program to over 2 million examinations. This move, if it ever became real, would exacerbate the problems of the program: bloated size, ill-qualified faculty and students, and growing failure rates, especially among minorities. Calculus BC is the exception that proves the rule about AP. This small program (31,000 exams in 1999) is still a success by all measures. Colleges still accept a "3" for AP, the pass rate is very high (79%), yet the student ability distribution on the PSAT is no higher than for calculus AB (Camara et al, 1997). The success of BC may be due to the same features of AP in its early days: self-selected, able, well-motivated faculty and students.

References

Berthelsen, C. (1999). Suit Says Advanced-Placement Classes Show Bias. *The New York Times. July 28, p. A18 YNE* (Western edition).

Bowen, W.G. and Bok, D. (1998). *The Shape of the River. Long-Term Consequences of Considering Race in College and University Admissions.* Princeton, N.J.: Princeton.

Bracey, G. W. (1991-1998). Why Can't They Be Like We Were? *Phi Delta Kappan* (Oct.), pp. 104-117, annual Bracey Report on the Condition of American Education, 1992-1998.

Bracey, G. W. (1997). Setting the Record Straight: Responses to Misconceptions about Public Education in the United States. Alexandria, VA: Association for Curriculum Development.

Camara, W. (1997) The Relationship of PSAT/NMSQT Scores and AP® Examination Grades. RN-02, November. New York: College Board.

Carson, C.C., Huelskamp, R. M., and Woodall, T.D. (1993). Perspectives on Education in America. J. Educational Research 86, No. 3, May/June, pp. 259-311.

Coley, R.J. and Casserly, P. L. (1992). *A Study of AP Students in High Schools with Large Minority Populations.* Princeton, NJ: The Educational Testing Service.

College Board. (1994). College and University Guide to the Advanced Placement Program. New York: The College Board, p. 32.

College Board. (1995). *Advanced Placement Program. 40th Anniversary Yearbook.* New York: The College Board.

College Board. (1996). *A Secondary School Guide to the Advanced Placement Program.* New York: College Entrance Examination Board and Princeton NJ: Educational Testing Service.

College Board. (1997). AP. Grading, Interpreting and Using Advanced Placement Examinations. New York: College Board and ETS.

College Board. (1998). *Advanced Placement State Initiatives.* New York: College Board.

College Board. (1999a). *More Schools, Teachers, and Students Accepted the AP® Challenge in 1998-1999.* Collegeboard.org/990831b.html

College Board. (1999b). *The College Handbook 2000.* New York: College Board.

College Board. (2000). HTTP://WWW.Collegeboard.org/press/html/index.html.

Daniel *et al. vs.* State of California *et al.* (1999).

Los Angeles: Superior Court of California.

Education Trust. (1999). *Thinking K-16.* Washington, DC.

Finn, C.E. Jr. (1991). *We Must Take Charge. Our Schools and Our Future.* NY: Free Press.

Haag, C. H. (1985). Using the PSAT/NMSQT to Help Identify Advanced Placement Students. Publication 273678. New York: College Board.

Hyser, R. H. (1999). Is a 3 a C? The Reliability of the Advanced Placement United States History Test for College Credit. *The History Teacher, 32,* No. 2, February, 223-235.

Lichten, W. and Wainer, H. (2000). The Aptitude- Achievement Function: An Aid for Allocating Educational Resources, with an Advanced Placement Example. *Educational Psychology Review, 12,* No. 2, 201-228. (Information available at http://www.wkap.nl//oasis.htm/222765.)

Mathews, J. (1988). *Escalante: The Best Teacher in America.* New York: Holt.

Mathews, J. (1997). A Math Teacher's Lesson in Division. Movie Inspiration Has Found That Fame Can Stand in the Way of Success. *The Washington Post.* May 21, D1

Mathews, J. (1998). *Class Struggle.* New York: Times Books.

Morgan, R. and Ramist, L. (1998). *Advanced Placement Students in College: An Investigation of Course Grades at 21 Colleges.* Princeton: ETS. Report No. SR-98-13.

National Committee on Excellence in Education. (1983). A Nation at Risk. The Imperative for Educational Reform. Washington: U.S. Department of Education.

Nieves, E. (1999). Civil Rights Groups Suing Berkeley Over Admissions Policy. New York Times, Feb. 3, p. A11.

Pfeiffenberger, W., Zolandz, A. M., and Jones, L. (1991). Testing Physics Achievement: Trends Over Time and Place. *Physics Today,* September, pp.30-37.

Princeton Review. (1998). *Complete Book of Colleges.* New York: Random House.

Ravitch, D. (1985). *The Schools We Deserve.* New York: Basic.

Rios *et al. vs.* the Regents of the University of CA, Case C.99-0525, U.S. District Court, Northern Dist. of CA.

Rosenfeld, S. (1999). Cal hit with race-bias suit. *San Francisco Examiner,* Feb. 2, pp. A1, A10.

Rothschild, E. (1999). Four Decades of the Advanced Placement Program. *The*

History Teacher **32**, No. 2, February, pp. 175-206.

Sahagun and Weiss (1999). Bias Suit Targets Schools Without Advanced Classes. *Los Angeles Times*, July 28, p. A1.

Tyack, D. B. and Cuban, L. (1995). *Tinkering Toward Utopia. A Century of Public School Reform*. Cambridge, MA: Harvard.

Woo, E. (1998). Calculus Test Scores Drop at Garfield; Film is Blamed. *Los Angeles Times*, August 19, 1:5

About the Author

William Lichten

Professor Emeritus of Physics
Fellow, Institution for Social and Policy Studies
Yale University
New Haven CT 06520-8209

Email: william.lichten@yale.edu

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina-- Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doc.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 442 times since June 29, 2000

archives | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 8 Number 30 June 29, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

The Use of Performance Models in Higher Education: A Comparative International Review

Janet Atkinson-Grosjean
University of British Columbia

Garnet Grosjean
University of British Columbia

Abstract
Higher education (HE) administrators worldwide are responding to performance-based state agendas for public institutions. Largely ideologically-driven, this international fixation on performance is also advanced by the operation of isomorphic forces within HE's institutional field. Despite broad agreements on the validity of performance goals, there is no "one best" model or predictable set of consequences. Context matters. Responses are conditioned by each nation's historical and cultural institutional legacy. To derive a generalized set of consequences, issues, and impacts, we used a comparative international format to examine the way performance models are applied in the United States, England, Australia, New Zealand, Sweden, and the Netherlands. Our theoretical framework draws on understandings of performance measures as normalizing instruments of governmentality in the "evaluative state," supplemented by field theory of organizations. Our conclusion supports Gerard Delanty's contention, that universities need to redefine accountability in a way that repositions them at the heart of their social and civic communities.

I. Introduction

In recent years, the imposition of performance models on institutions of higher education has become a widespread practice. National systems are in place in France, Britain, the Netherlands, Scandinavia, Australia, and New Zealand. In federations like Germany, the US, and Canada, individual *Länder,* states, and provinces have taken the initiative (Brennan, 1999; Woodhouse, 1996).

Performance models include, but are not limited to, social technologies like performance indicators. They are situated within broader, ideological mechanisms variously characterized as public sector reform, new public management (NPM), or what Neave, in the context of higher education (HE), calls "the evaluative state" (Neave, 1998; 1988). These mechanisms attempt to impose accountability on public sector institutions and improve service provision, by measuring performance against managerial, corporate, and market criteria.

Accountability and service improvement are common goals of all HE performance models. But different national systems adopt different combinations of supplementary goals. These include stimulating internal and external institutional competition; verifying the quality of new institutions; assigning institutional status; justifying transfers of state authority to institutions; and facilitating international comparisons (Brennan, 1999:223). The particular combination of goals depends on specific national contexts, and the balance within them of accountability, markets, and trust (Brennan, 1999; Trow, 1998).

But the foundations of these structural changes extend beyond ideological reform of public-sector institutions. They are rooted, as well, in the post-war transition from élite to mass systems of higher education (Scott, P. 1995). Arguably, the momentum of massification alone would have enforced restructuring of the HE system in most jurisdictions (Neave, 1998; Dill, 1998). The *combination* of HE expansion and the emergence of the evaluative state produces international convergence around the implementation of performance models.

Furthermore, convergence proceeds at a far-from-uniform rate. It is modulated by path-dependent national institutions and entrenched cultural traditions, and the divergent starting points of each national system. Broadly speaking, public universities in the Anglo-Saxon countries are moving from a position of strong autonomy to one of subordination to centralized, state control. For continental Europe and Scandinavia, where strong state control was the norm, more control of higher education is being ceded to the institutions.

These apparently contradictory trajectories converge at the level of institutional performance and accountability (Henkel and Little, 1999) where, as Newson (1998:113) has pointed out, "criteria such as 'efficiency,' 'productivity,' and 'accountability' are becoming embedded in the routine day-to-day decision-making that takes place in 'local' units throughout the university." At this level, the proliferation of a few dominant models can be explained, in part, by the operation of isomorphic forces within institutional fields, whereby "lead" organizations set the pace for "followers" (Powell & DiMaggio, 1983.)

Performance models have now been in place long enough for studies of consequences to be undertaken (Neave, 1998; Dill 1998). For example, a recent 15-country OECD study, under the direction of John Brennan and Tarla Shah of Britain's Open University, considers the impact of performance models in 40 participating institutions. On the basis of early analyses, Brennan (1999) reports that while impacts are conditioned by the nature of the individual institution and the distribution of authority in the HE system, performance mechanisms appear to have raised the profile of teaching and learning in HE institutions. He finds that overall impact is increased when the mechanisms gain legitimacy at the faculty and department level, and that increased centralization and managerialism is characteristic at the level of the institution. In some countries, Brennan suggests, evaluation and assessment mechanisms tilt the distribution of power away from faculty and towards senior managers and administrators. But in other countries, where the management layer is traditionally weak, the impacts of external evaluations are more important.

A potential weakness of this otherwise exhaustive study is its reliance on institutional self-reports. By surveying a wide range of methodologically diverse studies from different national contexts, we hope to distill a robust set of findings. We first construct the theoretical framework of the "evaluative state," through which

to view the policy and administrative implications of performance models. We then consider the theoretical importance of accounting tools in performance measurement, before defining the terms and trends in performance-based HE management. Next, utilizing a comparative international format, we summarize the impact of HE performance models in the United States, England, Australia, New Zealand, Sweden, and the Netherlands. Where appropriate, we add the results of cross-national studies. Finally, we attempt to synthesize our findings into a generalized set of consequences, identifying system-level effects, technical performance issues, institutional effects and management issues, impacts on teaching and research, and on faculty and academic departments.

II. The Evaluative State

Fundamental changes in the policies and practices of most OECD countries have followed a cultural shift in the public management paradigm over the last two decades. Public sector reforms induced fundamental changes, not only in policies and practices, but also in the culture underlying the public administration of nation-states (Strange, 1996; Aucoin, 1995; Charih and Daniels, 1997; OECD, 1995; Keating 1998). This new culture took as axiomatic market-like principles of cost-recovery, competitiveness, and entrepreneurship in the provision of public services (Power 1996; Charih and Rouillard, 1997). Criteria of economy and efficiency were supported by "broad accusations of waste, inefficiency, excessive staffing, unreasonable compensations, freeloading, and so forth" (Harris 1998:137). "Rational" corporate management techniques were installed incorporating accounting, auditing, accountability, and performance criteria. The intent was not only to make public institutions less costly and more effective, but also to normalize and entrench private sector principles (Hood, 1991, 1995; Savoie, 1995; Harris, 1998). The application of these criteria to HE produced elaborate exercises in "visioning," "re-engineering," and "quality assurance," structured on the basis of transparent and auditable accountability for performance (Power, 1996).

International convergence around these ideals renders the putative retreat of the state somewhat illusory (Dominelli and Hoogvelt, 1996; Strange, 1996; Dale, 1997). Rather than regulating directly, however, the state now regulates from a distance, assuring accountability through refined forms of "remote control" or steering (Burchell et al., 1991; Barry et al., 1996; Power, 1995). Neave neatly points to the paradox: "what some regard as a lighter form of surveillance...goes hand in hand with a veritable orgy of procedures, audits, [and] instruments of administrative intelligence which, in their scope and number...make those which upheld the state-control model appear rustic" (1998:266). By using these mechanisms to steer from a distance, the state ensures its performance agenda is internalized by the institution. Thus regulation becomes self-regulation, and state control becomes self-control—a type of self-disciplining Foucault (1978) called "governmentality."

In his study of Continental European HE systems, Maassen (1997) empirically identified this move. In the countries Maassen studied, detailed regulation of the inputs and processes of HE is no longer practiced. Instead, institutions themselves create the conditions for achieving the outcomes required by the state, thereby demonstrating the effects of "remote steering" (Maassen 1997:125). To induce self-regulation and self-surveillance in institutions, Maassen found that European governments are also abandoning existing rigid legal frameworks—a move Neave (1998) calls "dejuridification"—in favour of "framework laws." Maassen suggests that European HE is undergoing the most far-reaching transition since that from élite to mass systems. What we are seeing, he speculates, might be "only the beginning of a long-term trend that will change HE far more fundamentally than we can imagine" (1997:125).

According to Neave, the beginning of this long-term trend was the emergence of the evaluative state "from two very different discourses, the one European and political, the other mainly American and economic" (1998:278). In the first discourse, control of universities mirrored broader democratic issues, while the second was a direct bid to substitute market control for state control. The former

tended to predominate in France, Sweden, Belgium, and Spain, according to Neave, while the latter dominated in the UK and the Netherlands and rooted itself earlier. Both discourses converged, Neave says, around three major displacements in HE.

One displacement is increasing concentration on strategic planning and systems development. Another marks the emergence of powerful, intermediary "buffer bodies" to serve as the state's agents in evaluation and surveillance. The third is the proliferation of increasingly demanding performance models, including quality assessment and assurance; continuous improvement; performance-based funding, budgeting, and management; strategic planning and budgeting; and total quality management. In one way or another, all these models rely on measurements or "indicators" of performance.

III. Issues in Measuring Performance

Paradoxically, the evaluative state's self-regulating "governmentality" requires fidelity devices to measure and induce compliance. Largely, these calculative practices (Miller, 1994) or rituals of verification (Power, 1995) employ accounting tools, such as budgets, cost/benefit analyses, cost-centre comparisons, financial audits, and an increasing array of performance and compliance audits (Power, 1995; Porter, 1995; Harris, 1998). Accounting tools enable "actions on the actions of others...to remedy deficits of rationality and responsibility" (Miller: 1994:29). They are characterized by their surveillance and control capacities, i.e. ability to determine norms, then discipline performance against them (Hoskin and Macve, 1993).

Despite appearances, accounting techniques and numbers are not neutral reflections of "reality." Rather, they selectively construct reality from complex webs of social and economic negotiations. An accounting "fact" is actually a contingent and partial accomplishment. Yet contingency and partiality disappear in inscription. Tabulated, calculated, and double-underlined, accounting "facts" appear incontrovertible—the very essence of stability, objectivity, and impartiality.

In a university setting, the apparent objectivity of such "facts" can undermine autonomy, "open[ing] up the routine evaluation of academic activities to other than academic considerations, and...mak[ing] it possible to replace substantive judgements with formulaic and algorithmic representation" (Polster and Newson 1998:175). A financial calculus thus underpins the discourse of performance in HE, and constitutes its instrumental logic. The instrumentalities include performance indicators, quality indices, and benchmarking standards. In a detailed study of institutions in three commonwealth countries, Miller (1995:1) found that these market-based, managerial instrumentalities "have modified or come to dominate the governance and culture of universities in Australia, the United Kingdom, and Canada". Commenting on the lack of faculty resistance, Miller argues that as academics become constrained, monitored, and documented by performance criteria, they come to collude in the construction of their own fate (cf Harley and Lowe 1999).

Performance indicators (PIs) are the key instrumentality. Watts (1992) studied the major OECD countries, looking at accountability and performance measures. Of the eight commonalities he found, PIs were by far the most significant. PIs replace traditional input measures, like the number of students enrolled, with goal- or result-oriented estimates of outcomes or value-added, such as the quality and employability of graduates. Identifying one of their most contentious aspects, Watts (1992:87) comments that "many of these efforts have found...real problems in trying to measure quantitatively the unmeasurable."

Harris (1998:136) reminds us that despite their objectified and factual appearance, much of the accounting and other data used to construct PIs derives from the subjective exercise of judgement. Similar judgements are also exercised on the indicators themselves, which are interpreted to infer "facts" that then "create the domain of the factual" (Harris, 1998: 136). Because PIs focus on readily quantifiable inputs and outputs, they tend to neglect the more complex social variables that resist measurement (Newson, 1992; Harris, 1998). And, because of the difficulty of linking measurable outputs to inputs and processes, there is a danger is that "targeted

goals, as reflected in indicators, often become ends rather than means" (Harris, 1998:136).

El-Khawas and colleagues note that "academics have resisted the move towards performance indicators, arguing that [they] are reductionist, offer inaccurate comparisons, and are unduly burdensome" (1998:9). As a result, she notes, some governments are introducing PIs incrementally, requiring universities to generate an increasing amount of quantitative data for intermediary bodies. Others have embedded PIs in institutional contracts or other forms of conditional funding. While debate continues on their appropriate use, she says, in most countries public officials advocate the development of a few relevant performance indicators, together with comparisons among institutions and over time. She differentiates England, which "took a further step by linking the amount of research funding to performance scores of academic departments" (El-Khawas et al., 1998:9). In the studies cited later, we will find more variation than El-Khawas suggests in the numbers and types of indicators tracked. We will also see that the pattern of linking funding to performance extends beyond research to HE budgets more generally. And we will find performance-linked funding in, for example, the United States, Australia, and New Zealand as well as in England.

While there is no single, agreed-upon definition of PIs, the one developed by Cave, Hanney, and Kogan (1991:24) is still applicable:

> a performance indicator is an authoritative measure—usually in quantitative form—of an attribute of the activity of a higher education institution. The measure may be ordinal or cardinal, absolute or comparative. It thus includes both the mechanical applications of formulae (where the latter are imbued with value or interpretative judgements) and such informal and subjective procedures as peer evaluation or reputational rankings.

One of the principal causes of controversy surrounding the use of PIs is their link to performance-related funding and budgeting. It is important to differentiate between these terms. According to Burke and Serban (1998:2), "the advantages and disadvantages of each are the reverse of the other. In performance funding, the tie between results and resources is clear but inflexible. In performance budgeting, the link is flexible but unclear." Performance funding ties separate and usually small allocations of funding directly to institutional performance against a normally limited number of indicators. In performance budgeting, a longer list of indicators provides an overall picture of institutional performance; this then supplies the context in which a decision on the institution's *total* budget allocation is made. The former enhances the incentive to improve performance, but punishes circumstances beyond institutional control. Further, the small sums allocated are disproportionate to the effort required to generate the data. The flexibility of the latter allows for extenuating circumstances, but diminishes specific incentives to improve (Burke and Serban, 1998.)

Johnstone (1998) confirms these differences, and notes that both are rooted in conceptions of administrators as "rational actors" who will maximize whatever is rewarded. According to Johnstone, conventional budget drivers—particularly full-time equivalent enrollments—induce institutions to "over-enroll" at the cost of quality and can lead to a concentration on popular programs that can be taught cheaply (1998:16). In contrast, performance-based budgets use criteria such as degrees awarded, time to completion, graduates' external performance, faculty success in attracting competitive research grants, and faculty reputations with peers. However, says Johnstone, proponents of performance criteria are beginning to realize that there is a need to balance "multiple, difficult-to-measure, and not always compatible goals" (Johnstone, 1998:16). For example, to maximize student accessibility, institutions are encouraged to accept promising but less-qualified students. This goal is incompatible with maximizing completion rates or postgraduate examination performance.

The offsetting advantages and disadvantages of performance funding and performance budgeting helps to explain why increasing numbers of states in the U.S.A. are adopting both systems (Burke and Serban, 1998). While examples of

performance models could be found in some states (e.g. Tennessee) as early as the 1970s, by 1998 they were utilized in half the states in the U.S.A. Reported intentions predict that 70% of states will have adopted performance funding or budgeting models by 2002 (Burke and Serban, 1998).

There is more than rational judgement at work here; a "bandwagon" is rolling. Organizational theory assists our understanding of this phenomenon. Powell and Dimaggio (1983), for example, have pointed to the role of isomorphic forces in stabilizing institutional and organizational fields around a dominant model. The forces at work may be regulative, normative, cognitive, or any combination thereof, depending on the nature of the field (Scott, R. 1995). Thus the particular combinations of state policy, programs, and funding (regulative); academic values and norms of accountability (normative); and the way the social purpose of HE is framed (cognitive) might be expected to produce fairly similar institutional responses to performance criteria that may, nevertheless, differ in important respects in different national and sub-national contexts.

Further, formal organizations like universities and colleges tend to adopt prevailing "rituals of rationality" to increase their legitimacy and chances for survival (Meyer and Rowan, 1977; Kaghan, 1998). These rituals of rationality increasingly include principles of profitability and "good management" derived from the private sector. Public universities and colleges, therefore, can be situated in a larger institutional framework where the *system* of organizations is isomorphically aligned around ideological commitments to private sector principles of rationality.

But as Kaghan (1998:172) points out, institutional theories tend to focus at the macrostructural level and pay little attention the "microdynamics" of specific practices. To attend to this level of detail, we now consider the way performance models are enacted in different national contexts. A comprehensive examination of US and UK experiences is followed by less detailed analyses of Australia, New Zealand, Sweden and the Netherlands.

IV. Performance Models in Context

1.State Models in the United States

Policy-makers in the U.S.A. were among the first to experiment with monitoring the performance of publicly funded institutions of higher education. In the 1960s and 1970s, state officials began examining possibilities of allocating resources to institutions according to how well they achieved state objectives and outcomes (Layzell, 1998).

Tennessee was the first state to implement performance funding in higher education. Well regarded in the US, the program is considered a success. The Tennessee State Higher Education Board initiated a pilot program in 1975. By 1979, state officials, working with advisory groups, had developed a set of ten performance criteria. These, and the associated measurement and reporting procedures, were applied to all public universities and colleges (El-Khawas, 1998). During 1980-81, public institutions were able to earn up to 2 percent above formula allocations, based on performance against these criteria (Albright, 1997). The plan has been reviewed and updated at five-year intervals since then. Today, the amount of discretionary funding available to reward good performance stands at 5.5 percent of an institution's overall budget. Explicit goals are targeted over an extended period of time, allowing institutional behaviour to be shaped towards desired ends.

Because of isomorphic forces, the success of the Tennessee program led to the development of similar programs in Arkansas, Missouri and Ohio (El-Khawas, 1998). But conformity is far from total. Texas is among several states that have studied, proposed, and rejected performance funding—largely because of a lack of support from state legislators, combined with cumbersome reporting requirements, and reduced institutional autonomy (Albright 1997). On the other hand, the State of South Carolina has adopted measures that ties allocation of the state's entire budget for public higher education to institutional performance against 37 specific indicators (Burke and Serban, 1998) .

One notable characteristic of Tennessee-style performance funding is that it is

BEST COPY AVAILABLE

non-competitive. All institutions can access these supplemental "bonus" funds. If one fails to obtain its share of the supplementary funds, the others do not benefit. Generally, however, policy-makers today are less favourably inclined to voluntary institutional improvement; systems of mandated public accountability are becoming the norm. As with the introduction of the Tennessee model, we see a tendency to copy other states' systems, in an attempt to develop a common core of indicators to address common problems.

A study by the National Association of State Budget Offices (NASBO, 1996) reviews measures adopted by 38 states in addressing calls for HE improvement and accountability. These include budget reforms, restructuring of governance, performance-based funding, and privatization of teaching hospitals. We cannot report on this study in detail, or present the responses of all the participating states. However, certain states can be considered "indicators" of the changes induced by performance models in all states.

Arizona's Budget Reform Act of 1993 resulted in the development of a master list of state government programs in 1995, complete with mission statements of institutions, functional program descriptions, goals, performance measures, funding and staff information. This was the first opportunity for state analysts to determine budgets and funding sources for higher education. Subsequently, in an attempt to increase graduation rates without increasing the budget, a "short" Bachelor's Degree program (three-years) was implemented at Northern Arizona University. As well, certain programs implemented a twelve-month academic year. Faculty could elect to take their break in either fall or spring instead of summer. To ensure a steady supply of enrollees, the Arizona Legislature introduced a bill to provide HE scholarships to students who graduated high school in three consecutive academic years and retained a GPA of at least 3.0 (out of 4.0). State funding would be shifted from the K-12 system to the HE system to fund the new measures.

In 1995, Arkansas moved from an enrollments-based funding policy to one focused on productivity outcomes. The Institutional Productivity Committee and the State Board of Education developed sixteen performance measures. Amendments to the Revenue Stabilization Law resulted in the creation of a Higher Education Institutions Productivity Fund, authorized to provide an additional $5 million and $10 million in fiscal years 1996 and 1997 respectively, on the basis of institutional performance on these measures.

Also in 1995, the Governor of California agreed to provide lump-sum funding to the University of California, and California State University for a period of three years for general support, capital outlays, and to service debt requirements. In exchange, the universities were required to increase enrollments and the portability of courses between institutions; implement new productivity and efficiency increases each year; improve student graduation times; and restore faculty salaries to competitive levels. Meanwhile, in the Kansas fiscal 1997 budget, and the Kentucky 1994-1996 Appropriations Bill, appropriation increases to higher education were based on performance funding concepts and principles.

On July 1, 1995, Minnesota merged three of the state's public, post-secondary systems under a single governance structure. For 1995 and 1996, a portion of state appropriations to the University of Minnesota and the state's colleges and universities was made contingent upon achievement of performance goals. For example, for the University of Minnesota, $5 million of the 1996 appropriation was placed in a performance incentive account, to be released in $1 million increments for achieving each of five performance measures. The measures related to: a) recruitment and retention of freshman students with high academic averages in 1995; b) increase in the intake of minority students in 1996; c) increase in the number of women and minority faculty hired in 1995-96; d) increase in graduation rates between 1994 and 1996; and e) increase in the number of credits offered through telecommunications between 1995 and 1996.

Missouri adopted policies that ensure the recognition of institutional performance through appropriate incentive funding. In fiscal years 1995 and 1996, funding was appropriated to reward institutions based on their attainment of certain goals: a) assessment of graduates; b) graduation of minority students; c) number of students pursuing graduate education; d) teacher-education graduates scoring in the upper half of national exams; and e) job placement rates in major field. In fiscal year

1996, more that $7 million of the ongoing untargeted funding for four- year institutions was distributed according to these performance goals.

While other states, including New Mexico, New York, North Carolina, North Dakota, Oklahoma, South Carolina, Utah, Washington, and Wyoming have all undergone budget reform, restructuring, and the implementation of performance measures, none has gone to the extreme of South Carolina. In 1996, at the urging of a group of prominent business leaders, the State Commission for Higher Education implemented the most significant performance-based funding program to date. The program was phased-in. By the 2000 fiscal year, as stated earlier, 100% of state HE funding will be allocated on the basis of institutional performance on 37 specific indicators. This high number of indicators, as well as the total linking of funding to performance, runs counter to conventional wisdom on performance models.

Agendas beyond Performance

The above review of performance models makes evident the extent to which they can be used to advance state agendas other than those strictly concerned with accountability and performance. In the case of Minnesota and Missouri, for example, performance models are used to address state requirements for equity and equality in public institutions. Thus the state can use these models to force HE institutions to advance compliance with long-range state objectives. If the institutions successfully comply, they are rewarded. Otherwise, there is an implicit threat that the state will step in and take control of budgets and governance structures. But state policy is subject to change with each election. In between, there may be insufficient time for political objectives to be fully integrated into an institution's governance and funding structure.

A recent study by the State Higher Education Executive Officers (SHEEO, 1997), provides a snapshot of the experience of 48 states in implementing performance measures. The study indicates that:

- thirty-seven states used performance measures in some way
- this is more than double the number three years previously
- twenty-six states plan to expand or refine current efforts
- most states adopt performance measures for accountability purposes
- twenty-three states use performance measures to inform consumers about higher education
- twenty-three states use performance measures to distribute state funds to higher education institutions (Network News, 1998:1-2).

Most of the performance models referred to in this study fail to differentiate between longer-term state interests and short-term public demands. As well, in the twenty-three states where performance measures supply information to consumers of HE, the information reported is deemed more useful to policy-makers, than for assisting individual consumers to make informed educational choices.

Responses to US Performance Models

The SHEEO and the NASBO studies cited above seem to indicate a shared understanding between state officers and HE institutions about the importance of performance models. This may not be the case. In a survey of higher education policy issues (Ruppert, 1998), a total of 1008 respondents, consisting of political leaders (n=519) and higher education leaders (n=489), from 12 Midwestern states were asked to identify the most critical issues facing post-secondary education in the approach to the 21st century. Keeping higher education affordable was a major concern for both groups, but political leaders ranked it as their first priority, while higher education leaders ranked it second. Overall, how to pay for higher education (funding policies) was considered the Midwest's second highest priority. For higher education leaders this was the number one priority, while political leaders ranked it sixth out of nine issues. Capacity for change was the third priority for higher education leaders while political leaders ranked this item fifth. Not surprisingly, political leaders ranked ensuring accountability second, and productivity and cost efficiency third priority, while higher education leaders ranked these sixth and

eighth respectively. With such disparities on the relative priorities of key issues, will the two groups support one another? Or is the stage set for increased tensions, in the form of either active or passive resistance to state mandated measures?

In analyzing responses to the SHEEO survey, Albright (1998) reports that in states implementing performance-based funding, HE institutions accrue certain advantages. They benefit from increased communication with, and support from political leaders; the funding provides an alternative to enrollment-based subsidies, and acts as an incentive to improve performance. By aligning planning goals with budgets, institutions can respond to calls for accountability and reinforce confidence in higher education. However, the design and implementation of a performance model is not accomplished without difficulty. Ways must also be found to balance decreasing institutional autonomy and increasing state review and control. Qualitative methods must be used to supplement quantitative measures when studying institutional processes. There is a need to overcome the complexities of measuring "quality," particularly as it pertains to student learning, and to find measures that adequately reflect differences in institutional missions. While some states have been more successful than others in introducing performance measures, it is still too early to attempt to identify a single "best" US model.

In terms of future prospects, a survey of state finance officers reports data on legislative action plans for 1999 (McKeown-Moak, 1999). From the perspective of state officials, the financial outlook for US higher education is better now than in years. State appropriations reached the highest level ever in FY99, increasing four times faster than the Consumer Price Index. HE's share of state general funds increased for the first time in over a decade. Average tuition fees are rising steeply. State officials proclaim that such positive economic conditions for higher education have not existed in the last two decades.

At the same time, administrators in HE institutions prepare for reduced appropriations and increases in the use of performance models. Student debt loads continue to rise at an alarming rate, and institutions that originally welcomed new federal tax credits now face the added costs of compliance and record keeping. Added to this, are increased competition for state resources; demands for up-to-date curricula that keep pace with the economic and market change; approaching reelection campaigns for state legislators; tensions with faculty and staff about internal restructuring to accommodate performance criteria; and threats to restructure HE governance. Taken together, these factors indicate the prospect of continuing struggle for US higher education leaders.

A final note: Congress enacted changes to the Higher Education Act in October 1998. Beginning in the 2001 academic year, colleges and universities must submit comprehensive reports on attendance costs for students, to the National Committee on the Cost of Higher Education. NCHE will then publish trend information on tuition fees and financial aid by institution, and compare this information with the Consumer Price Index. Failure to comply will net the recalcitrant institution a fine of $25,000. Compared to the burgeoning costs of reporting, some might consider the fine the more fiscally prudent option for financially- starved institutions.

2. England

In England, performance models were first introduced in the early 1980s as an ideological initiative of the Thatcher government. Continuing under Thatcher's successor, John Major, they then, as in other countries, transcended the partisan divide into Tony Blair's New Labour administration.

A number of intermediary agencies are responsible for administering the performance agenda. These include the Higher Education Funding Councils of England (HEFCE), Wales (HEFCW), and Scotland (SHEFC), which administer the Research Assessment Exercise (RAE), and the Higher Education Quality Council (HEQC). Under the recommendations of the Dearing Report, the latter was succeeded by the Quality Assurance Agency for Higher Education (QAAHE) in 1997. QAAHE administers quality audits and the Teaching Quality Assessment (TQA).

The Research Assessment Exercises and Teaching Quality Assessments represent longstanding programs of performance assessment. Both are controversial, for various reasons. The purpose of the former is the highly selective distribution of

funding in support of high-quality research. It evaluates on the basis of perceived national and international standards. The latter justifies public support on the basis of quality and quality improvement, and rewards "excellence" in these areas. TQA evaluations are mission-dependent. They inform rather than determine funding, and are less oriented to quantitative data than the RAE, although both programs use performance indicators.

TQA indicators include student entry profiles; expenditures per student; progression and completion rates; qualifications obtained; and subsequent destinations. Institutions are assessed on six core aspects rated on a four-point scale. The RAE looks for indicators relating to research publications; research grant income; numbers of assistants and students employed; and the research environment. It rates seven categories and relies on the subjective judgements of peer panels concerning the national and international standing of the research departments assessed (Stanley and Patrick, 1998). In contrast to this "arm's length" determination, TQAs involve site visits by external assessors and encourage critical self-assessment of weaknesses as well as strengths. Much of the criticism focused on the RAE stems from the statistical ranking of institutional performance and the publication of those rankings in the media, with subsequent reputational and funding effects. Criticism is also leveled at the underlying methodology, the emphasis on outputs and the reliance on statistical data rather than qualitative assessments, as well as the additional workload institutions face in complying with performance models.

The 1997 National Committee of Enquiry into Higher Education (Dearing, 1997) made performance requirements even more explicit. Dearing recommended the development of performance indicators and benchmarks for "families" of institutions with similar characteristics, on the principle that the interpretation of performance should take account of sector context and diversity. In response, the Higher Education Funding Council (HEFCE) set up a Performance Indicators Study Group (PISG) to develop indicators and benchmarks of performance, rather than descriptive statistics. The latter, while they are "helpful in the management of institutions, can only be judged in the light of the missions of institutions and do not purport to measure performance" (PISG, 1999:8). In this regard, the group comments disparagingly on the publication of "misleading and inaccurate" league tables.

In the first stage of its study, the group focused on producing indicators for the government and funding councils that would also inform institutional management and governance. Its immediate priority was the publication of institutional-level, output-based indicators for research and teaching. Process indicators, such as the results of TQAs, were rejected. By the time of its first report (PISG 1999), the group had prepared proposals for indicators relating to: participation of under-represented groups; student progression; learning outcomes and non-completion; efficiency of learning and teaching; student employment; research output; and HE links with industry. All except the latter related to both institutional and sector-levels. Responding to Dearing's concerns about interpretive contexts, the group developed a set of "context statistics" for each indicator to take account, for example, of an institution's student intake, its particular subject mix, and the educational backgrounds of students. These will allow "the results for any institution to be compared not with all institutions in the sector, but with the average for similar institutions" (PISG, 1999:6).

The next stage of the study will look at the information needs of other stakeholders, particularly students and their advisers. The third stage will respond to a call from the Chancellor of the Exchequer to improve the indicators on student employment outcomes. The PISG acknowledges that PIs in HE are "complicated and often controversial" and that "the interpretation of indicators is generally at least as difficult as their construction" (1999:12). They note that PIs require agreement about the values (inputs) that make up the ratio, reliable data collection, and a consensus that a higher ratio is "better" or "worse" than a lower ratio. The literature supports that none of these is easily negotiable nor guaranteed in advance.

Faculty Responses to Performance Models in the UK

Among faculty and at the institutional level, responses to performance mechanisms tend to follow a "strategy of accommodation" that focuses on technical

rather than normative aspects, and involves participation in the development of measures to make them "more meaningful or less harmful" (Polster and Newson, 1998). Consequences of this strategy in the UK include: the imposition of performance accounting systems for rating faculty productivity; favouring of research that attracts funding; a competitive transfer market in the CVs of "high performing" researchers; heavier and lighter teaching loads for "less productive" and "more productive" researchers respectively; an associated deterioration in teaching conditions; and a reordered system of state-appointed buffer bodies to allocate funding on the basis of externally determined criteria (Polster and Newson, 1998: 177). These elements recur in the following detailed discussion of the findings of two UK studies. Each examines the implications of performance models for faculty in English universities

Henkel (Henkel, 1997) studied seven disciplines across six different types of universities, interviewing 105 adminstrators and academics at various levels in the hierarchy. The study sought the implications of three performance policies: the research assessment exercise (RAE); the Higher Education Quality Council's (HEQC) academic audits for quality assurance; and the Higher Education Funding Council for England's (HEFCE) teaching quality assessments (TQA). In five of the universities studied, Henkel found a significant trend to "centralized decentralization"—strong central management coupled with maximum devolution of responsibility. This involved the creation of well-defined new roles at the centre, and the proliferation of non-academic support units. In part, these were to mediate the state's performance expectations and policies, now interpreted as *corporate* standards. Budgets were being devolved, usually to the department level, and the iteration between the centre and departments was deemed increasingly important. The new challenges were creating adaptation and status problems for administrators in some universities. But in others, administrative roles were expanding to meet the requirements of the new state policies. One administrator referred to his new authority to "open the black box of academic decision making" (Henkel, 1997:140).

While those at the centre spoke of iteration, individual faculty and the basic units were more aware of centralized authority. Many academics expressed "bitter resentment" about the inordinate administrative requirements necessary to comply with performance models, and strongly objected to the amount of time taken away from academic work (141). Many expressed nostalgia for the élite system, and saw the new models as attempting to compensate for the consequences of that system's disappearance. Thus, performance models were viewed as connected with "an undervaluing of individualization, excellence, and risk, espousing instead a "predictable mediocrity"" (ibid). Some also saw the new models as facilitating instrumentalism and "satisficing" behaviour on the part of students, as well linking with market values of consumerism and customer-led education. At issue as well was the emergence of differentiated contracts "based on competitiveness, insecurity, the casualization of academic employment, and...the attenuation of institutional loyalty" (142).

Henkel's findings are affirmed in a study of what Dominelli and Hoogvelt (1996) describe as the "Taylorization" of academic labour. Taylorization is achieved through the fragmentation, sequencing, and commodification of faculty work "into component parts or activities, each part being translated or "operationalized" into empirically identifiable and quantifiable indicators or measures" (79) These discrete "technical competencies" may then be "subject to cost-efficiency scrutiny and put up for tender" (79). The elimination of professional autonomy is another key aspect. Functional analysis defines "competences," which are then further defined by performance criteria—the assessable outcomes.

What are the consequences of "Taylorization" and performance models for academics? Dominelli & Hoogvelt describe increased workloads; shrinking resources; dramatic declines in social status; and truncation of functions. They cite the following statistics:

- between 1987 and 1993, student numbers in HE increased by 50% while academic staff numbers increased by only 10%, and total spending per student fell by 50%. (p.82 and fns 35 and 36)
- in the same period, core staff increased by 1.2% while staff employed on

temporary and short-term contracts increased 23% (p83)
- in the OECD, between 1980 and 1990, the UK was the only country with real negative growth in pay (-3.8%) for academic teachers (p83)

Echoing Henkel's findings, these writers suggest that the English performance model is built on the following characteristics: (1) decentralized budget management; (2) peer pressure and peer scrutiny of "performance"; and (3) flexible production techniques.

The UK's Research Assessment Exercise (RAE)

The RAE is a major and recurring evaluation of research performance. For a comprehensive Foucauldian analysis of the RAE as a routine operation of surveillance and assessment dependent on coercion and consent, see Broadhead and Howard (1998). The last RAE was 1996; the next will be in 2001. The RAE directly affects the allocation of funds from the higher education funding councils. Council research budgets have not increased for some years so, for institutions, competition for research funds is a zero-sum game with winners and losers. And, since the binary system of universities and polytechnics was unified in 1992, this "flat" amount of funding now has to be allocated to more than 40 institutions—twice the original number (McNay, 1999). Reporting on the consequences of the 1992 and 1996 RAEs, McNay found that "money was a great driver in participating in the RAE and the money that flows from it was the main means by which it exercised influence for behaviour change" (1999:192).

Institutional submissions to the RAE describe research performance and plans for each academic area, and list by area all "research-active" staff, together with details of their research output—publications, discoveries, patents, and so on. A series of panels then judge performance—by a variety of different and not necessarily compatible means—against approximately 70 criteria. The scale runs from 1 (research of little consequence) through 5 (research of international renown), to 5* (outstanding) (Williams, 1998). Funds to support research in a particular institution are subsequently calculated from an aggregation of these determinations. Units that do well have funding for the next five years, while poorly rated units try to limit the damage resulting from lost income (ibid.).

To discover the impacts of the RAE, McNay conducted 30 institutional case studies; surveyed administrative and academic staff in 15 institutions; and interviewed external stakeholders in the funding councils, industry, learned societies, and professional bodies. Overall, he finds that the RAE's impacts extend beyond funding, to affect "institutional strategies, priorities, and use of general resouces, not just those flowing from RAE (1999:199).

He reports the following *institutional-level* impacts (1999:195-6). First, he found more refinement of research policy and strategy, with research now focused in a smaller number of priority areas. Next, the research function is better managed and more efficient but administrative requirements have increased, with an increase in centralized research management and the number of committees. Third, these changes are primarily expressed through strategic policies and practices relating to research staffing. For example, some universities adopted more exclusionary recruitment criteria favouring "proven" researchers, and used the same exclusionary criteria to designate some existing research staff "non-active." Contradicting other studies, McNay finds "*some* spending on attracting "stars" [the CV transfer market] but this was marginal" (1999:196).

Next, participation in the RAE caused an organizational restructuring that gradually but effectively separated research from teaching. Research centres freed staff from teaching responsibilities and graduate schools focused on research, leaving undergraduate teaching responsibilities to the departments. Overall, 71% of unit heads reported the RAE's positive impact on research, while 62% report its negative impact on teaching. These results are hardly surprising since, as McNay states, "the Dearing enquiry takes the breach [between teaching and research] as a *fait accompli*" (1999:198).

Finally, and paradoxically, the RAE generated a virement (reallocation) of funds from higher-graded to lower-graded departments. This reallocation was policy

in several of the institutions studied. Largely, the virement is a strategic response to an anomaly in the RAE framework. RAE funding flows from "improvement." Top-rated departments have no room for improvement on the RAE scale so receive no increase in funding. But lower-rated areas can improve their performance and increase their funding. Therefore "financially, improvers were better than star performers at the funding ceiling" (McNay, 1999:196). McNay also found internal reallocations of teaching funds to support research activities.

At the *unit level,* heads of research units were generally positive about the impact of the RAE on productivity but expressed concerns about the related increase in stress. Other concerns included: inhibition of new research areas and interdisciplinary research; increasingly conservative approaches to research; and the aforementioned rupture between teaching and research. Two other issues were important at the unit level. First, concern was expressed at the rewarding of publication rather than dissemination. It was felt that the RAE focused too exclusively on prestige journals "mainly read by other academics, including panel members making RAE judgements", whereas dissemination could often be more effectively achieved through professional and popular journals read by end-users (1999:198). McNay points out that there is a risk of "the academic world...talking only to itself and so sterilising its work" (201). Second, staff management was a major issue for unit heads—both the determination of researcher status (active or inactive), and the reorganization of individual researchers into teams.

At the individual *researchers' level*, only 34% in McNay's study believed the RAE had improved the quality of their research. Most said the exercise had had little or no impact on them, apart from the stress and time-loss associated with the administration of performance exercises. Nevertheless, half now worked more in teams and about a third reported some constraint on choice of research topics. About 58% believed that the research agenda and priorities were defined by people other than researchers, "despite the peer-review process of RAE and the prominence of academics in committees of the research councils and other funding bodies" (199).

Williams (1998:1079), a medical researcher involved in leading the RAE exercise for his research group, takes a more combative stance. He believes the RAE uses "restrictive, flawed, and unscientific criteria" and produces "a distorted picture of research activity that can threaten the survival of active and productive research units". He says the exercise is "unaccountable, time-consuming, and expensive" and should be made more objective. Williams identifies a number of major flaws in the RAE: restrictive survey criteria; dubious performance indicators; loopholes and abuses; inefficiencies and unnecessary expense; subjective unaccountable panel reviews; bias towards established groups; and damage to other aspects of scholarship like teaching.

McNay finally considers a number of *system level* impacts of the RAE. Through what Williams (1998:1079) calls "the double blessing of money and prestige", and the RAE's competitive nature, the state seems to have succeeded in increasing research achievements in exchange for little if any growth in the overall research budget. However, the costs are no less real.

McNay believes the research/teaching split was at least anticipated and probably intended. Each was funded and assessed separately and held separately accountable. Staff could be designated "teaching only" as well as "research only. And, increasingly, research and teaching were organized in different forms. McNay notes that in the 1996 RAE, the education panel was the only one that would accept teaching material as evidence of research output, and that "the teaching curriculum is being affected as senior staff in universities withdraw support from [departments] with low RAE grades, so that taught courses close" (200). Increasingly, staff rewards are research driven and some teaching funds are being reallocated ("raided") to finance research. Yet, as McNay points out, 80% of HE funding is for teaching. He questions the privileging of the "scholarship of discovery" over the "scholarship of transmission."

Another empirically based study investigated the RAE's impact on academic work in two social science and two business disciplines (Harley and Lowe 1999). In the study, some 80% of respondents identified changes and recruitment patterns in their discipline generally. Of these, three- quarters attributed the changes directly to the RAE and a further 18% held the RAE partly responsible. A quarter of the sample

characterized the changes in terms of *less emphasis on teaching skills;* just under two-thirds in terms of *greater emphasis on research;* and just over two-thirds in terms of *greater emphasis on publication.* More than three-quarters of the sample cited changes in recruitment and selection policies in their own departments as a result of the RAE. Asked about the changes taking place in their disciplines, 52% characterized them as "bad," 18% as "good and bad," and 23% as "good." In terms of impacts on their own work, 53% said the RAE had influenced it and only 10% indicated no influence whatsoever.

3. Australia

In Australia, the country's 40 public research universities and two private institutions are subject to a common framework of funding and regulation, that provides some 60% of their total funding and subjects them to the performance requirements of the Higher Education Funding Act (Marginson, 1998). Reform commenced in 1988, with the abolition of the binary divide between universities and colleges of advanced education, and has continued since that time. Reform included a number of early initiatives: a system of discipline reviews conducted by panels of experts reporting to the minister; the development and testing of a system of performance indicators; allocation of special funds to support performance initiatives; and establishment of a fund to improve teaching (Harman, 1998). There was strong emphasis on managerial modes of operation, adequate levels of accountability, and maximum flexibility in decision-making (Meek and Wood, 1998). Resulting changes have proved so extensive, the process is often referred to as the "Australian Experiment."

During 1993-95, a number of innovative performance features were introduced under the rubric of an annual academic audit focused on processes and outcomes (Harman, 1998). Participating universities would conduct a self-evaluation and prepare a detailed portfolio. Peer-review panels would visit and assess the institution's effectiveness in performance outcomes and processes. Universities would be ranked on the basis of effectiveness and outcome excellence and the rankings, together with detailed reports, would be published annually. As in England's RAE, these rankings and their publication were by far the most controversial element of the scheme. Results were widely reported in the media. High-ranked universities found their prestige had increased, while those who performed poorly experienced reputational damage. Finally, the process would be driven by the incentive of incremental performance funding, allocated according to the rankings, to a maximum of 5% of annual budgets for the top-ranked institutions (Harman, 1998).

Institutions have welcomed the additional funding and the program has garnered the support of institutional leadership and others who saw a need for management reforms and a greater client focus. Criticism has been severe however, much focused, as in England, around the contentious ranking system which favours the older, more-established universities; the underlying methodology and the reliance on narrow statistical data; the additional workload; and the negative effects on less-favoured institutions. Some have argued that, especially in teaching and learning, results are temporary. Others share Dill's (1998) opinion, that the cost/benefit ratio of the whole exercise is flawed, especially for the lower-ranked institutions where the consumption of scarce resources on these initiatives has bred staff resentment.

Nevertheless, the new government elected in August 1996 committed itself to continuing performance models, albeit with a 5% reduction in operating grants and other funding restraints (Meek and Wood, 1998). The Higher Education Council was made responsible for the government's new program, which includes the integration of various models; institutional reviews of performance improvements every three to four years; and public reporting of performance improvements. As of 1997, universities had been asked to submit a copy of their strategic plan, together with information on the key indicators they used to judge their own performance; current outcomes and intended improvements; and improvements since the last evaluation (Harman 1998:345).

A survey by Taylor and colleagues (Taylor et al., 1998) of Australian academics in three universities sought perceptions of the impacts of these and earlier reforms. The survey revealed a high level of concern in many areas and a fairly

dismal assessment of future prospects for teaching and research, as well as of the standard of undergraduate students and the extent of academic freedom. The quality of new students, teaching, and research are all identified as in decline, while the undervaluing of teaching in comparison with research persists. Changes in university management to a more corporate style are seen as a threat to academic freedom. More established research universities are concerned that scarce research funds are being stretched too widely. This perception is leading to new divisions in the unified higher education sector. The writers believe that "the tension between staff desire for academic freedom—with its often time-consuming collegial decision-making—and management's need for flexibility is set to continue" (269). Academics' entrenched distrust of administration "will not be ameliorated by the growing managerial desire to conceive of higher education as a corporate service industry". They conclude that "there is a real danger that management and academic staff will polarize" (ibid.)

Another study (Marginson, 1998) coined the term "new university" to capture the institutional impact of the constellation of changes introduced under the reform agenda. This extensive study of 17 universities found: the emergence of a new kind of strategic leader in the presidential office; eclipse of collegial decision-making and emergence of management-controlled, "post-collegial" mechanisms; changes in research management with consequent effects on academic work; commonalities and variations among the "new universities"; and that the changes corresponded with systems of "new public management." These results are confirmed in the study of governance and management by Meek & Wood (1998).

Currie and colleagues (Currie, 1998; Currie and Vidovich 1998) conducted a qualitative study based on interviews of 153 Australian and 100 American academics at six universities: Sydney, Murdoch, and Edith Cowan in Australia; Arizona, Florida State, and Louisville in the US. Additional data were drawn from studies and interviews in Canada and New Zealand. Currie's theoretical framework was constructed around Foucault's concept of governmentality; Lyotard's ideas on performativity; and theories of globalization and pervasive neoliberal market ideals. The focus was managerialism in Australian and US universities. A large majority (+85%) of respondents in the study reported increases in accountability and surveillance over the last five years. There was a sense that performance data were being gathered without any clear perception of how they were to be used.

Other perceptions included: declining budgetary control by faculty; predominance of private-sector approaches to management; the sense that universities no longer thought of themselves as primarily educational institutions; and a suspicion that salary and administrative costs for senior and middle management were burgeoning. Divisions between faculty and central administration were reported to be widening, with the academic function becoming subordinated to the administrative function. Full-cost recovery was a major theme (Fisher and Rubenson, 1998), as were efforts to run the university like a business. Those areas closer to the market flourished while the rest had to battle for survival. A majority of faculty (73% in US; 59% in AUS) said decision-making had become "more bureaucratic, top- down, centralized, autocratic, and managerial" (Currie, 1998:26). Of the rest, 19% in the US and 17% in AUS identified democratic decision-making as present at the unit level, while bureaucratic and corporate managerial procedures predominated at the institutional level.

4. New Zealand

New Zealand's 32 post-secondary institutions currently enroll some 200,000 students, just over half at the seven national universities. In September 1997, the New Zealand government released a green paper on tertiary (higher) education. The proposals were radical enough to prompt student protests in the streets of Auckland, Christchurch, and Wellington. Some 74 students were arrested attempting to break through a police barricade at the Parliament Buildings in Wellington. A student leader said that the proposals, if enacted, would turn the NZ into the "most right-wing country in the world" in terms of HE funding (Cohen, 1997:A44). An earlier, leaked version of the document used the term "corporatization," and painted a picture of "voucher-bearing students attend[ing] higher education institutions that were more private than public. The institutions would be expected to turn a profit" (ibid.). The language of the official version was more temperate.

Its release was followed by a year of extensive consultation and policy development—almost 400 submissions were received—culminating in a November 1998 white paper. In substance, the new policies have been compared to the UK's Dearing Report. Both the UK and NZ documents "suggest a future in which institutions will bear much more responsibility for their own affairs, particularly their financial affairs" (Cohen, 1997:A44). The white paper establishes the ground rules for what the government calls "a high-performing tertiary sector" (Creech, 1998). The policy direction follows the "evaluative state" model long established in New Zealand. It calls on universities to "lock-in quality" and sets up a number of mechanisms to ensure performance will occur.

A new intermediary body—Quality Assurance Authority New Zealand (QAANZ)—will "rigorously test" the teaching and research of every institution in the sector. Funding will depend on performance tests being met. As well, university governance will be reformed. Governing councils will be limited to twelve members, including faculty, outside experts, and students. The government reserves the right to intervene in the affairs of any institution deemed at risk, whether academically or financially, "to protect the taxpayers' investment". All institutions will have to demonstrate their financial viability before receiving further government funding.

The awarding of government funds for research will also be modified, along the lines of Britain's RAE, to introduce competition. Of the $100 million annual research budget, 20% will be set aside initially as a "contestable pool". To qualify, researchers will need a demonstrated track record in their fields and a "strategic" focus that both benefits the national interest and is cost-effective. In 2001, after a review of the country's research requirements, the plan is to increase the contestable portion of the annual budget to 80%.

These recent moves continue the process of cultural change in the New Zealand Higher Education System, that began with the "neoliberal experiment" in 1984. In a program of radical social and economic restructuring, successive governments have reconfigured the country once called "the welfare capital of the world" (Roberts, 1998:3). As in Australia, and Britain under Thatcher and Major, welfare benefits were slashed, user-pay systems were introduced in the public sector, and state assets privatized. The public sphere was transformed by the introduction of quasi-markets (Marginson, 1997). The trend towards devolution with strong state steering is that of the "evalulative state." Bureaucrats now talk the language of "inputs," "outputs," and "throughputs" (Roberts, 1998). Students pay a higher proportion of their educational costs and are designated as "customers." The teacher-student relationship has become contractual rather than pedagogic (Codd, 1997). The emphasis on performance and accountability for results is pervasive. The discourse is of "international competitiveness" and "enterprise culture" (Roberts, 1998:3). Transforming educational institutions into corporate entities "geared toward the ideal of making a profit or at least minimizing losses and efficiencies" has been an important objective (Roberts, 1998:3). Regular performance reviews—based on a variety of performance indicators—are mandated for all levels of the institution, to ensure efficiency objectives are met. The development of a National Qualifications Framework, which breaks down the "educational product" into "unit standards," facilitates the Taylorization (Dominelli and Hoogvelt, 1996) and commodification (Peters and Marshall, 1996) of higher education in New Zealand.

5. Sweden

The evaluation movement arrived in Sweden later than elsewhere in Europe, with performance models first appearing on the political agenda towards the end of the 1980s (Nilsson and Naslund, 1997). It is also developing somewhat differently than in other Nordic countries with a clear trend linking program reviews, institutional evaluations, and national evaluations. Considerable movement can be detected away from the system of highly centralized state control of HE, that saw the country through the expansive period of the 1960s and 1970s. Decentralization was the *motif* of the 1980s. In 1989, the Minister of Education appointed a national commission to begin investigating the quality of higher education. The Liberal-Conservative government of 1991-94 signalled continuing commitment to deregulation of HE policy, with their 1992 proposition: *Universities and Colleges of Higher Education—Freedom for Quality.* They disbanded the central HE authority

(Universitets-och höhskoleämbetet—UHÄ) and allowed individual insitutions to communicate directly with the Ministry of Education regarding funding.

Infused with neoliberal ideology, the new government sought to provide institutions with more autonomy in their dealings with the state. They established a national Secretariat for Evaluation of Universities and Colleges (subsequently to become the Office of the Chancellor) with a mandate to determine "various indicators of quality which can be used as the basis for allocating funds for undergraduate education" (SFS, 1992, cited in Nilsson & Naslund, 1997: 7). When this proved unrealistic at a national level, each institution was given responsibility for establishing a program of quality development. With the institution of the 1994 proposition (*Teaching and Research—Quality and Competitiveness)* 5% of each institution's resource allocation was based on an evaluation of its quality development program and iimplementation efforts (Nilsson and Naslund, 1997). When the Social Democratic government assumed power in 1994 they did away with this premium, declaring that "quality enhancement is not simply something that is expressed in special programmes but is basically an attitude which must characterize the day-to-day work of each institution (Nilsson and Naslund, 1997:7).

The Social Democratic Government also restructured the intermediate authority into separate free-standing units— including the National Agency for Higher Education (Högskoleverket)—to ensure that institutional performance programs were reviewed regularly. Thus, beginning in 1995, efforts to improve the quality of performance, rather than the quality of education, became the focus of assessment. Concurrent with this decision came the announcement that total funding of undergraduate education was being cut by 10%. Bauer and Kogan (1997) argue that while there appears to be a general trend in devolution of authority from the state to institutions, and while the notion of a national system of performance indicators has been abandoned, the State has actually increased its performance requirements. Feedback of results is an important function in the new steering system. Greater autonomy has thus been obtained at the costs of increased demands for accountability, and a more systematic approach to assurance. This is described by Wahlén (1998), as a shift from a system of management by rule, to one of management by goals or results. The system includes the evaluation of individual educational subjects at a National level, the evaluation of education programs for accreditation, and an emphasis on the development of a professional culture in which university staff take responsibility for their work and its results. Recently, as well, a new requirement calls on universities to report student outcomes according to class, ethnicity, and gender. In performance models generally, social engineering ambitions are never far away.

Finally, all 36 institutions of higher education in Sweden must undergo a quality audit to ensure that mechanisms are in place, before the year 2000, for the efficient use of resources. From early indications, university reactions to these moves are mostly positive (Wahlen, 1998:38).

In a study of performance systems in the Nordic countries, Smeby & Stensaker (1999) found evidence in all four countries of balance between internal institutional needs and external societal needs. None of the countries link assessment with resource allocation nor are there direct attempts at political steering. Rather, the intent seems to be ameliorative and, as such, may bolster academics' trust in these systems (1999:13). Despite surface similarities, however, differences in design and practice are apparent, reflecting the differing institutional and political endowments of each country. While the authors accept that performance models represent the new "meta-discourse" of HE policy, they suggest that "the processes involved imply, at least in the Nordic countries, very incremental changes to existing structures of power within higher education" (1999:13). In Norway and Finland, for example, these systems are considered "policy experiments." In Denmark, the process is undergoing reassessment at the end of the first round, while in Sweden the history of decentralization and delegation predates the new meta- discourse, extending back to 1977. The authors conclude that "changes to the existing external and internal "power balance" between state and institutions...occur very slowly in all four countries" (ibid.). This study therefore supports a "historical institutionalist" interpretation of path-dependent policy change (Hall, 1997).

6. The Netherlands

Together with France and Great Britain, the Netherlands was among the first European countries to institute a formal performance model system in the mid-1980s. The original approach combined self-evaluation with peer review by visiting expert committees. The focus was the program, rather than the institution. The state strongly advocated performance indicators, but these were resisted by universities. The model was refined in the Ministry of Science and Education's 1985 publication *Higher Education Autonomy and Quality,* which set out a new coordination relationship between the HE sector and the state (Maassen 1998). More autonomy would be granted, but in exchange for cooperation in the development of a comprehensive system designed to regularly assess the performance of university performance. The state would not completely devolve its authority, but would be selective about the arenas of its involvement. As well, the coordination relationship was open to other stakeholders such as employers and local authorities. According to Maassen, the system incorporated a drift towards market-oriented criteria (1998:20). Universities were to develop strategic, performance-based self-knowledge—*institutional profiles*—and were encouraged to adopt managerial modes of behavior and business principles.

Originally, the state intended the Inspectorate of Higher Education to administer the performance model. But through a compromise deal in 1986, the universities and higher professional schools (the Netherlands has a dual system) were able to involve their own representative organizations in the process, and the IHO was bypassed. In practice, two separate systems were developed: one for universities coordinated by the Association of Cooperating Universities in the Netherlands (VSNU); the other for the higher professional sector coordinated by the HBO-Council (Maassen, 1998:21-2). Both emphasized the dual performance goals of quality improvement and accountability. The VSNU's pilot project began in 1988 and the full system became operational in 1989.

While adapted from the North American model, the Dutch system differs because it is collectively owned by the institutions. Largely because of this, over time, the emphasis has shifted from the accountability end of the spectrum towards the improvement end. As well, evaluation results do not feed into the policy or funding process; there are no political consequences. It is felt direct links would lead to strategic behaviour and tend to undermine the improvement process (Maassen, 1998:25). This creates something of a dilemma since real incentives are lacking, yet if incentives were introduced, power games would prevail. According to Maassen, the Ministry's response has been to abstain from short-term interventions, but with the threat of medium- to long-term consequences in the absence of results. Thus the IHO plays a meta-evaluative, monitoring role. So far, the trust invested in institutions appears not to have been misplaced. Faculties and departments seem to take their responsibilities under the system seriously.

But, in the absence of incentives, what does "taking responsibilities seriously" mean? Has the low-key approach to performance produced any real change? A study of Dutch higher education by Frederiks & Westerheijden (1994) concluded that the quality of teaching is receiving considerably more attention than before the reforms. Many programs and faculties now have "special committees or specially appointed staff members for the quality management of education" and the topic "has certainly gained an important place on the agenda of [university} decision makers" (1994:200). As well, in contrast to the former singular focus on pedagogy, the input and output characteristics of education—informing potential students, and investigating the labour market prospects for graduates—are now receiving attention. Frederiks & Westerheijden suggest that a "quality culture" is emerging in Dutch higher education.

In terms of responses to self-evaluations and the recommendations of visiting peer-review committees, the authors find that while measures *are* taken to address outstanding issues, the relation between taking measures and observing improvement is obscure. There is no evidence that "the large amount of resources invested leads immediately to an equally large improvement in the quality of education" (ibid.). Nevertheless, the authors find a surprisingly high level of satisfaction with the Dutch performance model. Surprising for two reasons: the traditional reluctance of autonomous organizations to submit to external scrutiny,

and the heavy administrative burden involved in constructing an adequate self-evaluation.

Despite generally high levels of satisfaction, however, Maassen forecasts change. Specifically, this relates to Holland's role in the EU, and the general harmonization of HE under EU rules. Some type of accreditation approach may well replace the peer review system in the coming decade.

V. Summary and Conclusions

The politics of performance is deeply embedded in the "evaluative state" and the trend to performance measurement is unlikely to be reversed. Indeed, with the normalization of performance expectations and the broadening of knowledge missions beyond teaching and research, accountability and performance criteria are likely to become ever more complex and embedded. Gibbons predicts "new bench-marking methodologies and the production of a range of bench-marking studies right across the higher education sector" and the use of quality indicators to rank universities "by region, by country and even globally" (1998: 50).

With the globalization of performance in prospect, our study shows deep flaws in the conceptualization, measurement criteria, and impacts of these models (see Appendix for more details.) At the technical level, for example, we report lack of clarity in definitions of what constitutes "good performance," and absence of agreement on the adequacy of specific indicators. At the broad system level, we identify increasing differentiation and stratification as universities were defined by their performance rankings as "good," "bad," or "indifferent" performers, and as either "research" or "teaching" institutions. Increasingly, teaching and research are being defined as measurable products rather than processes of learning or enquiry. The proliferation of buffer bodies to mediate compliance with performance models was a feature of all systems studied.

In terms of institutional effects, we find a performance- linked focus on missions and visions that promote increased efficiency and calls for more effective, centralized management. Funding is increasingly linked to performance on various measures, variously defined, few of which account for traditional moral or social imperatives. A consistent complaint is the amount of time and expense involved in conforming to proliferating compliance requirements. Individual departments and faculty members report erosion of disciplinary boundaries and decline of collegiality, as well as polarization between departments and the locus of administrative control. Throughout, we find a strong consensus that the costs of compliance with performance regimes far outweigh the benefits.

Our review of the experience of different states and institutions raises a number of empirical questions deserving of further study. Is there any evidence that performance-based funding will actually improve institutional performance in the long run? Is the money allocated in these programs a large enough incentive for participation, or is the implied threat of greater state intervention and the loss of autonomy sufficient motivation? Does compliance indicate agreement with the concept and process? Are the ways states deal with non-compliance effective? Do attempts to meet general, institution-level performance measures create goal dissonance and other difficulties at different internal levels? To what extent is the increased demand for detailed reporting an additional burden? Will institutions engage in aggressive competition in attempts to demonstrate compliance? If funding is at stake, is there a possibility that quality of education will be sacrificed in the rush to meet external standards and access additional funds?

Only longitudinal empirical research can answer questions like these, and determine whether performance models have enduring value for the conduct of higher education. Further study is clearly needed. Given the evidence to date, there seems to be no "ideal" model or mix. However, if one country stands out, it is the Netherlands. Of those national systems reviewed here, the Dutch seem to have mastered the positive aspects of performance models while avoiding many of the more negative consequences. This is the reason, no doubt, that many countries in Continental Europe follow a "softer" Dutch-style model, involving qualitative

measures and far less prominence for performance indicators than in the UK and US. States, territories, and provinces that have yet to implement these models, might want to consider the contrasting understandings of "performance" in the European and Anglo-Saxon systems, and review relative strengths and weaknesses, before committing resources.

In conclusion, few would argue against the ethic of accountability that animates performance models, nor would they disagree that what performance models measure is important. But the "fatal flaw" of performance models is that they reduce performance to what is measurable, when so much of importance is not. Because performance models focus on instrumental and utilitarian concerns, the fear is that the intrinsic value of education may be lost.

As it becomes more accountable in a "knowledge society," can the university survive in its traditional form? Survival may depend on a much broader definition of accountability, according to Delanty (1999); one that encompasses public and civic commitment. The best way to guarantee the future of the university, he says, is to reposition it at the heart of the public sphere, "establish[ing] strong links with the public culture, providing the public with enlightenment about the mechanisms of power and seeking alternative forms of social organization." Further, with university knowledge becoming such a central social, economic and political resource, why be "a tool of the state and market forces"? Why not, instead, become an agent of social and political change? (ibid.). The central task, we would argue, is to embrace a social mission, banish lingering élitism, and advance the democratization of knowledge.

Appendix: Summary of issues and impacts of performance models internationally

In the tables below, we itemize the consequences, impacts, and issues attached to the performance models we reviewed in a set of tables. As this article makes clear, some of these effects are more pronounced in Anglo-Saxon systems, others in European systems. We do not differentiate among the systems nor do we make a determination whether the consequences are good, bad, or indifferent, since these are open to interpretation and will be conditioned by the reader. We have organized the effects into five categories: (i) overall system-level effects; (ii) technical performance issues; (iii) institutional effects and management issues; (iv) impacts on teaching and research; and (v) impacts on faculty and academic departments. Clearly, many of the effects "spill-over" into other categories and may even appear mutually contradictory. It is worth reiterating that, whatever the commonalities, legacies count. Whether cultural, institutional, national, or ideological, the differences between systems are as great as the convergence among them. Finally, the classification scheme is both provisional and heuristic and should not be read otherwise. No attempt is made to rank-order the effects or to exhaustively reproduce every element previously discussed. We try, instead, to convey generalities.

System-level effects

- possible differentiation of universities into research institutions and teaching institutions
- increased stratification, as rankings differentiate "good," "bad," and "indifferent" performers
- more isomorphism as valid differences are erased by conformance to a limited number of indicators
- "newcomers" have to compete with established institutions for limited funds
- established institutions have to share "steady state" funding with newcomers
- proliferation of external intermediary bodies to administer performance and quality programs and mandate consequences of noncompliance and "poor performance"
- more "rational" basis for funding decisions therefore better justifications for HE funding
- bilateral systems unified
- social engineering ambitions

- broad frameworks replace regulation (dejuridifation)
- proliferation of stakeholders to be accommodated

Technical performance issues

- lack of agreed-on definitions of what constitutes "good performance" (quality)
- lack of agreement concerning the adequacy of specific performance indicators
- incompatibilities between performance measures, so that maximizing some means underperforming on others
- inability of quantitative measures to capture contextual and institutional differences
- use of dubious proxies of performance
- reduction of complexity
- subjective bias in construction and interpretation of measures
- appearance of "objective" neutrality
- more, and more directly useful data; revelations about previously unknown aspects of performance
- increased ability to "prove" accountability for public funds
- susceptibility of measures to changing political agendas

Institutional effects and management issues

- increased efficiency and more effective management
- focus on "missions," priorities, and identification of strengths
- growth of non-academic management-support functions with the power to intervene in academic decisions
- funding increasingly linked to performance, on various measures, variously defined
- increased competition, both within and between institutions
- increased surveillance, both internal and external
- centralized, corporate decision making, supported by budgetary and performance-based criteria
- increased time and costs to administer and conform to proliferating compliance requirements
- possibility that short term gains from compliance will produce "long-term pain"
- possibility that the "short term pain" of compliance will produce long term gains
- evidence that universities are becoming more market-like; strategic behaviour to maximize market gains
- evidence that universities are abandoning traditional societal and moral imperatives
- better understandings of institutional missions and new, more dynamic perspectives on the management of institutions
- better responsiveness to the needs of public, political, and other stakeholders
- limited financial incentives

Impacts on teaching and research

- performance defined as measurable product (publications; external research funding; job-ready graduates) rather than process (learning; inquiry)
- separation of research and teaching
- more-rigorous definitions of "active research"
- focus on quantity rather than quality of research
- focus on quantity rather than quality of publications
- devaluation of teaching in some systems, with shift of resources to research
- less time for *performing* teaching and research due to *conforming* with compliance procedures
- peer-reviewer "burn-out" as more are called on to participate in assessments and audits
- preference for research with measurable outcomes, within a defined time

frame, that carries external funding
- shift in pedagogical emphasis as students demand more "relevance"
- value-for-money approach: students are no longer learners in pursuit of understanding, but customers taking delivery of a commodity
- impact of cost/benefit and cost-recovery constraints on course diversity
- narrow definitions of research performance discourage risk-taking and innovation

Impacts on faculty and academic departments

- erosion of disciplinary boundaries
- decline of collegiality
- individual projects discouraged in favour of "team efforts"
- polarization between faculties/departments and central administration
- detrimental effect of compliance exercises on faculty workloads
- decreased faculty time for students and community service
- increased stress, anxiety, uncertainty, and resentment
- resistance to the measures although this tends to be passive rather than active
- "Taylorization" of faculty work means more short-term contracts and less security
- loss of autonomy over individual work
- demands for more productivity

Note

The authors wish to acknowledge the financial support of the Humanities and Social Science Federation of Canada, which funded the foundational study for this paper. The authors also thank Professors Kjell Rubenson and Donald Fisher, co-directors of the Policy Centre, for their guidance.

References

Albright, B. N. (1997) Of Carrots, Sticks, and State Budgets, available online at <http://www.nacubo.org/website/accountable/agb.html>. In *Accountability and Regulation on Public Higher Education*. August 16

Albright, B. N. (1998) Performance-based Funding. *Network News*, **17**(1), February

Aucoin, P. (1995) *The New Public Management: Canada in Comparative Perspective*. Montreal: The Institute for Research on Public Policy

Baker, J. (1997) Conflicting Conceptions of Quality - Policy Implications for Tertiary Education. Paper presented at. AIC Tertiary Education in New Zealand Conference, 27 & 28 May 1997

Banta, T., Rudoph, L. B., Van Dyke, J. & Fisher, H. S. (1996) Performance Funding Comes of Age in Tennessee. *Journal of Higher Education*, **67**(1)

Barry, A., Osborne, T. & Rose, N., Eds. (1996) *Foucault and Political Reason: Liberalism, Neoliberalism, and Rationalities of Government*. Chicago: University of Chicago Press

Bauer, M. & Kogan, M. (1997) Evaluation Systems in the UK and Sweden: Successes and Difficulties. *European Journal of Education*, **32**(2)

Bjarnason, S. (1998) *"Buffer" Organizations in Higher Education: Illustrative Examples in the Commonwealth*, Study supported by the Commonwealth Fund for Technical Cooperation. London: Association of Commonwealth Universities, available online at <http://wwwacuacuk/chems/onlinepublications>

Brennan, J. (1999) Evaluation of Higher Education in Europe. In *Changing*

Relationships Between Higher Education and the State, ed. M. Henkel & B. Little, pp. 219-35. London: Jessica Kingsley Publishers

Broadhead, L. and Howard, S. (1998). "The Art of Punishing": The Research Assessment Exercise and the Ritualisation of Power in Higher Education. *Education Policy Analysis Archives, 6*(8). (Entire issue.) Available online at http://epaa.asu.edu/epaa/v6n8.html.

Burchell, G., Gordon, C. & Miller, P., eds. (1991) *The Foucault Effect: Studies in Governmentality*. Chicago: University of Chicago Press

Burke, J. c. & Serban, A. M. (1998) *Funding Public Higher Education for Results: Fad or Trend?* Results from the second annual survey, July 1998. State University of New York: The Nelson A Rockefeller Institute of Government

M., Hanney, S. & Kogan, M. (1991) *The Use of Performance Indicators in Higher Education: A Critical Analysis of Developing Practice, Second Edition*. London: Jessica Kingsley Publishers

Charih, M. & Daniels, A. (1997) Introduction: Canadian Public Administration at the Crossroads. In *New Public Management and Public Administration in Canada*, ed. M. Charih & A. Daniels, pp. 13-24. Toronto: The Institute of Public Administration of Canada

Charih, M. & Rouillard, L. (1997) The New Public Management. In *New Public Management and Public Administration in Canada*, ed. M. Charih & A. Daniels, pp. 27-45. Toronto: The Institute of Public Administration of Canada

J. (1997) Knowledge, Qualifications, and Higher Education: A critical view. In *Education Policy in New Zealand: The 1990s and Beyond*, ed. K. M. Matthews & M. Olssen. Palmerston North: Dunmore Press

Cohen, D. (1997) New Zealand prepares for a major shake- up of its higher education system. *Chronicle of Higher Education*, December 19, A44, International section

Creech, W. (1998) *Students, Quality, and Fairness— Key to Future for Tertiary Education*, Minister of Education's Press Statement on the Tertiary Education White Paper, November 18. Wellington: Government of New Zealand

Currie, J. (1998) Globalization Practices and the Professoriate in Anglo-Pacific And North American Universities. *Comparative Education Review*, **42**(1), February, 15-29

Currie, J. & Vidovich, L. (1998) Microeconomic Reform through Managerialism in American and Australian Universities. In *Universities and Globalization: Critical Perspectives*, ed. J. Currie & J. Newson. Thousand Oaks: Sage

R. (1997) The State and the Governance of Education: An Analysis of the Restructuring of the State-Education Relationship. In *Education: Culture, Economy, and Society*, ed. A. Halsey, H. Lauder, P. Brown & A. S. Wells. Oxford: Oxford University Press

Dearing, Lord (1997) Report of the National Committee of Enquiry into Higher Education, Hayes, Middlesex: NCIHE

Delanty, G. (1999) The transformation of knowledge and the ethos of the university: Outline of a theory of epistemic shifts. Plenary paper. *Re-Organizing Knowledge/Transforming Institutions: The University in the XXI Century*. University of Massachusetts-Amherst, September 17-19

D. D. (1998) Evaluating the "Evaluative State": Implications for Research in Higher Education. *European Journal of Education*, **33**(3), September, 361- 78

Dominelli, L. & Hoogvelt, A. (1996) Globalization, Contract Government and the Taylorization of Intellectual Labour in Academia. *Studies in Political Economy*, **49**(Spring), 71-100

El-Khawas, E. (1998) Strong State Action but Limited Results: Perspectives on University Resistance. *European Journal of Education*, **33**(3), September, 317- 31

El-Khawas, E., DePietro-Jurand, R. & Holm-Nielsen, L. (1998) Quality Assurance in Higher Education: Recent Progress; Challenges Ahead. Preparation supported by the World Bank; available online at <www.worldbank.org/html/extdr/educ/postbasc.htm>. UNESCO World Conference on Higher Education. Paris, France, October 5-9, 1998

Fisher, D. & Rubenson, K. (1998) The Changing Political Economy: The Private and Public Lives of Canadian Universities. In *Universities and Globalization: Critical Perspectives*, ed. J. Currie & J. Newson. Thousand Oaks: Sage

Foucault, M. (1978) Governmentality. In *The Foucault Effect: Studies in Governmentality*, 1991, ed. G. Burchell, C. Gordon & P. Miller, pp. 87-104. Chicago: University of Chicago Press

Frederiks, M. & Westerheijden, D. (1994) Effects of Quality Assessment in Dutch Higher Education. *European Journal of Education*, **29**(2), 181-200

Gibbons, M. (1998) Higher Education Relevance in the 21st Century. Paper prepared for UNESCO World Conference on Higher Education, Paris, France, October 5-9, 1998. Secretary General, Association of Commonwealth Universities, supported by the World Bank. 70 pages

Hall, P. (1997) The Role of Interests, Institutions, and Ideas in Comparative Political Economy of the Industrialized Nations. In *Comparative Politics: Rationality, Culture, and Structure*, ed. M. I. Lichbach & A. S. Zuckerman. Cambridge: Cambridge University Press

Harley, S. & Lowe, P. (1999) Academics Divided: The Research Assessment Exercise and the Academic Labour Process. Paper presented at. Critical Perspectives on Accounting. Baruch, CUNY, April 25-27

Harman, G. (1998) Quality Assurance Mechanisms and their Use as Policy Instruments: Major International Approaches and the Australian Experience since 1993. *European Journal of Education*, **33**(3), September, 331- 49

Harris, J. (1998) Performance Models. *Public Productivity & Management Review*, **22**(2), December, 135-40

Henkel, M. (1997) Academic Values and the University as a Corporate Enterprise. *Higher Education Quarterly*, **51**(2), April, 134-43

Henkel, M. & Little, B. (1999) Introduction. In *Changing Relationships Between Higher Education and the State*, ed. M. Henkel & B. Little, pp. 9-22. London: Jessica Kingsley Publishers

Hood, C. (1991) A Public Management for All Seasons? *Public Administration*, **69**(Spring), 3-19

Hood, C. (1995) The "New Public Management" in the 1980s: Variations on a Theme. *Accounting, Organizations and Society*, **20**(2/3), 93-109

Hoskin, K. W. & Macve, R. H. (1993) Accounting As Discipline: The Overlooked Supplement. In *Knowledges: Historical and Critical Studies in Disciplinarity*, ed. E. Messer-Davidov, D. R. Shumway & D. J. Sylvan, pp. 25-53. Charlottesville, VA: University Press of Virginia

Johnstone, D. B. (1998) The Financing and Management of Higher Education: A Status Report on Worldwide Reforms. Preparation supported by the World Bank; available online at <www.worldbank.org/html/extdr/educ/postbasc.htm>. UNESCO World Conference on Higher Education. Paris, France, October 5-9, 1998

Kaghan, W.N. (1998), "Court and Spark: Studies in Professional University Technology Transfer Management," Diss, Seattle, WA, University of Washington.

Keating, M. (1998) *Public Management Reform and Economic and Social Development*. Paris: OECD

Maassen, P. A. M. (1997) Quality in European Higher Education: Recent Trends and their Historical Roots. *European Journal of Education*, **32**(2), 111- 25

Maassen, P. A. M. (1998) Quality Assurance in the Netherlands. In *Quality Assurance in Higher Education: An International Perspective*, ed. G. H. Gaither, pp. 19-27. San Francisco: Jossey Bass

Marginson, S. (1997) *Markets in Education*. Sydney: Allen and Unwin

Marginson, S. (1998) The Best of Times and the Worst of Times; Research Managed as a Performance Economy—the Australian Case. Paper presented at. Association for the Study of Higher Education. Miami, FL, November 4-5

McDaniel, O. C. (1996) The Theoretical and Practical Use of Performance Indicators. *Higher Education Management*, **8**(3), November

McKeown-Moak, M. P. (1999) *Financing Higher Education: An Annual Report from the States*, online publication at <http://www.sheeo.org/pubs/pubs-finance99.htm>. SHEEO

McNay, I. (1999) The Paradoxes of Research Assessment and Funding. In *Changing Relationships Between Higher Education and the State*, ed. M. Henkel & B. Little, pp. 191-218. London: Jessica Kingsley Publishers

Meek, V. L. & Wood, F. (1998) Higher Education Governance and Management: Australia. *Higher Education Policy*, **11**(2-3), June, 165-81

Meyer, J.W. and Rowan, B. (1977), "Institutional Organizations: Formal Structure as Myth and Ceremony," in W.W. Powell and P. DiMaggio (Eds.), *The New Institutionalism in Organizational Analysis*, 1991, Chicago, University of Chicago Press.

Miller, H. D. (1995) *The Management of Change in Universities: Universities, State and Econonomy in Australia, Canada, and the United Kingdom*. Buckingham: Open University Press and Society for Research into Higher Education

Miller, P. (1994) Accounting as a Social and Institutional Practice: An Introduction. In *Accounting as a Social and Institutional Practice*, ed. A. G. Hopwood & P. Miller, pp. 1-39. Cambridge: Cambridge University Press

NASBO. (1996) *State Innovations in Higher Education Finance and Governance*. National Association of State Budget Officers Information Brief, vol. 4, No.1, April. Washington, DC: NASBO

Neave, G. (1988) On the Cultivation of Quality, Efficiency, And Enterprise: An

Overview of Recent Trends in Higher Education in Western Europe 1986-1988. *European Journal of Education*, **23**, 7-23

Neave, G. (1992) On Bodies Vile and Bodies Beautiful: The Role of "Buffer" Institutions between Universities and State. *Higher Education Policy*, **5**(3)

Neave, G. (1998) The Evaluative State Reconsidered. *European Journal of Education*, **33**(3), September, 265-85

Network News. (1998) State Survey on Performance Measures: 1996-7. *Network News: A Quarterly Bulletin of the SHEEO/NCES Communication Network*, Vol.17, No.1 (February)

Newson, J. (1992) The Decline of Faculty Influence: Confronting the Effects of the Corporate Agenda. In *Fragile Truths: Twenty-Five Years of Sociology and Anthropology in Canada*, ed. W. K. Carroll, L. Christiansen-Ruffman, R. Currie, F & D. Harrison, pp. 227- 46. Ottawa, ON: Carleton University Press

Newson, J. (1998) The Corporate-Linked University: From Social Project to Market Force. *Canadian Journal of Communication*, **23**(1), Winter, 107-24

Nilsson, K. A. & Naslund, H. (1997) Towards a Swedish Evaluation and Quality Assurance System in Higher Education. Http://www.evaluat.lu.se/dokument/evaluation.htm. August 15 1999

OECD (1995) *Governance in Transition: Summary*, PUMA Committee, 1990-95. Paris: OECD

OECD (1999) IMHE: Institutional Experiences of Quality Assessment in Higher Education. Available online at <http://www.oecd.org//els/edu/imhe/instexp.htm>, for information contact: John Brennan at <J.L.Brennan@open.ac.uk>. August 14 1999

Peters, M. & Marshall, J. (1996) The Politics of Curriculum: Busnocratic Rationality and Enterprise Culture. *Delta*, **48**(1), 33-46

PISG (1999) *Performance Indicators in Higher Education*, First report, Performance Indicators Steering Group, First report—February. London: Higher Education Funding Council for England

Polster, C. & Newson, J. (1998) Don't Count your Blessings: The Social Accomplishments of Performance Indicators. In *Universities and Globalization: Critical Perspectives*, ed. J. Currie & J. Newson, pp. 173-91. Thousand Oaks: Sage

Porter, T. M. (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press

Powell, W. W. and DiMaggio, P. (1983) "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields," in W.W. Powell and P. DiMaggio (eds) *The New Institutionalism in Organizational Analysis*, 1991, Chicago, University of Chicago Press

Power, M. (1995) *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press

Power, M. (1996) Making Things Auditable. *Accounting, Organizations and Society*, **21**(2/3), 289-315

Roberts, P. (1998) Rereading Lyotard: Knowledge, Commodification and Higher Education. Http://www.sociology.org. *Electronic Journal of Sociology*, **3**(3)

Ruppert, Sandra (1998) Survey Results of State Political and Higher Education Leaders. Available online at <http://www1.umn.edu/mhec/summit/issues.htm>

Savoie, D. J. (1995) What is Wrong with the New Public Management? *Canadian Public Administration*, **38**(1), Spring, 112-21

Scott, Peter (1995) *The meanings of mass higher education*. Buckingham: Open University Press

Scott, W.R. (1995), *Institutions and Organizations*, Thousand Oaks, Ca, Sage.

SHEEO. (1997) *Focus on Performance Measures*, available online at <http://www.sheeo.org/network/nn- feb98.htm>

Smeby, J.-C. & Stensaker, B. (1999) National Quality Assessment Systems in the Nordic Countries: Developing a Balance between External and Internal Needs. *Higher Education Policy*, **12**(1), March, 3-14

Stanley, E. C. & Patrick, W. J. (1998) Quality Assurance in American and British Higher Education: A Comparison. In *Quality Assurance in Higher Education: An International Perspective*, ed. G. H. Gaither. San Francisco: Jossey Bass

Strange, S. (1996) *The Retreat of the State: The Diffusion of Power in the World Economy*. Cambridge: Cambridge University Press

Taylor, T., Gough, J., Bundrock, V. & Winter, R. (1998) A Bleak Outlook: Academic staff Perceptions of Changes in Core Activities in Australian Higher Education. *Studies in Higher Education, Oct98, Vol 23 Issue 3, P255, 14p*, **23**(3), October, 255-69

Trow, M. (1998) American Perspectives on British Higher Education under Thatcher and Major. *Oxford Review of Education*, **24**(1), March, 111-30

Wahlen, S. (1998) Is there a Scandinavian Model of Higher Education? *Higher Education Management*, **10**(3)

Watts, R. (1992) Universities and Public Policy. In *Public Purse, Public Purpose: Autonomy and Accountability in the Groves of Academe*, ed. J. Cutt & R. Dobell, pp. 75-91. Ottawa, ON: Institute for Research on Public Policy and Canadian Comprehensive Auditing Foundation

Williams, G. (1998) Misleading, Unscientific, and Unjust: The United Kingdom's Research Assessment Exercise. *BMJ: British Medical Journal*, **316**(7137), 4 April, 1079-83

Woodhouse, D. (1996) Quality Assurance: International Trends, Preoccupations, and Features. *Assessment and Evaluation in Higher Education*, **21**(4), December, 347-57

About the Authors

Janet Atkinson-Grosjean

Centre for Policy studies in Higher Education & Training
2125 Main Mall
University of British Columbia
Vancouver, BC, Canada, V6T 1Z4

604-377-8155
Email: janetat@interchange.ubc.ca

Janet Atkinson-Grosjean is conducting interdisciplinary research at the intersection

of science and innovation policy, higher education policy, science studies, and institutional and organizational sociology. Her dissertation examines the status of "post-academic" science at the public/private divide, where intellectual property rights move publicly-funded discoveries into the private domain. Her fieldwork concerns Canada's Network of Centres of Excellence program. Other research interests include accountability, performance and governance in higher education, and the sociology of the professions. She holds a doctoral fellowship from the Social Sciences and Humanities Research Council of Canada and has presented her work at national and international venues.

Garnet Grosjean
Research Coordinator
Centre for Policy Studies in Higher Education and Training

Garnet Grosjean's research draws on the policy and practice of adult and higher education, to study co-operative education as a bridge between the academy and workplace. The central focus is the effect of learning context on university students' conceptions of learning and work. Other research interests include governance and performance of higher education, and the changing vocational role of the university. He is a member of Canada's Western Research Network on Education and Training.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 231 times since July 4, 2000

archives | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 8 Number 31 July 4, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Class Pictures: Representations of Race, Gender and Ability in a Century of School Photography

Eric Margolis
Arizona State University

Abstract

This article examines photographs taken of American public school classes between the 1880's and the 1940's. Most of the images were found in two virtual archives: The American Memory site at the Library of Congress and The National Archives and Record Center. These very large photograph collections were searched for representations of race, gender, and physical ability. The photographs were compared and contrasted and analyzed for elements of hidden curricula using techniques drawn from the social sciences and humanities. It was found that these large photo collections have significant gaps and historical amnesias. Collections made under conditions of racial segregation are themselves segregated and continue to reproduce images of hierarchy and dominance. To the extent these sites function as important resources for teachers and students searching for primary source documents for history and social studies projects, the archives convey significantly biased views of the history of education and minority groups in America.

It is a common experience of childhood in America. Teachers tell their class to wear dress clothes tomorrow because the photographer is coming to take the class picture. School photography was a regular source of income for local photography studios, a source of pride for schools, and a memento for students and their families. Most of these photographs did not withstand the tests of time — faded, lost, or thrown out with the rest of our childhood things. Others survived and found their way to local or state historical collections or historical archives. Often the only thing preserved in the process was the image itself, with little provenance or documentary material to understand the image (see, Figure 8 below, for an intriguing example). Occasionally entire studios with tens of thousands of negatives were donated to or purchased by state historical societies or museums. Advancing technology that includes digitized images, databases allowing fast search and retrieval, and the Internet for dissemination has spurred a secondary development as entire collections are being swept into ever more enormous virtual archives that are open to anyone with a personal computer and access on line. (Note 1)

An article in the New York Times (November 29, 1998) entitled "Digitized Artifacts are Making Knowledge Available to All, on Line" suggests the scale of a new resource:

> The Library of Congress, which has 117 million items in its archives, hopes to have four million items digitized and accessible on the World Wide Web by the turn of the century. The Denver Public Library expects to put 95,000 photographs of the old west on-line. California has linked 35 universities and museums into one on-line archive.

Clearly, in a very short time most of the major historical photograph collections will go on-line thus creating a searchable data base of millions of historic images. Future developmen' will include search engines designed specifically to retrieve photographic images, not indirectly by a key word system but by seeking images directly. (Note 2) Mega-sites like the Library of Congress's "American Memory" digital archive with 42 separate collections and hundreds of thousands of images and the National Archives and Records Administration with 54,000 images are enormously popular. These and similar electronic archives are free and open to the public twenty-four hours a day and seven days a week. Image banks have quickly become an invaluable source of primary source data for students doing research and gathering material for reports and class projects, and they are a remarkable resource for teachers and others preparing lectures, doing research, or just browsing. (Note 3) If, as the Library of Congress name suggests, they have literally become a representation of our collective memory, an essential question becomes: What is the nature of that memory?

A Simulacra of History? Historical Photographs on the Internet

These technological developments have opened an entirely new niche to historians and scholars of visual communication, making possible research which was unimaginable only a decade ago. (Note 4) While this is a remarkable technological advance and a general benefit for scholars and researchers, there are a number of caveats to this development, of which I will mention just two that are particularly salient to this discussion. (Note 5)

The first has to do with the uses to which such freely available images may be put. As the Internet develops into what will be in effect a single archive, the meanings of the individual collections (and photographs) will tend to become submerged. Alan Sekula, posed central questions for those interested in understanding and using historical photographs: "How is historical and social memory preserved, transformed, restricted, and obliterated by photographs" (Sekula, 1983:193)? Having raised those questions, Sekula (1983:195) warned that "Photography constructs an imaginary world and passes it off as reality." He drew attention to some of the sources of error and misrepresentation in collections of historic photographs. He mentioned the fallacies of assuming that photographs "transmit truths"; "reflect reality"; or are "historical documents." "The very term

document," explained Sekula, "entails a notion of legal or official truth, as well as a notion of *proximity to* and verification of an original event" (Sekula, 1983:198).

Sekula (1983:194) has also given a great deal of thought to photographic archives, observing that ownership of photographs or photographic archives and their subsequent alienation or sale, can have important ramifications for historians and photo researchers:

> ... not only are the pictures in archives often *literally* for sale, but their meanings are up for grabs....This semantic availability of pictures in archives exhibits the same abstract logic as that which characterizes goods on the marketplace.

In other words, regardless of the intent of the photographer, captions and documentary evidence preserved with the image, or attempts by the repository to control or restrict usage, these digital images can be downloaded and used in ways that may be quite antithetical to the original meanings (cf., Margolis, 1994). Ripped free from context, photographs become free floating signifiers that appear to be little snippets of reality and can be used to bolster or "prove" a variety of contradictory theses. (Note 6)

The second warning has to do with meaning of such enormous archives as a whole—that is, with the ontology of the archive. What does it mean to have a media collection called "American Memory?" Jean Baudrillard (1983), the French sociologist, described the developing image world as a "simulacrum," a "hyperreal" media world of copies of copies where there is not and has never been an original. Everything in this symbol system refers to other symbols. Basic to the discussion of photographic archives is Baudrillard's (1983) observation that

> Abstraction today is no longer that of the map, the double, the mirror, or the concept. Simulation is no longer that of a territory, a referential being or a substance. It is the generation by models of a real without origin or reality... (p. 3)

In place of the two-dimensional concepts in written history, we are faced with an (imag)inary model of history. Baudrillard described a world of allusion and trope, maps referring not to territories but only to other maps, news referring to other news, photographs referring to photographs and so on. As millions of photographs are digitized and placed online in the "American Memory," this carefully constructed and selective simulacrum will be thought of more and more as something similar to Durkheim's "*conscience collectif*." (Note 7)

Precisely because of these twin issues, it is vital that scholars begin to seriously explore the photographic data banks (morgues?) that are growing on line. What is in the American memory? What has been forgotten? What survives in unconscious or unexamined form? What is myth, what is reality? Photographic images *do* provide a fresh source of data about our past, but this data has as much power to obscure as it does to reveal. It is essential to temper the "semantic availability" that stems specifically from the conversion of photographs produced with particular use values into commodities with an abstract equivalence dictated by their exchange value, by studying the development of the virtual archive and providing the kind of social and historiographic scholarship necessary to understanding. In this effort it is necessary to study both available meanings and the lacks and oversignifications of the images and the data banks: as I shall demonstrate, whole classes of photographs are not represented, while others exist in such replication and proliferation that they crowd out alternative meanings and critical perspectives. We will need to develop a new paradigm to discuss the developing simulacrum itself. How shall we conceive of a web site with hundreds of thousands of images and documents that calls itself "American Memory?" Is it a thing, a process, a reflection? What research tools might one employ to study such a complex entity and the people who use it?

The Hidden Curriculum in Black and White

This project began as a search for photographs to be used as illustrations for a series of lectures on the history of American education. At first the enormous numbers of photographs of schools, students, and teachers available on line seemed overwhelming. In an evening I found more images than I needed for three lectures. A closer look at the photographs, and the collections that they were found in, raised a simplified set of research questions informed by the issues asked by Alan Sekula: What photographs have been included? How can we understand the meaning of these photographs? What photographs were made that are not in the archives? What was not photographed?

The research on class pictures was theoretically informed by an interest in socialization processes and hidden curricula having to do with the reproduction of race and gender hierarchy (Margolis and Romero, 1998). The term "hidden curriculum" was coined by Philip Jackson after he observed public school classes. He noted the peculiar disciplines and behaviors in classrooms and embedded in school practices that do not necessarily further intellectual development. Jackson (1968, p. 33) observed that students are awarded credit for "trying," rewarded for "neatness, punctuality and courteous conduct," and that negative sanctions are levied for the violation of institutional rules. The concept of hidden curriculum came to refer to the socialization that takes place in school but is not written into the formal curriculum.

Socialization functions of the hidden curriculum have been further analyzed as encompassing three distinct functions. Apple and King (1977) building on the work of Elizabeth Vallance (1973) termed the first two "weak" and "strong": 1) a "weak" Durkheimian concept of the socialization essential to social life —reproducing the connections to civil society that transform children into social beings able to live and work together, form social institutions, and agreed upon meanings; and, 2) a "'strong' sense of control wherein education in general and the everyday meanings of the curriculum in particular were seen as essential to the preserving of the existing social privilege, interests, and knowledge of some elements of the population at the expense of other less powerful groups. Most often this took the form of attempting to guarantee expert and scientific control in society, to eliminate or 'socialize' (acculturate, assimilate) unwanted racial or ethnic groups or characteristics or to produce an economically efficient group of citizens..." (Apple and King 1977, p. 34). Strong controls are highly visible in gender role socialization practices, in segregation and different curricula provided to different racial/ethnic groups and in the reproduction of social classes (Anyon, 1989). The third function of the hidden curriculum is the direct production of ideological belief systems, for example patriotism, certain forms of representative democracy, market capitalism, heterosexual family structures and so on.

While the education literature refers to socialization curricula as "hidden" they are actually quite visible and have readily been photographed. From a critical perspective, class pictures can be viewed as an historical record of certain elements of the hidden curriculum. The photographs show bodies with certain race, gender, age, and ability characteristics spatially arranged in an environmental setting. As social scientists, historians, and educators we interpret these visible relationships as representations of social relations learned about elsewhere: segregation, integration and hierarchy, gender socialization, social class structures. Moreover, we infer that the images were not randomly produced but were carefully fashioned using agreed upon conventions of representation to be symbolic representations of such social qualities and others including: order, discipline, purity, equality, patriotism, and community pride and stability. In these photographs we can see attempts to denote social processes such as socialization, assimilation and acculturation which cannot be directly photographed. Clearly this interpretative enterprise is fraught with peril. Precisely because one cannot actually photograph social relationships, there is a fundamental issue of ethnographic sense making: we cannot be sure if we understand "from the native's perspective" what the project of photographer and her subjects entailed; nor can we ever be sure that our reading is not an error, a misplaced abstraction, or an aberrant decoding.

 BEST COPY AVAILABLE

In the Archives

Once upon a time newspapers called their collections of photographs assembled for the future obituaries of persons still living, "morgues." Now photograph collections are becoming our collective memory. This paper will focus on two of the federal government's major archives each encompassing a number of collections. The various collections were created for different purposes, in different geographic locations, in different historical periods and provide distinct and different views of school life. In essence, much like schools and America itself, the photographic collections are segregated. Separate collections offer divergent and sometimes confusing or contradictory views of race and ethnicity, social class, rural/urban life, and ability/disability. As previously discussed these collections are to some extent losing their identity and becoming submerged in the digital archive. Even though each image retains its citation and whatever provenance exists, the fact that one can search across collections by topic begins a process of homogenization. The National Archives and Records Administration, for example is not organized by collection. There are about 54,000 photographs currently available and nearly 1600 of them can be retrieved with key words "teacher, student, school" (although not all are linked to digital images). Some major collections were discovered this way: photographs from the relocation camps for Japanese Americans, photographs from the Roosevelt Library depicting African American schools in the south, photographs of the Albuquerque Indian Boarding School, and so on.

The "American Memory" site run by the Library of Congress is organized by collection. While one can choose to search the entire site, one can also search each collection individually. The following chart describes some of the collections in the "American Memory" site that have large numbers of photographs of schools.

Table 1
School-related Images Available Through the American Memory Site, January 1999

Components of the American Memory Site	Number of Images in Component	Images Found with Keywords "school," "teacher," or "student"
Architecture and Interior Design for 20th Century America: Photographs by Samuel Gottscho and William Schleisner, 1935-1955.	29,000	1,479
Touring Turn-of-the-Century America Photographs from the Detroit Publishing Company, 1880-1920.	25,000	302
America from the Great Depression to World War II: Photographs from the FSA-OWI, 1935-1945.	56,600	463
Built in America: Historic American Buildings Survey/Historic American Engineering Record, 1933-Present.	35,000	542
American Landscape and Architectural Design, 1850- 1920, A Study Collection from the Harvard Graduate School of Design.	2,800	46
The Northern Great Plains, 1880-1920: Photographs from the Fred Hultstrand and F.A. Pazandak Photograph Collections.	900	30
Taking the Long View: Panoramic Photographs, ca. 1851-1991.	4,000	149
Washington as It Was: Photographs by Theodor Horydczak, 1923-1959.	14,000	374

In what follows, I will show and discuss a small number of photographs drawn from several of these sources. The goal will be to ask what can be learned from the class photographs found in these great archives. This is not an attempt to present a statistical analysis, although we are rapidly approaching the point in sheer numbers where such an undertaking would be fruitful. Rather, it is more a qualitative and ethnographic study in which a few images have been selected as indicative of specific categories and will be quoted and analyzed in an attempt to capture the scope and detail of this source of data. One other note. The archives contain many photographs of school related subjects like sports, recess, school dances, etc. The images selected for analysis are those that would generally be considered "class photographs." Some images were selected because they are representative, but as in the selection of quotations from interviews in more conventional qualitative research, images were frequently chosen because they were unique—particularly articulate, well-composed, and interesting. A number of techniques will be employed in the analysis. Photographs will be compared to other photographs and collections to other collections. Meanings will be elucidated by current perceptions and theories of schooling, as well as by symbolic and literary understandings. Concepts such as status, body language and position, discussed by many analysts of photographs (Goffman, 1976; Solomon-Godeau, 1991; Trachtenberg, 1989), will be utilized. Additional data about the social world in which these photos were made will be brought to bear, for instance, the social settings in which they were produced and consumed. Thus, "class pictures" will be treated as social constructions and will be analyzed using techniques developed in diverse fields including literary criticism, art theory and criticism, semiotics, deconstructionism, ethnography, and symbolic interaction.

White Students

The first public school law in the Dakota Territory was passed in 1883. The Northern Great Plains Collection contains photographs of the rural, one room schools that were built in the townships. These photos from the 1880's and 90's were generally posed outside the school in the sunshine. The shot reprinted here is part of the Fred Hultstrand collection that was donated to North Dakota State University. Hultstrand was born in 1888 and would have been eight when his class picture was taken; he photographed extensively from 1905 through the 1950's, collected photographs of frontier life, and spent much time hand tinting. While these are photographs of real schools, they also helped constitute a pervasive, nearly mythological, image of American public schools. The common school, with its modest architecture, ungraded classrooms, local control, strong community support, and curriculum limited to primary instruction, is often credited with being the backbone of America.

Figure 1 reveals a number of possible meanings. The building in the background is visually less important than the people. There are forty-seven children; boys and girls are not casually mixed, nor were age groups. Everyone dressed for the portrait. Men and boys wore black or somber colors; all the males stand except for three older boys who were posed on horseback. Women and girls were wearing clothes that appear white in the black and white photograph but the hand-tinted copy shows dresses painted in pastel colors. A row of little girls was seated in front in a decorative and passive pose. Overall the people were arrayed in an open semi-circle facing the camera with younger and smaller pupils placed in front and older and larger students and teachers in back. The created image—very much in keeping with the model of the one room school—suggests the older protecting or shielding the younger.

Figure 1. "Soper school, Soper Post Office, North Dakota, 1896 G.G. Grimson, teacher; Mandus, Fred, Bernard Hultstrand" Typical of rural schools found in the Northern Great Plains Collection. Fred Hultstrand History in Pictures Collection, NDIRS-NDSU, Fargo. North Dakota State University Institute for Regional Studies PO Box 5599, Fargo, ND 58105-5599. American Memory, Library of Congress. **(Click on the image to view a larger version.)**

Interestingly the image is also one of equality in that the teachers and adults are standing among the students and not indicating superior status by clothes, body language or position. Despite the fact that a few of the children are barefooted, this is not highlighted as a marker of poverty (but see Figure 6). Images such as these, from Walton's Mountain to Little House on the Prairie, shape an American mythology of a bucolic golden age of schooling that inspires our periodic longing for a return to basics, simplicity, morality and so on.

Everything is not quite what it seems. In the case of white immigrants, nationality and linguistic proficiency are invisible, but according to text at the Hultstrand web site, many of the children were recent immigrants speaking Swedish, German, Norwegian, etc. These meanings disappear in the photographs, as they disappeared in society where white immigrants became invisible through assimilation in a generation. It is important to note that all the people in the photograph are white, not because one would expect racial diversity in the territorial communities of the Northern Great Plains but because "whiteness" is precisely part of the taken for granted quality of the American Common School. (Note 8) It was lucky that the Northern Great Plains collection preserved these particular images, but in doing so the images of specific schools begin to pass over into an archetype of the one-room school. Photos like this raise a question: where were the others? Did the African American, Native American, and Asian communities that existed at that same historical moment in the South, the Northeast, or on the West Coast also educate their children in one-room schools? What did they look like? What kind of historical or cultural amnesia accounts for the fact that these photos are not present in the American memory collection or National Archives? In fact without substantial historical research we do not know if the photos and not present, because they were not made (or not made in the same volume), because they were not preserved, or because they were not archived

Figure 2 is one of more than sixty images from the Detroit Publishing Collection depicting urban high schools. The picture was selected because of the children; most of the other views of urban high schools show buildings only. Despite the rather grandiose title: "Touring Turn-of-the-Century America," images in this collection were not created as an overview of the nation. These views, as they were thought of, were made by professional photographers to be reproduced as postcards — that is, they had to sell. Other than date and location there was little documentation. The collection site describes it this way:

> The Detroit Photographic Company was launched as a photographic publishing firm in the late 1890s by Detroit businessman and publisher William A. Livingstone, Jr., and photographer and photo-publisher Edwin H. Husher. They obtained the exclusive rights to use the Swiss "Photochrom" process for converting black-and-white photographs into color images and printing them by photolithography. This process permitted the mass production of color postcards, prints, and albums for sale to the American market.

According to Bogdan and Marshall (1997, p.6), in the early years of the century more than a billion postcards were mailed each year and many cards depicted architectural monuments and large-scale institutions. They were able, for example, to collect more than sixteen hundred *different* views of asylums and institutions for the mentally ill or retarded.

Figure 2 was made ten years after the North Dakota photograph and contributes a countervailing view of American schools around the turn of the century.

Figure 2. High school, Hancock, Mich. C. 1906 Detroit Publishing Co. American Memory, Library of Congress. **(Click on the image to view a larger version.)**

A large mass of students stand in front of an imposing stone building. While apparently called out of school for the photograph, the students seem to be casually milling around and much less orderly than in the rural school. No teachers or adults are in evidence; neither was an attempt made to arrange the students by size. Here too the students are all white but more homogeneous in age than in the prairie school. The images of shelter and protection are completely missing; in fact, students in the street and lounging against a telephone pole suggest urban toughness and self-sufficiency. Overall this is a photograph of a school; the building was emphasized over the students who form a faceless mass. Comprehensive high schools like this were expensive public works that were sources of civic pride. The high school views were perhaps similar to the mental institution and asylum photos discussed by Bogdan and Marshall (1997) who observed that:

> The initial impression the postcard pictures leave is that these institutions were orderly and therapeutic environments. One way to understand the cards is that they were part of the visual rhetoric of hegemony — they helped manage the public's understanding of the legitimacy of professional control of deviance. (p. 5)

High schools were, of course, not asylums, but when these schools were built and these postcards circulated, the notion of universal high school education was new. Images such as these were reassuring, lending *gravitas* and legitimacy to bold social institutions that were taking professional custody over all children — ending family control and child labor practices that had marked history to this point.

Figures 3 and 4 provide additional insight into the ways in which school and children were imagined at the turn of the century. The two shots were made in the same doorway, presumably the same day. The photographer has used the steps as risers and the doorway as an ornate frame, carefully posing the children to create images of order and obedience. White children are dressed in white, a symbol of innocence and purity. The imposing door, itself a metaphor for the doorway to knowledge, is forbidding. Knowledge is not depicted as an open process of personal growth or something gained in family and community. It is the property of the awe inspiring institution behind the children through which they must pass. The children are ready for the challenge. They stand at attention, equidistant, not quite touching, the girls in bonnets and white dresses the boys in what appear to be uniforms with short pants, leggings, shirts and caps.

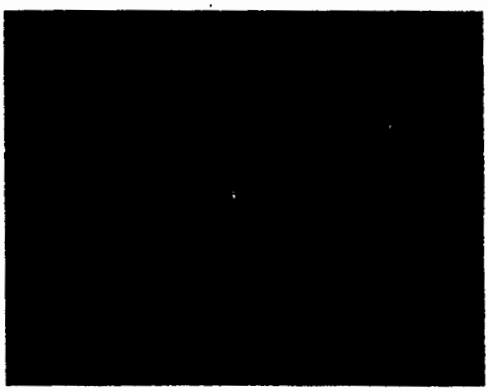


Figure 3. "School Girls" created between 1900 and 1905. Detroit Publishing Co. American Memory, Library of Congress. **(Click on the image to view a larger version.)**

Figure 4. "School Boys" created between 1900 and 1905. Note says "Students holding swords at their sides." Detroit Publishing Co. American Memory, Library of Congress. **(Click on the image to view a larger version.)**

The caption informs us that the little boys have swords at their sides. This is a particularly telling example of the ways in which gendering, one of the strong elements of the hidden curriculum, and school discipline, one of the weak elements, were represented on film.

Even though nearly all of the thousands of photographs of schools in these collections are photographs of white students and teachers, they were not identified as such. Figure 5 is particularly interesting because of its caption which identifies children of white migrant workers. In the United States "white" is the taken-for- granted category. White has been the color of invisibility, the norm, the regular and average (Frankenburg 1993). There are no hits in either "American Memory" or the National Archives site for "white students" or "white teachers." "White schools" produced a single hit from "American Memory," a 1938 Marion Post Wolcott photo of a dark school building with the caption "'White school house, Chaplin, Scotts Run, West Virginia." The National Archive site produced three hits on white school. One was a "Sunday School Indians and Whites" Indian Territory (Oklahoma) 1910. The other two were segregated schools. One photograph from 1941 is a picture of a building with the following caption: "Harmony Community, Putnam County, Georgia.... The Harmony white school was closed down for several years because there were not enough children to make its continued operation worthwhile. Two years ago it was reopened, and last year it had an enrollment of 11, three of whom were from outside the Community. The few high school age children in Harmony go to Eatonton in a bus operated by the County—but no transportation is furnished for children of grade school."

I reproduced the other as Figure 5. "Whites" are only identified as such in opposition to people of color, whereas people of color always have their ethnicity attached as a marker and identifier.

Figure 5. A school for the children of white migrant farm workers is maintained at the Osceola Farm Labor Supply Center operated by the Office of Labor, War Food Administration. This is the second grade, taught by Mrs. Emma Greenwood. Department of Agriculture. Office of Information. Photographer, Osborne 3/1945. **(Click on the image to view a larger version.)**

African American Students

As Elliott Eisner (1985, p. 97) suggested, it is important to consider the "null curriculum"—that which is missing. It is, of course, not news that schools in turn-of-the-century America were segregated by race and ethnicity. But complete invisibility is surprising. *None* of the 300 school photos in the Detroit Photography collections showed Black, Native American or Asian children in school. If children of color were not in school, it occurred to me to look for them elsewhere in that collection. Searching the 25,000 images of the Detroit collection for "Black Children" yielded half a dozen photographs. Figure 6 is typical of these stereotyping images. This is not a candid shot nor is it documentary; it was made by the same postcard company that posed the White children in the doorway. These four children were also posed, arrayed in a line in front of their house. The image constructed had the intention of emphasizing their "otherness." They were not dressed-up, even though they may well have owned Sunday-go-to-meeting clothes. They were not posed in a meadow where bare feet might have been read as a youthful or romantic symbol.

Figure 6. "Four black children in yard" from the Detroit Publishing Collection. Created between 1890 and 1910. American Memory, Library of Congress. **(Click on the image to view a larger version.)**

The tableau of clapboard house and fence with clothes thrown over emphasizes their poverty. Photographic postcards of African Americans, produced for white audiences, were not as overtly racist as the popular cartoon cards of alligators, pickaninnies, and mammies (Turner 1994; Mellinger 1992). Still, Figure 6 is a clear example of what Turner termed "contemptible collectibles," postcards produced for white consumers that conformed to certain racialized stereotypes: Black children were frequently photographed outside dressed in rags and tatters. As Turner (1994, p. 16) observed: "Picture postcards featuring poorly dressed little black children romping in cotton fields suggests that if they had been given a choice, they would have chosen to spend their days in the field rather than in the schoolroom." Images of diligence, order, and innocence were never included.

Curiously, while the "American Memory" site allows one to search a large number of individual collections as a group, the "The African American Odyssey," which is part of the site, must be visited separately and is not searchable for photographs. (Note 9) An expanded search of the entire "American Memory" collection for "Negro Children" produced about fifty hits, all the photographs of African American students, teachers or schools dated from the Farm Security Administration collection in the 1930's. Figure 7 is representative of a series made

BEST COPY AVAILABLE

by Marion Post Wolcott at Prairie Farms school in Montgomery Alabama in 1939. Germany was already making war in Europe and the worst days of the depression were behind America.

Figure 7. Spring 1939. "Primary class in new school, Prairie Farms, Montgomery, Alabama" Marion Post Wolcott, photographer. Still Picture Branch (NWDNS), National Archives. **(Click on the image to view a larger version.)**

The job of Farm Security Administration photographers was shifting from the focus on depression misery to an emphasis on America's strength and resiliency. By the 1930's, advancing photographic technology made it easier to take photographs inside, and the image Post-Wolcott made shows African American students seated reading at a table with their African American teacher standing over helping a student. The class is small with books and tables and chairs instead of rows of student desks. Boys and girls seem to be working together, perhaps reading. The choice of a new and apparently well-equipped but segregated school creates an affirming vision of Black America as "separate but equal." The photograph similarly creates an image of teaching as an active and caring activity. Other images in Post Wolcott's proof sheet include playing basketball and volleyball in which the teacher also takes an active role.

A better source for historic photographs of African Americans in school is the Schomburg collection of the New York Public Library. The Schomburg offers a searchable archive of 19th Century images of African Americans. A search of the "education" category produced fifty images. The earliest of these are wood block engravings made for *Harper's Weekly* and published in the 1870's.

Figure 8. "Class of school children posing outside with their teacher, Espy, Pa. Spring 1912. Thomas 8, Eleanor 7, donor: Eleanor Drayton," Photographs and Prints Division, Schomburg Center for Research in Black Culture, The New York Public Library Astor, Lenox, and Tilden Foundations. **(Click on the image to view a larger version.)**

They show "freedom schools" for emancipated slaves. There are a number of photos of famous educators, Booker T. Washington, for example, and there are many photographs from the historical black colleges: Hampton Institute and Tuskegee Institute. Figure 8 is unique in depicting what is apparently an integrated school class in Pennsylvania in 1912. It is one of those important images that appear in historic collections with inadequate captions and provenance. The caption identifies the photo as a gift to the Schomburg by Eleanor Drayton, and we might assume that she is the Eleanor age seven in the photo. Other meanings are more problematic. Thirty students including Whites, Blacks and apparently Non-White ethnics (Native Americans? Eastern European immigrants?) were clustered together shoulder-to-shoulder on a bleacher with the African American teacher standing on the left with her arm symbolically embracing the entire class. The students probably dressed for the photograph. They do not seem sorted by race or gender. This is the clearest image of equality and diversity that I found in any of the collections searched. A good deal of research would be necessary to discover whether integrated classes were common in Espy, Pennsylvania in 1912, or if the Anglo-appearing students were immigrants whose "otherness" set them apart as well.

Latinos were even more invisible than Blacks in schools. A search of the American Memory collection for Spanish American, Puerto Rican, or Mexican schools, teachers or students yielded nothing before a single image taken by Russell Lee in July 1940 with the caption: "Spanish-American farmer who is also justice of the peace and teacher in local grade school, Chamisal, New Mexico." (Note 10) The National Archive site produced a group of seven shots taken by Irving Rusinow for the Bureau of Agricultural Economics in Peñasco, New Mexico in 1941. Figures 9 and 10 are representative selections from that shoot.

Figure 9. Taos County, New Mexico. Children play in the Peñasco schoolyard. Photographer, Irving Rusinow, December 1941 Department of Agriculture. Bureau of Agricultural Economics. Still Picture Branch (NWDNS), National Archives. **(Click on the image to view a larger version.)**

In Figure 9, a long low adobe school building stands against a line of arid mountains in the background marking the geography as the Southwest and establishing a Spanish feel. The students are clearly aware of the camera; some appear to have been posed in a circle holding hands, others are wandering around as if at recess. Overall this is not an image of order like Figures 3 and 4, or of the specific relations of teaching and caring evidenced in Figures 7 and 8. In place of order, book learning or scholarship, we see playfulness. A Dominican nun approaches the circle from the right, but she is not working with or embracing the students. The image is especially interesting because of the caption: "School was built by the Catholic Church, then deeded over to the State, and most of the teachers are Catholic Sisters, though this is a public school. Sisters' salaries are paid by the State directly to the Church. Though religious teaching does not take place during the regular school period, the Sisters "naturally express the Catholic way of life, and by association with them the children cannot but receive some of the religious essence." (Father Morgan)

In the last half of the 19th century Spanish speaking families in the southwest tried to escape anti-Mexican sentiments, and in particular "English only" school requirements, by sending their children to Catholic schools that they found more welcoming and less hostile to their culture. The situation reported by the photographer Rusinow suggests that by the middle of the 20th century the state was beginning to reassert control.

Figure 10, depicting a class of older female students in a home economics class, is a familiar image of women's traditional gender roles. The young women are apparently making clothes for dolls as Christmas presents. Sex segregated home economics classes are a form of vocational education, preparing Mexican-American girls to be domestics and mothers. (Note 11) Similar pictures were made regularly at the Indian boarding schools showing Indian girls using sewing machines or cooking.

Figure 10. Taos County, New Mexico. Home Economics class at Peñasco High School make toys for Christmas. Photographer, Irving Rusinow December, 1941. Department of Agriculture. Bureau of Agricultural Economics. Still Picture Branch (NWDNS), National Archives. **(Click on the image to view a larger version.)**

Native American Students

Compared with other racial/ethnic groups, Native American Indians were dramatically over-represented in the photo archives. They were frequently photographed as part of the documentation of federally-funded Indian boarding schools, and as official records these images were preserved in large numbers. The American Memory site produced about sixty hits on "Indian School" and the National Archives and Record Center site yielded 106. Figure 11 is a panorama of the Mt. Pleasant Indian Industrial School. In the collection, there are a number of additional panoramas showing the buildings and grounds of Indian Schools in Phoenix, Arizona, Santa Fe, New Mexico, and Carlisle, Pennsylvania, and other places. The Mount Pleasant panorama is an interesting composition. Female students in white dresses were placed in small groups and circles around the grounds.

Figure 11. Mt. Pleasant Indian Industrial School c. 1910 Taking the Long View: Panoramic Photographs, 1851-1991. American Memory, Library of Congress **(Click on the image to view a larger version.)**

The image of the "industrial school" belies its name by presenting a peaceable view of grounds including a formal pond and young girls holding hands ("Ring-a-ring-a-roses, A pocket full of posies"). While clearly we are looking at an institution, nothing in the image tells us that Mount Pleasant was an "Indian" school. The pastoral scene, manufactured by architecture, costume, gendering and photography, suggests gentility and civilization without any hint of the struggle for the hearts and minds of Indian children: removed from family and community; locked in this institutional compound; sent to boarding school to become White.

Indians were subjected to forced regimes of acculturation/assimilation unique in American history. Students were taken far from their parents and community, had their hair cut, were required to wear Euro-American dress and forbidden to speak their mother tongue. Alongside quasi-military discipline, cultural "re-education," and cleverly articulated attempts at cultural genocide to "Kill the Indian and save the man," Indian schools provided vocational training, art and music education, and sports. (Note 12) These were well-funded federal institutions with a coherent curriculum. Compare the Indian school movement, for example, with the treatment of African Americans who were denied schooling in the South until the end of slavery. Although some northern abolitionist women teachers opened "freedom schools" for freed slaves, there was no federal program to provide education to emancipated

Blacks. Instead, southern states rights imposed the jim crow system of segregated schools, and northern urban school districts were segregated "de facto" by housing practices and gerrymandered districts. The legacies both of the Indian boarding schools and of segregation have yet to be overcome.

Figure 12 depicts young Indian Boys at the Albuquerque Indian School. The image is one of symmetry and order. Wearing uniforms and holding American flags, the children were posed quite formally, arrayed as a design around an Anglo American woman (teacher? supervisor? guard?) who stands in the center of the composition. Uniforms are a very important element both of the schooling experience and of the photographic images.

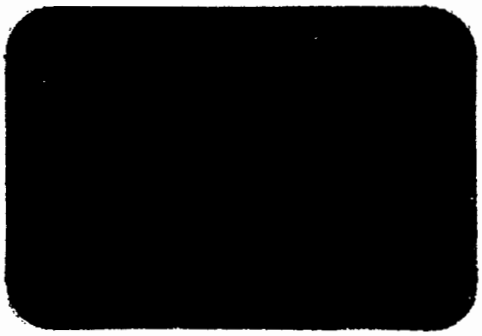


Figure 12. Very early class of young boys with flags at the Albuquerque Indian School. c.1895 National Archives and Record Center. Still Picture Branch (NWDNS), National Archives. **(Click on the image to view a larger version.)**

Uniforms were part of the original concept for Indian schools: Captain Richard Pratt who originated the concept, dressed the losers in uniforms similar to the cavalry that defeated them, and then regimented them like soldiers (PBS Video, 1991). In the photo, uniforms submerge individuality and produce an image of both conformity and interchangeable parts; moreover, they accomplish what Goffman (1976, p. 32) termed "function ranking" removing any ambiguity or status inconsistency. They also serve to strip the children of their native identity. (Note 13) The woman's lack of uniform makes her the only individual and sets her apart. Taller than any of the children, eyes fixed firmly on the lens, the woman holds her arms stiffly at her side. In the midst of a group she stands alone, not touching any of the students. Her position is quite different from the teachers in Figure 1, who stand among the students but to the side and are depicted on the same level; or Figure 7 where the teacher seems to make a gesture of inclusion; or Figure 8 where the teacher is symbolically lowering herself to the students' level. The caption material in the Archive reads: "This is one of a small collection of photographs of the Albuquerque Indian School, which was established in 1881 to provide off-reservation industrial training to the Indians of the Southwest. By 1912, the school had 8 primary grades and over 300 students; by 1925 enrollment increased to over 800 students and grades 11 and 12 were added. The Albuquerque Indian School continued operating until 1982, when its program was transferred to the Santa Fe Indian School."

As the photographs make clear, the Indian school's curriculum of socialization and acculturation was not at all hidden. They were consciously created as industrial training centers to train the students for working class occupations and jobs in white society. The fact that most returned to reservations where these jobs did not exist was conveniently overlooked.

Asian Students

A small set of photos of Chinese children emerged from a search of the "American Memory" site. Figure 13 is representative of a single shoot showing an unnamed group of Chinese at about the turn of the century posed on a rooftop. They were made by the famous western photographer William Henry Jackson.

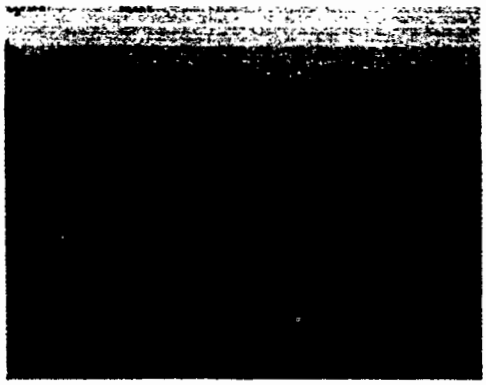


Figure 13. "Chinese Subjects" Photograph by Wm Henry Jackson c. 1901. No location given. American Memory, Library of Congress. Detroit Publishing Co. American Memory, Library of Congress. **(Click on the image to view a larger version.)**

He made a number of exposures of the same family, but he left no firm date, location, or discussion of the occasion for the shoot. (Note 14) I find these photographs similar to the Detroit collection's images of Blacks: they share the stereotyping feel of photographs of the exotic "other." The first segregated school for Chinese students was opened in San Francisco in 1885, and rigid segregation was enforced until 1905 when the board of education allowed Chinese students into a regular city high school. In 1906 a separate school in San Francisco was established for Japanese, Korean, and Chinese children (Spring 1997, p. 76). Perhaps photographs of these schools can be found in California.

No photographs of Japanese children appear in either American Memory or the National Archives collections before 1941, when a slew of photographs were made to document the relocation procedure. Dorothea Lange and other Farm Security Administration (FSA) photographers were now working under the auspices of the Office of War Information (OWI) and completed assignments to show Japanese students in California schools and orphanages on the eve of relocation. This was followed by a long term campaign to document the internment camps, including many images showing Japanese children in school in Tule, Manzanar, Salt River and the other sites. Figures 14 and 15 are representative of these efforts. Figure 14 was made by Dorothea Lange at an integrated San Francisco public school with large numbers of Japanese students.

Figure 14. "San Fransisco, California. Flag of allegiance pledge at Raphael Weill Public School, Geary and Buchanan Streets. Children in families of Japanese ancestry were evacuated with their parents and will be housed for the duration in War Relocation Authority centers where facilities will be provided for them to continue their education." 04/20/1942 Department of the Interior. War Relocation Authority. Photographer, Dorothea Lange. Still Picture Branch (NWDNS), National Archives. **(Click on the image to view a larger version.)**

The occasion was the rounding up of Japanese families so that they could be shipped to relocation camps. The choice of patriotic images, saluting the flag, clearly advanced a view of Japanese as patriotic and law-abiding Americans. Figure 15, also by Lange, shows a school class in Manzanar. Students with what appears to be a Japanese teacher are hard at work reading and writing. They have the same sort of modern desks and chairs that can be seen in Figure 7. The hidden curriculum portrayed in the relocation photographs is an unabashed patriotism illustrative of school's role in the direct reproduction of ideological belief systems. As the captions indicate, these are a class of photographs taken not to showcase school children but demonstrate to the world that the United States relocation camps for Japanese citizens were much different from concentration or POW camps. They featured images of well-equipped schools, caring teachers, and happy willing students.

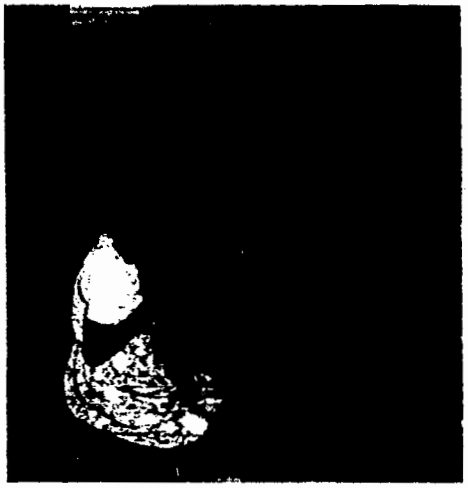


Figure 15. "Manzanar Relocation Center, Manzanar, California. These young evacuees are attending the first elementary school at this War Relocation Authority center. There are six grades with volunteer teachers and voluntary attendance." 07/01/1942 Department of the Interior. War Relocation Authority. Photographer, Dorothea Lange. Still Picture Branch (NWDNS), National Archives. **(Click on the image to view a larger version.)**

There is another hole in the American Memory. Children with disabilities were as invisible as children of color. Based on my survey of these two mega-archives, America's photographic images of schools, and the historical memories they engender, consist nearly entirely of able- bodied white children and teachers. A search for deaf schools retrieved a single Detroit Publishing view of the outside of the "Deaf and Dumb School, Columbus, Ohio." This was a familiar institutional view with no persons present. The search also retrieved and a number of potential ("not yet digitized") photos of deaf and dumb schools from the Historic American Buildings Survey/Historic American Engineering Record. These too are likely to be photographs of architecture. Figure 16 from the National Archives is the only photograph found in either site depicting crippled, deaf or blind children in school. It is interesting that this photograph was attributed to Franklin D. Roosevelt who was himself crippled by polio.

Discussion

Figure 16. WPA, "Blind children at work in Art Center Workshop in Salem, Oregon." Creating Indiv. Roosevelt, Franklin D. Federal Arts Project; Works Progress Administration; 1941. Still Picture Branch (NWDNS), National Archives. **(Click on the image to view a larger version.)**

Along with all the other historical photographs in the archives, class pictures are becoming part of a modern hidden curriculum as well. Web access in schools is making historical photographs into a "curriculum" of primary source materials for teachers preparing classes and for students doing projects. There are many implications of this development for how students are taught about American history in general and specifically about the schools and students who came before them. As the preceding analysis demonstrates, the record encompassed by these photographs is full of holes. Some views are over-represented, while whole groups of students and types of schools are simply absent. This is likely the case whether one searches for school-related photographs or photos in any other category.

There are two central issues in the implied critique of the World Wide Web: In the first place one must consider what exists at this point in time. Clearly the historical photograph collections currently available on line reproduce the familiar historic amnesias, lapses and sins of omission, while continuing to overemphasize powerful, dominant and hegemonic structures. In this way it resembles the historiography of the first half of the 20th century with its great men theories and inattention to workers, to women, and to people of color. The photo archives

valorize assimilation models, a peaceful bucolic past, upward mobility, and order at the expense of cultural diversity, domination and conflict. The second question has to do with the potential of the Web to offer a different vision. Because it is global, decentralized, and offers open access it is quite probable that some of these deficiencies will be overcome. If archives are opened to images from all sources: personal collections, small local history societies, private collectors, newspapers and so on, it is easy to imagine that a search for schools, teachers, students would return a far more heterogeneous selection.

However, even if *all* extant photographs of schools were to be made available as digital on-line images, we would still be confronted by the deficiencies of photography itself. Many things were not photographed. I found, for instance, no views of teacher unions or organizing activities, no photographs of school boards or teacher meetings where the central decisions shaping schooling were made. There were no photographs of conflicts and tensions in schools—between teachers and students, among students, between school boards and communities. No pictures of discipline and punishment. No photographs of boredom. And even if such photos did emerge, they would not solve the central problem of the photograph; photography is powerless to represent some things. I argue in an earlier piece (Margolis, 1999) that it is not possible to photograph social relationships. My example in that article was, that although photographs could represent the coal mining process and technological divisions of labor, they could not capture the social relations of production which remain invisible: ownership, alienation, exploitation, fear, and so on. Similarly, photography can capture the physical relationships of schools, but cannot make visible the social relationships of education: failure, intellectual excitement, oppression, resistance, or teaching/learning. These are multidimensional concepts that cannot be reduced to a visual icon.

Recognizing the inherent limitations of visual images is critical if one intends to use them as other than propaganda vehicles. Given that, there are many ways that photographs can be used by historians of education, not just as illustrations to make textbooks and lectures visually interesting, but as primary source data. The preceding analysis should be taken as only suggestive, as most of the issues raised need to be investigated on their own and in more depth. This paper is meant simply as a provocative introduction, indicative of new avenues for educational research. In effect, it opens a space analogous to an environmental niche which can be explored and settled in a number of ways. As suggested earlier, there is room for the application of additional analytical techniques including quantitative methods to many of these issues. One might ask questions about the frequencies and ratios of certain types of representations, and about their correlations. It should be possible to statistically compare geographic regions and/or historical periods. It is likely that changes in representation can be seen over time. For example, one might hypothesize that the number of photographs showing integrated classrooms increases since 1954.

In some locations there exists a fairly dense and complete photographic record, allowing a kind of retrospective rephotography project to be done. Researchers could collect and arrange in sequence photographs taken at the same school over decades in order to examine and analyze social change. For instance, some of the Indian schools appear to have left a fairly detailed photographic record from the 1880's through the 1930's. It would be interesting to examine the change in these images over half a century. (Note 15) Additionally, much might be learned from cross-cultural investigations. One might compare, for example, images of order and discipline in class pictures taken in England, Japan, and the U.S.

As well as asking diachronic and comparative questions, synchronic questions need to be addressed in more depth. Careful historical analysis of the people, places, and occasions photographed is necessary. What can be discovered about the actual school, the children, teachers and communities? What can be learned about the photographers, the occasions upon which the photographs were made? How can other documentary evidence shed light on the images, and vice versa? The conventional touchstones of historical research: newspapers, school and government records, census data and so on need to be consulted and cross-referenced with the images (Margolis 1988). (Note 16) Where possible, it might be extremely fruitful to

employ oral history and ethnography to gather additional information. It seems likely that it would still be possible to find and interview students depicted in pictures made in the 1930's, for instance. More recent history, for example the period following Brown v. The Board of Education in 1954, could be even more useful both because participants are available to study and because the sheer volume of photographs probably increased. The techniques of the visual anthropologist — photo elicitation, inventories of various types and surveys can be employed to examine, for instance, issues relating to the inequalities of "separate but equal" or same gender schools (cf., Collier and Collier 1986).

It should also be clear that the study of school photographs is not only a historical undertaking. Social science researchers can examine current collections of school photographs: year book photographs, sports pictures, class pictures, the huge collections of snapshots and vernacular pictures found in virtually every school. One might do interesting research simply with the bulletin boards (or more recently web sites) found in many grade school classes. These constitute different simulacra, image worlds manufactured by students, parents, and school personnel. These images can be studied in much the same way, examining both the actual occasions and intentions governing the production of the photographs, the apparent symbolic meanings, and selection, juxtaposition and arrangement for display. Photographs produced as part of school culture, like historical photos, can be analyzed as icons with symbolic, iconic and indexical meanings. (Note 17)

Notes

The author would like to acknowledge Jon Wagner and Mary Romero whose comments on earlier drafts of the article were extremely helpful in framing the argument. Marina Gair helped with copy editing on the final draft and obtaining photograph permissions. This article first appeared as pages 7-38 in "Seeing Kid's Worlds," a special issue of Visual Sociology (14), 1999. Additional information about Visual Sociology and the "Seeing Kids' Worlds" special issue can be found at the web page for the International Visual Sociology Association (IVSA): http://www.sjmc.umn.edu/faculty/schwartz/ivsa/

1. There are sixteen images included in this article and most of them are photographs of classes although they were not all examples of school photography Three were produced by a professional photography company to be reproduced as postcards (Figures 2, 3 and 4). Eight of the photographs were made by various government documentary projects (Figures 5, 7, 9, 10, 14, 15, and 16). Two of the images (Figures 11 and 12) were part of the ongoing photographic documentation of the federally financed Indian boarding schools. Two of the images were not school photographs at all, but an attempt to find photographs of children of color who did not appear in any of the class pictures: Figure 6 was made as a postcard and shows four African American children and Figure 13 is a William Henry Jackson photograph of Chinese children. Figure 8 depicting an integrated class in Pennsylvania in 1912 appears to be a school photograph, but has little provenance to clearly identify its genre.
2. For a useful review of some of the issues of searching for photographic images see Steiner, Kathy "Finding Photographs."
3. "The most thorough audience appraisal resulted from an end-user evaluation conducted in 1992-1993. Forty-four school, college and university, and state and public libraries were provided with a dozen American Memory collections on CD-ROMs and videodisks. Participating library staff, teachers, students and the public were polled about which digitized materials they had used and how well the delivery systems worked. The evaluation indicated continued interest by institutions of higher education as well as public libraries. The surprising finding, however, was the strong showing of enthusiasm in schools, especially at the secondary level." American Memory pilot—seed of a universally available Library http://lcweb.loc.gov/ndl/nov-dec.html#pilot
4. Where historians and social scientists have typically used photographs to illustrate reconstructions of the past that are entirely language based, I have

advocated the use of photographs as primary source material. For many years I have been collecting, paying attention to, and thinking about American historical photographs. This work was expensive to undertake and extremely labor intensive. It required traveling to libraries, museums and photograph collections and obtaining permission to make copies -- using a film camera and copy stand to photograph each image. Much of this work was part of a study of coal miners for which approximately 12,000 historic photos were collected from archives all over the country (Margolis, 1988; 1994; 1998). Cataloging, studying and working with a collection that by necessity included slides, prints, and negatives all associated with data about captions and provenance has been a very slow and inefficient process. This process is rapidly becoming as obsolete as the card catalog, handwritten note card, and carbon paper. In a few minutes once can visit a web site, search thousands of images by keyword, download the images one is interested in and paste them into your document.

5. Many critics of the image have drawn attention to problems inherent in photography and the creation of a mass culture "image world:" cf. Adorno and Horkheimer 1973), Baudrilliard (1983), Rossler (1990), Sekula (1990), and, Solomon-Godeau (1991).
6. An anonymous reviewer correctly pointed out the reflexive confirmation of this quality of the image with the observation that: "One possible 'antithetical' use, of course, is the sort that occurs in this article: critical social analysis of pictures not made with this purpose in mind."
7. People have misconstrued Durkheim's notion of "collective consciousness" to mean some kind of group mind. But this is inaccurate. The engendering of collective consciousness is both an abstract and theoretical lesson and a practical activity. It is represented in collective action and in schools, libraries, museums, repositories, and now the Internet: "Society is not the work of the individuals that compose it at a given stage of history, nor is it a given place. It is a complex of ideas and sentiments, of ways of seeing and feeling, a certain intellectual and moral framework distinctive of the entire group. Society is above all a consciousness of the whole. It is therefore, this collective consciousness that we must instill in the child" (Durkheim, [1925]1961:277).
8. Many scholars have been working to make whiteness into a visible category. See Frankenberg, 1993 for one of the pioneering analyses of whiteness.
9. The site described it this way: "This Special Presentation of the Library of Congress exhibition, The African American Odyssey: A Quest for Full Citizenship, showcases the Library's incomparable African American collections. The presentation is not only a highlight of what is on view in this major black history exhibition, but also a glimpse into the Library's vast African American collection. Both include a wide array of important and rare books, government documents, manuscripts, maps, musical scores, plays, films, and recordings. This presentation is not yet searchable."
10. I actually have seen many photographs dating from the turn of the century or before that show Mexican and Spanish American children in school. Such photographs can be found in nearly every state historical society, local history museum and library in the Southwest. As is no doubt the case with the other racial/ethnic groups it is not the absolute lack of photographs that is problematic. It is the curious selection process that has produced the simulacrum of "National Archives" or "American Memory" that is the issue. Moreover, the problem that is so obvious in photographs of school is no doubt present in many other categories.
11. As Mary Romero pointed out: "The 'cult of domesticity' advocated sex roles that were not really applicable to working-class Mexican Americans whose economic circumstances did not allow the maintenance of gender-specific spheres of activity—that is, women in the private sphere of the home and men in the public sphere of production and trade." Programs such as this did, however, produce trained and "Americanized" domestic workers to work for nearby Anglo families (Romero,1992, p. 81-82).
12. These were the words of Captain Richard Pratt who established the Carlisle Indian School. He believed in subjecting Native American youth to

quasi-military discipline: uniforms and drill exercises alongside instruction in English and industrial training. (Cf. PBS Video: In The White Man's Image, 1991).

13. This is equally true for the children posed in Figures 3 and 4. Uniformed students are the perfect images of "product" for the industrial-efficiency model of schooling that was the hallmark of late 19th early 20th century education.
14. In one of the shots there is a sign and the caption: "On tablet (translated from Chinese): Today we are the owners of money, yesterday we were the owners of the territory" which may suggest the occasion for the shoot.
15. The technique of using photographs taken over time to examine social change was pioneered by Mark Klett and given a more sociological interpretation by Jon Rieger (Klett et al, 1984; Klett 1991; Rieger, 1996;)
16. Such research might provide important information for the interpretation of Figure 7, for example.
17. Semiotics, the science of signs, has developed a complex and highly technical language that can be useful in the interpretation of photographic images. Images and texts are analyzed along multiple dimensions described as Indexical (pointing), Iconic (representative) and Symbolic (cultural) meanings. Serious students might consult the works of Ferdinand de Saussure, Umberto Eco or Roland Barthes. These analytic tools of semiology can also be employed in the construction of images designed to produce certain impacts. See for example: Nadin, Zakia, and Nadin, (1995

References

Adorno, T.W. and Max Horkheimer (1973). The Culture Industry: Enlightenment as Mass Desception. In James Curran et. al. (Eds) *Mass Communication and Society* London: Edward Arnold.

Anyon, J. (1989). Social Class and the Hidden Curriculum of Work. In J. H. Ballantine (Ed.), *Schools and Society* (pp. 257-279). Mountain View, CA: Mayfield Pres.

Apple, M. & King, N. P. (1977). What do schools teach? In R. H. Weller (Ed.), *Humanistic Education* (pp. 29-63). Berkeley, CA: McCutcheon.

Baudrillard, J. (1983). The Precession of Simulacra. *Art and Text*, no. 11, September, 3-47.

Bogdan, R. and Marshall, A. (1997). Views of the Asylum: Picture postcard depictions of institutions for people with mental disorders in the early 20th century. *Visual Sociology, 12*, 4-27.

Collier, J. and Collier, M. (1986). *Visual Anthropology: Photography as a Research Message Method.* Albuquerque, NM: University of New Mexico Press.

Digitized Artifacts are Making Knowledge Available to All, On Line. (1998, November 29). *The New York Times*, p. 21.

Dreeben, R. (1986). *On What is Learned in School.* Reading, MA: Addison-Wesley.

Durkheim, E. (1961). *Moral Education.* (Everett K. Wislon and Herman Schnurer, Trans.). New York: The Free Press. (Original work published 1925)

Eisner, E. (1985). *The Educational Imagination* (2nd ed.). New York: MacMillan Publishing.

Frankenberg, R. (1993). *White women, race matters: the social construction of whiteness.* Minneapolis: University of Minnesota Press.

Goffman, Erving. (1976). *Gender Advertisements.* New York: Harper and Row.

Jackson, P. (1968). The daily grind. In P. Jackson (Ed.), *Life in classrooms* (pp. 33-37). New York: Holt, Rinehart and Winston.

Klett, M. *et al.* (1984). *Second View: The rephotographic survey project.* Albuquerque, NM: University of New Mexico Press.

Klett, M. (1991). *Rephotographing Oklahoma 1889/1991.* Oklahoma City, OK: Oklahoma City Art Museum.

Lesiak, Christine an Jones, Matthew (producers). (1991). *In the White Man's Image* [PBS video]. (An episode of the series on "The American Experience").

Margolis, E. (1998). Picturing Labor: Photographs of the Coal Mine Labor Process. *Visual Sociology*, 13 (2), 5-35.

Margolis, E. and Romero, M. (1998). "The Department is Very Male, Very White, Very Old, and Very Conservative": The Functioning of the Hidden Curriculum in Graduate Departments. *Harvard Educational Review, 68* (1), 1-32.

Margolis, E. (1994). Images in Struggle: Photographs of Colorado Coal Camps. *Visual Sociology, 9*, (1), 4-26.

Margolis, E. (1988). Mining Photographs: Unearthing the Meaning of Historical Photos. *Radical History Review, 40* (January), 32-48.

Mellinger, W. M. (1992). Representing Blackness in the White Imagination: Images of 'Happy Darkeys' in Popular Culture, 1893-1917. *Visual Sociology, 7* (7), 3- 21.

Nadin, M. Zakia, R.D. and Nadin, M. (1995) Creating Effective Advertising : Using Semiotics Consultant Press.

Rieger, J. H. (1996). Photographing Social Change. *Visual Sociology, 11* (1), 5-49.

Romero, M. (1992). *Maid in the U.S.A.* New York: Routledge.

Rosler, M. (1990). in, around, and afterthoughts (on documentary photography). *The Contest of Meaning.* Cambridge, MIT Press.

Sekula, A. (1990). *The Contest of Meaning.* Cambridge: MIT Press.

Sekula, A. (1983). Photography Between Labor and Capital. In Benjamin H. D. Buchloh and Robert Wilkie (Eds.), *Mining Photographs and Other Pictures: A Selection from the Negative Archives of Shedden Studio, Glace Bay, Cape Breton. 1948-1968* (p. 194). Halifax, Nova Scotia: The Press of the Nova Scotia College of Art and Design.

Solomon-Godeau, A. (1991). *Photography at the Dock: Essay on Photographic History, Institutions, and Practices.* Minneapolis: University of Minnesota Press.

Spring, J. (1997). *Deculturalization and the Struggle for Equality* (2nd edition). New York: McGraw Hill.

Steiner, Kathy. (October 15, 1995). Finding Photographs. (Kathy Steiner/ILS 603)/Besser). Can be found on the Web: http://www.sims.berkeley.edu/impact/f95/Papers- projects/Papers/steiner.html

Tractenberg, A. (1989). *Reading American Photographs: Images as History Matthew Brady to Walker Evans.* New York: Hill and Wang, The Noonday Press.

Turner, P. A. (1994). *Ceramic Uncles and Celluloid Mammies: Black images and their influence on culture.* New York: Anchor Books.

Vallance, E. (1973-1974). Hiding the hidden curriculum. *Curriculum Theory Network, 15.*

Web Addresses

Library of Congress:
http://memory.loc.gov/ammem/phcoll.new.html

National Archives and Records Administration:
http://www.nara.gov/nara/searchnail.html

New York Public Library Schomburg Center for Research in Black Culture: Digital Schomburg:
http://digital.nypl.org/schomburg/images_aa19/

The Hultstrand site:
http://memory.loc.gov/ammem/award97/ndfahtml/hult_home.html

The Detroit collection:
http://memory.loc.gov/ammem/detroit/detcoll.html

About the Author

Eric Margolis
Division of Educational Leadership and Policy Studies
College of Education
Arizona State University
Tempe, Arizona 85287-1411

Email: eric.margolis@asu.edu

Eric Margolis is a sociologist and Assistant Professor in the Division of Educational Leadership and Policy Studies, Arizona State University College of Education. Recent publications include: The Department is Very Male, Very White, Very Old, and Very Conservative: The Functioning of the Hidden Curriculum in Graduate Sociology Departments (with Mary Romero) *Harvard Educational Review* (Vol. 68 1998); AIDS Research/AIDS Policy: Competing Paradigms of Science and Public Policy, *Research in Social Policy,* Vol. 6 (1998) JAI Press; and *Hidden Curricula in Higher Education* (in press) Routledge.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 240 times since July 11, 2000

Education Policy Analysis Archives

Volume 8 Number 32 July 11, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Student Assessment as a Political Construction: The Case of Uruguay

Luis Benveniste
The World Bank

Abstract
This article reveals the interplay between assessment policies in Uruguay and the nature of State-societal relations. The central State has been historically a staunch defender of public education and has championed the cause of equalizing opportunities for the most disadvantaged sectors of society. The national evaluation system of student performance has been constructed as an expression of this tradition. The Uruguayan government sought to build a wide level of consensus with respect to the assessment instruments by encouraging educators to participate and buy into the assessment initiative. Moreover, the national government shifted the focus of the national evaluation from measuring schooling outcomes to addressing the social wants that condition student learning. Hence, the national evaluation has come to symbolize an agreed-upon mechanism of social accountability by which the central government upholds its responsibility for educational provision as it intervenes on behalf of impoverished communities. (Note 1)
This article is also available in Spanish in Adobe Acrobat format at http://www.grade.org.pe/gtee-preal/docs/Benveniste.pdf

BEST COPY AVAILABLE

This study reveals the interplay between assessment policies in Uruguay and the nature of State-societal relations. The central State has been historically a staunch defender of public education and has championed the cause of equalizing opportunities for the most disadvantaged sectors of society. The national evaluation system of student performance has been constructed as an expression of this tradition.

The first section describes the educational system as a highly centralized organizational structure. Then, it provides a brief overview of the education reform initiative launched in 1995 by the National Administration of Public Education to promote and consolidate social equity.

The second section portrays the *Unidad de Medición de Resultados Educativos* (the evaluation agency of primary education) as a temporary unit created in 1996 within the framework of a project financed by the World Bank. In spite of its short history, the assessment system has garnered substantial popular support and spurred a curricular and pedagogical renovation among teachers, principals and supervisors.

The third section explores the reasons behind the public embrace of the national assessment system. This has been no slight accomplishment in light of the fact that the evaluation of student performance may potentially exert a destabilizing role by highlighting deficiencies in educational service provision. First, the central State circumscribed teacher liability over poor performance, largely assuming itself the responsibility for the character of schooling. Second, the national government built a wide level of consensus with respect to the assessment instruments by encouraging educators to participate and buy into the assessment initiative. Third, the national government shifted the focus of the national evaluation from measuring schooling outcomes to addressing the social wants that condition student learning. Hence, the national evaluation has come to symbolize an agreed-upon mechanism of social accountability by which the central government upholds its responsibility for educational provision as it intervenes on behalf of impoverished communities.

Assessment may in fact reify centralized control by imposing standards that must be uniformly enforced throughout the country. Paradoxically, in Uruguay's highly concentrated model of governance, the national evaluation proves that centralization need not be incompatible with democratic participation.

The process of education reform in Uruguay

The Uruguayan educational system

The educational system of the Republic of Uruguay is organized in three levels. (Note 2) Initial education caters to children between 3 and 5 years of age. Preschool instruction is not compulsory presently, but the government plans to make it obligatory for 4 and 5 year-old children in the proximate future. Primary education consists of six grades and services 6 to 11 year-old children. Secondary education consists of two sub-cycles. The *Ciclo Básico Único* (Unique Basic Cycle) is a three-year course common to all students between 12 and 14 years of age. Students may then opt to proceed for baccalaureate or technical-professional instruction to round off their secondary education. Training at this level may last between 2 and 7 years depending on the course. Primary schooling and the Unique Basic Cycle constitute the national compulsory educational requirements (Uruguay—Ministerio de Educación y Cultura, 1996).

The administration of the education sector is highly centralized, but falls under the jurisdiction of several independent de-concentrated councils. The Ministry of Education and Culture is responsible for devising broad national educational policies. Despite its overarching mandate, this Ministry has a subsidiary role in the operations of the education sector. The *Administración Nacional de Educación Pública* (ANEP), the National Administration of Public Education, is the agency responsible for the management of the public educational system. The ANEP is fully autonomous from the Ministry of Education and Culture and it is configured by several bodies: (a) the Central Board Council (CODICEN), (b) the Council of Primary Education (CEP), (c) the Council of Secondary Education (CES), and (d) the Council of Technical-Professional Education. The Central Board Council is the highest administrative authority in the education sector. It is comprised of 5 members elected by the President and approved by the Senate. The other three councils are subordinate to the CODICEN, but they function largely autonomously. They are responsible for imparting, administering and supervising educational services. The directors of these councils are appointed by the CODICEN (see Figure 1).

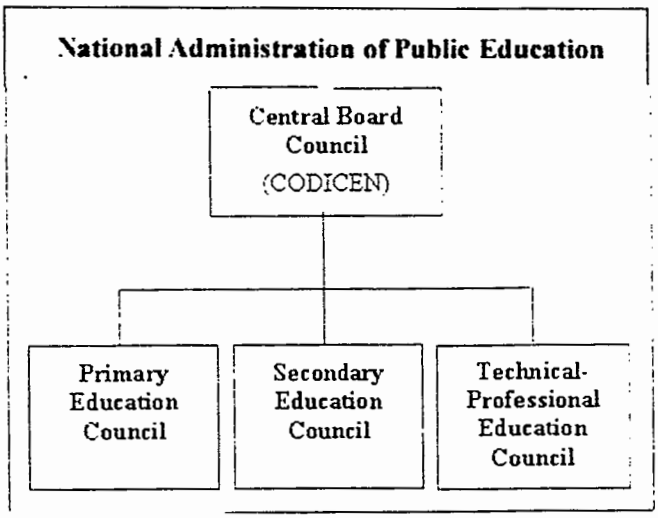


Figure 1. Organizational Structure of the ANEP

Educational policy is also shaped by several independent official advisory bodies to the ANEP. The Coordinating Commission of Education consists of the Minister of Education and Culture, the highest authorities of the autonomous councils as well as by representatives of universities and post-graduate institutions. It propounds guidelines and draft agreements for the coordination of the education sector. The *Asambleas Técnico-Docentes* (Technical-Pedagogical Assemblies or ATDs) are national and regional deliberative bodies comprised of teachers elected through secret compulsory voting. ATDs pronounce opinions regarding the conditions of education and may initiate educational policy directives (González Rissotto, 1997).

Basic education has reached universal proportions in Uruguay. In 1995, net enrollment rates at the primary school level encompassed 95% of the 6 to 11 year-old cohort. At the Unique Basic Cycle level, matriculation rates averaged 67% for the relevant school-aged population in Montevideo and 57% for all other urban areas in the rest of the country. Participation rates drop sharply in the second cycle of high-school instruction. Net enrollments at this level were below 30%. Total expenditures in education amounted to US$ 578 million in 1995, which represents 3.4% of the gross national product. The private sector caters to 13% of primary school students and 14% of secondary school enrollments (Uruguay—Ministerio de Educación y Cultura, 1997).

Uruguay has a shortage of teachers. The imbalance between teacher supply and demand has prompted governmental authorities to allow instructors to work double shifts. Teachers' real income has deteriorated steadily, even declining during periods of private real income recovery. Between 1960 and 1989, real salaries for teachers declined by 46.6%. Monthly wages in 1996 ranged between US$ 270 and US$ 407 (Uruguay—Ministerio de Educación y Cultura, 1996). Low salaries have forced teachers to search for alternative sources of income.

The Uruguayan education reform

A concern for the inequities in the Uruguayan educational system has prompted the government to embark on an ambitious reform initiative. Net enrollment rates for the population in chronic poverty reach 27% for preschoolers and 34% for high school students. The dropout rates for the poorest children in the first cycle of obligatory secondary education surpass 37%. There is also growing weariness about the deterioration of the quality of education. The national assessment of student achievement revealed that 6th graders in extreme poverty responded correctly to 37% and 17% of a language and mathematics test on average. The national means are nearly 20 percentage points above these levels. Primary school repetition rates have remained stable at around 10% during the past fifteen years. The repetition rate in the first grade, however, has reached 22%. In Montevideo, 63 out of 257 schools have a repetition rate in the first grade above 30%, and another 67 establishments

between 20% and 29% (Rama, 1998).

The current administration of the ANEP has adopted four guiding principles to transform the educational system (Rama, 1998; Uruguay—Ministerio de Educación y Cultura, 1996):

1. The consolidation of social equity,
2. The appreciation of teacher professionalism and training,
3. The improvement of educational quality, and
4. The strengthening of institutional management

The *consolidation of social equity* effort directs services and compensatory actions to underprivileged children. The ANEP seeks to extend public preschool services to 95% of the 5 year-old population and conduct an outreach program to incorporate 85% of 12 to 14 year-olds to the first cycle of secondary schooling. The poorest students receive more hours of instruction, including "full-time" schooling. They also have access to a comprehensive school meal program.

The *appreciation of teacher professionalism* effort strives to double the graduation rates of primary school teachers and triple that of secondary school instructors by the end of 1999. Approximately 90% of the elementary school teacher corps and 4,300 non-certified high school instructors will receive in-service professional development training. Teacher salaries were planned to undergo an increase of 13% in 1996, 10% in 1997, 15% in 1998 and 18% in 1999. In actuality, teacher salaries did rise over the yearly inflation rates, but did not reach the goals originally contemplated. Nonetheless, education was the only social sector that received an appropriation to increase salaries and its general operating budget in August 1998.

The *educational quality enhancement* effort focused around the widespread distribution of textbooks, instructional materials and pedagogical resources to public establishments. Curricular programs at the secondary level are also undergoing an in-depth review and renovation. In addition, the ANEP finances school-based projects to address specific needs within educational communities. Finally, the government has launched a program, "All Children Can Learn," to reduce primary school repetition rates. This program consists of a series of integrated social activities that endeavor to facilitate the access and permanence of children in schools, to strengthen the coordination between preschool and primary education, to enhance teacher training and to use textbooks as "an instrument for open learning" (Rama, 1998).

The *strengthening of institutional management* effort encompasses specialized training for school principals as well as the creation of computerized systems to assist administrators in their functions. Rural schools with less than ten students are being consolidated in order to reduce wastage and promote a more efficient use of resources.

These four initiatives are funded by a 22% increase in the education sector appropriation. The 1996-2000 budget has grown by US$ 75 million from the 1991-1995 budget, to US$ 430 million. The government of Uruguay also receives substantial aid from the international donor community to implement these reforms. The Inter American Development Bank and the World Bank have lent $140 million dollars to the modernization of the educational system. The Project for the Improvement of the Quality of Primary Education (MECAEP), (Note 3) funded by the World Bank, has contributed to the construction of preschools, the in-service training of elementary school teachers, and the provision of textbooks and pedagogical resources. It also supports the *Unidad de Medición de Resultados Educativos* (UMRE), the agency responsible for assessing educational quality at the primary level. The Project for the Improvement of the Quality of Basic Education and the Instruction and Training of Teachers (MESyFOD), funded by the Inter American Development Bank, has supported the creation of five regional teacher training centers, the in- service development of high school instructors, and the maintenance of secondary school infrastructure. In addition, MESyFOD has conducted the national assessment of student achievement at the secondary level in 1999.

2. Student assessment practices in Uruguay (Note 4)

A. The measurement of student achievement

Initial experiences with student assessment

Between 1990 and 1994, the United Nation's Economic Commission for Latin America and the Caribbean (CEPAL) conducted a series of studies requisitioned by the National Administration of Public Education. These studies were based on two examinations administered in 1990 to a small sample of 4th and 9th grade students in language and mathematics. CEPAL also collected socioeconomic and background information from parents, teachers and principals. The purpose of these tests was to explore the conditions of basic and secondary education in Uruguay (Comisión Económica para América Latina y el Caribe, 1994; 1993; 1992; 1991; 1990).

The primary school evaluation revealed that on average students could respond correctly to 58% of the questions (Comisión Económica para América Latina y el Caribe, 1991). The results from the secondary school evaluation were significantly inferior. Less than 22% of public school students reached an adequate level of proficiency in mathematics or language, as opposed to over 50% in the private sector. The mathematics test showed that "students learn very little in the courses of the Unique Basic Cycle." The language scores exposed that "the probability of success of the great majority of public establishments is so low that failure is almost certain" (Comisión Económica para América Latina y el Caribe, 1992: 90, 122).

The reports produced by CEPAL, however, abstained from making curt accounts or generic descriptions of student outcomes. Rather, test scores were the starting point for in-depth analyses of the impact of socioeconomic variables on student learning. Predictably, CEPAL found that low- income children tend to have lower levels of academic attainment. After an exhaustive review of the effect of various sociocultural indicators on school performance, the CEPAL underscored that maternal educational level is the best predictor of student achievement (Ravela, 1997b).

The research agenda of this study also included the identification of schools that, despite serving disadvantaged populations, have attained high levels of academic performance. These educational establishments were denominated "exemplary schools." The CEPAL carried out a qualitative investigation of these schools and posited that there were four factors that explain scholastic excellence in underprivileged environments:

1. the ability of the principal to assume a leadership role in the school as well as in its community,
2. the knowledge and experience of the classroom teacher combined with the satisfaction and commitment to his/her work,
3. a dynamic pedagogical culture within the teacher cadre, and
4. the existence of significant bonds between the educational establishment and parents (Ravela, 1997a).

Finally, the CEPAL emphasized that low test scores were symptomatic of a systemic crisis in the education sector.

> The reason for the results is not the fault of educational establishments or their authorities.... They are the outcome of a prolonged social process, during a prolonged historical period, during which the quality of education ceased to be a priority as an objective of State action (Comisión Económica para América Latina y el Caribe, 1992: 123).

In other words, the deterioration of educational quality was ascribed to a lack of commitment from the central State to make adequate investments in schooling services. According to this report, the reversal of this situation would follow from the initiative of the national government towards promoting policies and programs that support the labor of teachers and principals.

The construction of a national assessment system

It could be said that Uruguay does not have an institutionalized national assessment system. UMRE, the unit responsible for the measurement of academic

achievement at the primary education level, is not a formal "line- agency" of the National Administration of Public Education. It is an *ad hoc* unit initially constituted to implement the evaluation sub-component of the MECAEP Project financed by the World Bank. UMRE must abide by the directives of the Central Board Council, but it is exonerated from following certain civil service regulations. Similarly, the secondary education evaluation was developed autonomously within the framework of the MESyFOD Project, funded by the Inter American Development Bank. Although there are plans to make student assessment a permanent entity within the governmental organizational structure, the appraisal of academic performance currently operates from quasi-independent transitory agencies. This situation has provided to the evaluation of student achievement certain degree of independence and freedom—in relation to its organization, operation and personnel selection—by means of its ability to proceed outside the strict channels that regulate public offices. On the other hand, and as it will be described in a later section, this "extra-official" character has generated concern among certain sectors of the educational community, and particularly among the school inspectorate, who perceive UMRE as a parallel entity, alien to them.

The systematic and periodic measurement of schooling outcomes was not an initiative of the Uruguayan government. It was a conditional clause for the appropriation of the MECAEP World Bank loan (Interview UGN1). Although initially greeted with some resistance, the Uruguayan government eventually welcomed the creation of an evaluation unit (Interview UGN34). Germán Rama, who became Director of the ANEP in 1995, had been responsible for the design and implementation of the CEPAL study on student achievement aforementioned. Under his leadership, the Central Board Council decreed a resolution in March 1996 stipulating that "one of the prioritized lines of action of this Council is the implementation of assessment systems of [student] learning ... with the objective to appraise the performance of this Organism and the quality of service it provides to the population" (Uruguay—Administración Nacional de Educación Pública, 1996b).

UMRE has been in operation since 1994. Pilot tests for a 3rd and 6th grade evaluation were conducted late that year, with the intention to launch the first national assessment in 1995. When Dr. Rama assumed control of the ANEP in mid 1995, however, he replaced the technical leadership of UMRE and resolved to postpone the exam for one year. The national assessment underwent an important reformation. First, the ANEP would evaluate all public and private school students in 6th and 9th grades, the terminal years of the primary and secondary educational levels, every three years. Second, the test would veer from appraising curricular contents to measuring skills and competencies (such as reading comprehension or problem resolution). Third, the evaluation would incorporate a detailed sociocultural survey to be completed by parents, teachers and principals. Fourth, UMRE would seek feedback about its mission and operations from the various stakeholders involved in the provision of schooling services. Fifth, governmental authorities committed to maintaining secrecy about individual school test results. The ANEP guaranteed that only aggregate data would be made public (UMRE, 1996e).

UMRE is constituted by 3 full-time and 5 part-time professionals. It is responsible for the design, implementation, analysis, and devolution of results of the primary education assessment. From practically its inception, public and private school authorities as well as policy makers, supervisors and teachers were consulted about the development of instruments, test administration practices, and the uses of assessment results. The government also held regular informative workshops and produced several publications to raise awareness about the objectives of collecting student data (UMRE, 1996e). UMRE devoted significant effort to securing support and building consensus for the national assessment across the gamut of educational actors. In 1996, an "Advisory Group" was consolidated to review the work of UMRE and promote cooperative participation. This committee is conformed by national and regional representatives from the Council of Primary Education, the supervisory cadre, teacher training institutions, the Technical-Pedagogical Assembly, the Association of Private Education Establishments, the Uruguayan Association of Catholic Education, and the Uruguayan Federation of Teachers (the national teachers' union).

UMRE administered the first standardized evaluation in mathematics and language to all 6th grade students in 1996. Rural schools with less than six pupils in the sixth grade classroom were exempted from participation. Absenteeism rates were below 3.5% of the total enrollment in the mathematics test and 6.2% in the language test. In addition, educators and parents were required to complete socioeconomic background surveys. The rate of parental response to this survey was 98.5% (UMRE, 1998c).

The exams consisted of multiple choice items and open-ended questions. Teachers and supervisors participated in the formulation of test items, but technical staff from UMRE ultimately devised the exam. (Note 5) Independent proctors monitored the administration of the assessment. Students were allotted one hour and thirty minutes to complete the test, but those who required additional time to finish were allowed to do so (UMRE, 1996f). UMRE was responsible for correcting the exams and analyzing the results.

Forty days after the application of the test and prior to the culmination of the academic year, schools received an individualized confidential report with aggregate school results item by item. The socioeconomic background surveys served as a basis to categorize schools into five categories, from very unfavorable to very favorable contexts. Student outcomes were compared to the national average, the departmental/regional average and that of schools that service students from similar socioeconomic conditions. Educational establishments also obtained two technical manuals to interpret results. In the following academic year, educators received a second confidential report with a sociocultural profile of their school, based on background questionnaire data. UMRE also produced methodological guides with pedagogical suggestions and recommendations to redress weaknesses identified in mathematics and language (UMRE, 1997b; 1997c; 1997d; 1997f; 1997g; 1996c; 1996d; 1996g).

UMRE tailored several reports for the supervisory cadre. School inspectors participated in workshops where they received a regional profile of local schools and a "socioacademic map" that classified educational establishments under their oversight in terms of achievement levels and socioeconomic context. These instruments would allow supervisors to identify exemplary schools that exhibited high test scores in spite of being resource poor. They were also meant for targeting compensatory interventions to low performing educational establishments.

UMRE results

The national assessment of 6th grade students showed that 57.1% of students were able to respond to more than 60% of the language test correctly. The success rate in mathematics was considerably lower. Only 34.6% of students were able to answer over 60% of the questions satisfactorily. The percentage of students that did not reach the 60% "adequacy level" in both tests was 37.9%.

The first official report of results for public dissemination highlighted the role of contextual variables in the acquisition of knowledge. Students were classified into four categories according to their sociocultural context. Sociocultural context was defined in terms of maternal educational level. Schools from "very favorable" contexts were characterized as those with over 50% of students whose mothers completed at least secondary education. Schools from "very unfavorable" contexts were characterized as those where less than one out of two mothers had received only a primary education, and at most one out of ten mothers had received a secondary education.

As the CEPAL studies had demonstrated earlier, students from underprivileged backgrounds scored significantly below students from more affluent families (see Tables 1 and 2). While over 85% of children from "very favorable" contexts answered correctly to at least 60% of the language test correctly, less than 40% of students from "very unfavorable" contexts attained the same level of achievement. In mathematics, the gap between high- and low-income children widened.

Table 1
Percentage of Students by Performance Level in Mathematics and School Sociocultural Context

	Very Favorable	Medium High	Medium Low	Very Unfavorable	Total
Highly Satisfactory (scores above 80%)	21.0%	8.4%	3.4%	2.0%	6.8%
Satisfactory (scores 60% to 80%)	45.6%	35.3%	23.2%	15.7%	27.8%
Unsatisfactory (scores 30% to 60%)	30.6%	49.7%	60.7%	64.4%	54.5%
Very unsatisfactory (scores below 30%)	2.8%	6.7%	12.7%	17.9%	10.9%
Total	100.0%	100.0%	100.0%	100.0%	100.0%

Source: UMRE (1996g), p. 10.

Table 2
Percentage of Students by Performance Level in Language and School Sociocultural Context

	Very Favorable	Medium High	Medium Low	Very Unfavorable	Total
Highly Satisfactory (scores above 80%)	41.9%	19.5%	9.8%	5.0%	15.8%
Satisfactory (scores 60% to 80%)	43.3%	48.1%	40.9%	32.8%	41.3%
Unsatisfactory (scores 30% to 60%)	14.0%	29.7%	43.2%	52.7%	37.7%
Very unsatisfactory (scores below 30%)	0.8%	2.8%	6.1%	9.5%	5.2%
Total	100.0%	100.0%	100.0%	100.0%	100.0%

Source: UMRE (1996g), p. 10.

UMRE produced a second report exploring the relationship between sociocultural factors and student achievement. This study categorized the Uruguayan educational system into five subsystems according to geographical and sociocultural variables. This study revealed that private schools in Montevideo generally attracted students with the highest maternal educational levels, followed by, in decreasing order of maternal educational level, private schools in the interior, public schools in Montevideo, public schools in the interior, and rural schools. School performance in these subsystems was closely correlated to sociocultural context, with the exception of rural schools that evinced academic achievement levels slightly greater than expected for their low sociocultural context (UMRE, 1997f). More importantly, this report provided proof that academic achievement levels were not directly tied to the public or private nature of schooling, but rather to the sociocultural composition of the student body. In other words, the average scores of public schools from very favorable contexts were similar to those of their private counterparts within this context. The outcomes of private schools that served underprivileged populations were also analogous to those of public schools that assisted students from very unfavorable contexts. (Note 6)

A third national report was released late in 1997 providing a meticulous institutional profile of educational establishments. This document was based on the background surveys provided by principals, teachers and parents. It depicted the attributes of building facilities, school materials, class size, years of experience of

principals, teacher training, pedagogical approaches favored, staff turnover, parental involvement, and student self-esteem (UMRE, 1997g). As in previous inquiries, the analysis gravitated around the relationship between sociocultural context and schooling conditions.

Overall, the Uruguay government emphasized consistently throughout its public reports the role played by contextual factors in student learning. Average student scores, as all comparisons between geographic regions or between the public and the private sectors, were presented in direct relation to the sociocultural level in which learning took place. School-level data was kept rigorously confidential.

Other assessment activities

In addition to the sixth grade assessment, the Uruguayan government has undertaken two other evaluation exercises. Firstly, the government conducted an experimental assessment to a stratified sample of 3rd grade classrooms late in 1998. This test was available to other educational establishments outside the controlled sample for self- administration on a voluntary basis. The Central Board Council, however, exhorted all educational establishments to take part of this initiative (UMRE, 1998a). The purpose of this evaluation was to appraise student competencies at mid- point of their primary schooling. It also pursued to signal teachers about the expected competencies pupils ought to master by the third grade and provide them with an early- warning system to reformulate programmatic contents and pedagogical strategies (UMRE, 1997a).

The exam consisted of open-ended questions that integrated concepts from a variety of disciplines (mathematics, language, social studies, natural sciences, moral education, art) without compartmentalizing them into different spheres of knowledge. In response to teachers' demands for greater participation in the formulation of the test, UMRE established working groups with educators selected by the supervisory cadre, the regional Technical-Pedagogical Assemblies, and the associations of private independent and private Catholic schools. These working groups identified curricular areas to be evaluated and collaborated in the development of test items.

An informational document providing detailed information about the proposed testing scheme and objectives was drafted and distributed to all teachers and school inspectors. UMRE later requested teachers to respond to an opinion survey regarding the assessment instrument and competencies to be evaluated. Ninety two percent of respondents declared that the test was "adequate" and there was complete agreement about the competencies selected (UMRE, 1998b). As in the 6th grade assessment, the measurement instrument included background surveys for parents, teachers and principals in order to obtain data regarding the conditions in which student learning took place.

Every educational establishment received a report with national aggregate averages by competencies (reading comprehension, resolution of problems, processing information). Test scores were also broken down by socioeconomic context (rural, very favorable, favorable, medium, unfavorable and very unfavorable). A supplementary report detailed average background information (maternal educational level, home overcrowding, books in the house, preschool training) tabulated by sociocultural context. Schools that did not participate in the controlled sample received as well a standardized correction manual so that they could tally their own in-house results and compare them to the official national average scores.

Secondly, the 6th grade cohort evaluated in 1996 was re-tested in 1999 as students completed their 9th grade. MESyFOD, the project responsible for the administration of the test, espoused a methodology similar to that implemented by UMRE. The evaluation team sought to conduct informational sessions with supervisors, private and public school instructors, the Technical-Pedagogical Assemblies (ATD) and the teachers' union to gain their support. MESyFOD also intended to establish an advisory group conformed by representatives from every sector of the educational system that would review its operations. At the time the data collection for this study was conducted, it was unclear whether MESyFOD would be able to build consensus for the evaluation, especially from the ATDs and the *Federación Nacional de Profesores de Enseñanza Secundaria*, the national

secondary school teachers' union. Secondary school teachers had adopted a more contentious stance towards the central government's reform initiatives than primary school educators. ATD representatives had refused in the past to collaborate in projects spearheaded by MESyFOD (Interviews UGN3, UGN3b). (Note 7)

The MESyFOD team, however, concedes that the national experience with UMRE had greatly eased their work nonetheless. In most instances, educational establishments offered little resistance. They had not questioned the government's rationale for conducting this initiative nor were they concerned about being penalized for poor performance.

> Our undertaking has been facilitated due to the fact that MECAEP has been very careful about the confidentiality of test results, about the prompt devolution of scores, about the provision of individualized reports to each educational center. They took a series of precautions that, for instance, have encouraged private schools to open their doors. ... The realities of secondary education are not the same as those of the primary level, and there's still all the prejudices about standardized evaluations, but we're going along (Interview UGN3).

The assessment involved approximately 40,000 students. It appraised achievement in language, mathematics, social studies and natural sciences. Tests were administered by independent proctors and corrected centrally by MESyFOD.

B. The uses of assessment data

The findings uncovered by the first national measurement of student achievement are aimed at three distinct audiences: (a) the central government, (b) the school inspectorate, and (c) teachers and principals. Parents are informed indirectly about the general conditions of schooling through the press. A few schools, mostly in the private sector, have taken the initiative to publicize their scores to the families they serve.

The central government

> The national evaluation of student learning has as its official mandate: to produce information about the extent to which primary school graduates have been able to develop the skills and fundamental understandings in Language and Mathematics that every Uruguayan child ought to have incorporated regardless of his social origin, economic condition, or local context (UMRE, 1996b: 1).

This mission statement underscores the diagnostic objectives of assessment. "To have this information available," claims the ANEP, "is crucial to recuperate the democratizing role of the national educational system." Equity considerations lie at the heart of the central government's involvement in the measurement of academic outcomes.

The ANEP has relied on data gathered by UMRE primarily to guide and inform compensatory policies. There are three autonomous agencies within the national government that are consumers of information generated by UMRE: (a) the Council of Primary Education (CEP), (b) the MECAEP project (which is administered independently from the CEP), and (c) the Planning Area of the ANEP, a unit that depends directly from the Central Board Council.

The MECAEP project has been the most active patron of assessment data. On one hand, MECAEP has played a key role in promoting reflection among educators regarding the results of the first national evaluation. Technical discussions about the meanings of UMRE's findings have become a standard feature of institutional planning or professional development workshops organized for school inspectors and principals (Uruguay—ANEP-MECAEP, 1997). On the other hand, test scores and UMRE's classification of schools according to sociocultural context guide many of the initiatives undertaken by MECAEP. For instance, MECAEP disburses US$ 3,000 government grants for school-based projects. The selection process takes into

account how these projects may address shortcomings identified by the UMRE evaluation. Moreover, priority is awarded to schools from "unfavorable" sociocultural environments (Uruguay—ANEP-CODICEN, 1998). Sociocultural context, as defined by UMRE, has also become a salient criterion for the allocation of resources. The official press release detailing the outcomes of the first evaluation to the general public, for example, announced that MECAEP earmarked US$ 1 million to the purchase of pedagogical materials, targeting specifically 400 schools from unfavorable contexts (Uruguay—Administración Nacional de Educación Pública, 1996a).

"The Council [of Primary Education] permanently solicits information from UMRE," states a senior government official. "We are interested in learning about the strengths and weaknesses in language and math achievement, as well as about the relationship between school and family variables" (Interview UGN6). In practice, although the CEP's school inspectorate has been an important end user of test data, the central CEP office has given at best limited application to the UMRE results. School test scores have been used as educational quality indicators for the program "All Children Can Learn." This initiative strives to reduce repetition rates below the 20% mark in 160 schools through a comprehensive set of activities that include teacher training, providing health care services, reaching out to parents, and supplying textbooks (Uruguay—ANEP-CODICEN, 1998). Achievement levels have not been a parameter for bringing schools into the program, but test outcomes are occasionally used to tailor specific remedial actions in some establishments. Outside this initiative, the Council of Primary Education does not rely on UMRE data for other purposes. This has been a source of disappointment for some UMRE officials (Interviews UGN1, UGN2).

Finally, the Planning Area of the ANEP has depended on UMRE's school socioeconomic data for several of its own activities as well. In 1998, it conducted a research project on variables associated with primary education repetition rates (Área de Planeamiento de ANEP, 1998). This study demonstrated a close relationship between sociocultural context and the likelihood that students will be held back in the first and second grades. In addition, school background information has been "a fundamental referent" in the identification of establishments that could benefit from recent government initiatives, such as in-school meals, school infrastructure maintenance, or classroom construction (Interview UGN7). It is expected that once the MECAEP project comes to its conclusion, UMRE will become part of the Planning Area of the ANEP. (Note 8)

UMRE's own policy initiatives

The Council of Primary Education maintains that UMRE's role "is bounded to describing what happens" and "providing statistical data," so that, in turn, this knowledge can serve "the relevant organisms to make pertinent decisions" (Interview UGN6). In practice, UMRE has been more than just an information-gathering agency. It has been intimately involved in the design and promotion of educational policies for schools from "very unfavorable" contexts.

UMRE, with support from regional Institutes for Teacher Training, developed a Saturday workshop series for 541 urban primary schools serving underprivileged communities (approximately 40% of all public establishments). Participation in this four-month seminar was voluntary, but in order to qualify, at least half of a school's professional staff must have agreed to participate. Teachers were remunerated for the time they dedicated to this venture with a monthly monetary bonus equivalent to 30% of the average teacher salary.

Furthermore, UMRE established a fund to finance propositions that could enhance educational quality. Teacher training institutes received $1,000 awards to foster "the accumulation of knowledge about [student] learning in unfavorable environments and the implementation of professional development activities in teacher training institutions around these themes" (UMRE, 1997c: 1). Low-income schools could solicit $1,000 grants for the implementation of intervention projects destined to improve achievement levels in that educational community. The resources made available, however, would only allow for 50 school awards altogether.

Lastly, UMRE, in collaboration with the Program for the Strengthening of the

Social Area (FAS) from the Office of Planning and Budget, conducted a qualitative research project in 12 schools from unfavorable sociocultural contexts. Eight of these establishments excelled in the first national evaluation. The purpose of this study was to uncover the attributes of those establishments that inspired high attainment levels in underprivileged children. In particular, the dimensions explored were: (a) institutional characteristics, (b) pedagogical focus, and (c) linkages to the family and surrounding community. This study has become the basis for a comprehensive pedagogical proposal for "full-time" schooling to be implemented in 10% of public educational establishments serving the poorest children in the nation (Uruguay—ANEP-MECAEP, 1997).

School supervisors

The school inspectorate is organized hierarchically from the national-central level to the departmental-regional level to the local-zonal level. Although theoretically organized in a decentralized fashion (Macedo, 1995), school supervisors abide closely by the mandates established centrally at the Technical Inspection unit of the Primary Education Council (World Bank, 1994).

The supervisory cadre has a long tradition of evaluative activities at the school level. Schools are required to self-design and self-administer initial, mid-year and final exams in mathematics and language at all grade levels in order to appraise academic attainment. Inspectors must report on student test scores and specify the percentage of students that can master specific competencies, such as oral expression, orthography, reading comprehension, production of a text, resolution of algorithms, or recognition of geometrical figures. (Note 9)

In addition, inspectors are instructed to conduct their own institutional assessments in order to look beyond academic achievement as "the only objective testimonial of the level and quality" of educational services (see, for example, Uruguay—ANEP-CEP-Inspección Técnica, 1991a; 1991b; 1991c). They collect data on a wide a variety of measures related to educational quality, including student attitudinal qualities (respect, self- confidence, tolerance), absenteeism rates, repetition rates, classroom pedagogical approaches, availability of didactic materials, in-service professional development opportunities, and extent of parental involvement (see, for instance, Inspección Departamental de Montevideo, 1998). Supervisors produce a comprehensive school profile on the basis of this information and elaborate in conjunction with school authorities a strategic plan to address the shortcomings identified in this process.

The national assessment conducted by UMRE summed itself to the battery of school diagnostic information available to the inspectorate. UMRE elaborated reports tailored for the supervisory cadre categorizing schools by sociocultural context and performance level (UMRE, 1998c). Supervisors also had access to the scores of the schools under their tutelage. UMRE developed a series of workshops to familiarize inspectors with the results of this standardized evaluation and suggest potential courses of action that they may take to enhance educational quality.

Overall, the inspectorate gives high marks to UMRE. They underscore that it "has been extremely useful" (Interview UGN16) and has spurred a transformation throughout the educational system at various levels.

> We discovered that, it is important to have these data at the national level. In second place, this information is not only useful for the [educational] system, but for schools themselves. There were certain guarantees respected of all operations conducted. [The assessment] is not assigning blame in the face of potential deficits or anything like it. It is simply an objective measure that goes beyond [curricular] contents. and looks at much broader processes. ... In general terms, everybody is conscious that this is something valuable (Interview UGN9).

The assessment was a starting point to begin to understand the weaknesses in schooling, particularly for low-income children. Furthermore, it paved the way for the adoption of specific remedial actions to address these shortcomings.

> This mass evaluation of [student] achievement has put on the table quite

> clearly what all teachers have been perceiving for many years: how little children in situations of social exclusion learn. The evaluation took into consideration the educational level of the mother, home crowding, or the number of children in the family. [This systematized information] gave us, at the educational system level, some tools to correct in part this situation of low student achievement by updating teachers ... and proposing useful strategies in the areas of psychology, language and mathematics. From this point of view, it served an important professional upgrading role throughout the nation. It allowed many teachers to connect [with their students], because many knew that things were going poorly but it wasn't clear the reason why. It was useful to find new pathways (Interview UGN10).

Supervisors praise the technical reports and pedagogical recommendations put forward by UMRE. They are described as "filled with proposals for action" and "based on solid theoretical foundations" (Interview UGN15). "For me, [UMRE] has been very advantageous because of the exchange of materials. Their contributions are very helpful ... Really, they have been a great technical support" (Interview UGN13).

The national assessment has also served as a model towards a new educational paradigm. Traditionally, educators have emphasized memorization drills of curricular contents. The UMRE test, instead, moved away from appraising curricular contents to assessing competencies. A supervisor suggests that the UMRE test "took place precisely at a time when other pedagogical changes were taking place, and UMRE was able to appropriate itself of all this ... and motivate a re-elaboration of [educational] processes" (Interview UGN23).

> *Inspector 1.* [UMRE] moved us. It put us into contact with [new] literature, with another modality of evaluation that in turn implied another modality of [curricular] planning (Interview UGN19).

> *Inspector 2.* The results obliged us to think about the way curricular proposals were being implemented in educational establishments and how children were learning. The failure of students ... suggested that perhaps it was necessary to reformulate the educational project (Interview UGN14).

The inspectorate has played a crucial role in bringing the lessons from the first national evaluation into the classroom. Across regions, the supervisory cadre was required to organize in commissions to reflect upon student outcomes and devise plans of action that responded directly to the needs identified. These sessions focused on "the role and mission of the inspector" as a catalyst for change (Interview UGN12).

> The departmental inspector asked [us] to conduct a study, an analysis of the results, and see what we, as a departmental inspectorate, could do. I was recently reviewing this, and we had accorded to work with institutional projects ... Every supervisor, following these general guidelines, could request for funds to implement an intervention project in reference to the [UMRE] test results (Interview UGN30).

Inspectors were encouraged to adapt the guidelines outlined in departmental commissions to the social realities of the establishments they oversaw. In certain localities, supervisors organized 2- to 3-month seminars "to support educators with the findings of new research, and a theoretical framework" that delved not only on how students learn but what is relevant learning (Interview UGN13). In most districts, the favored approach has been to intercede directly with school administrators. "We work on specific proposals with our principals, who in turn pour this effort into institutional projects developed together with their teachers" (Interview UGN21). Lower scoring schools have received preferential attention over higher achieving establishments.

There is a growing sense that UMRE has imbued the educational system with a reflexive and renovating spirit. Regardless of the actual transformations that may have occurred as product of the first national evaluation, supervisors concur that UMRE has been responsible for bringing to the fore a national dialogue on the effectiveness of educational services and practices.

> Personally, I perceive that there have been changes. Changes in the good sense. There has been an evolution, in theory and in practice. There is a theoretical discussion about [educational] issues, which gets translated into daily activities. ... I have never seen such quick change. I believe this is positive (Interview UGN20).

Despite this strong endorsement to the work and outcomes of UMRE, inspectors do express reserve towards the national evaluation system. First, they underscore that the measurement of student achievement is not a new activity in the Uruguayan educational landscape "We have always evaluated," attests one supervisor unequivocally (Interviews UGN22).

> *Inspector 1.* In terms of evaluation, I believe that teachers have been working a lot previously on this subject. And so have inspectors. Yes, I share with others that the [UMRE] materials we received have triggered reflection among educators, but I believe that we have been working continuously on evaluation (Interview UGN17).

> *Inspector 2.* I suggest that it is not new to evaluate. [The UMRE assessment] is not new nor is it the only kind of evaluation. Of course, this was an evaluation at the macro level and by an external agent to the school. But we have never stopped evaluating within schools because this is inherent to teachers' practices: evaluating, planning, and researching (Interview UGN15).

Second, the supervisory cadre is concerned about the lack of coordination between the central Technical Inspection and the national assessment. Although all levels of the inspectorate (national, departmental and zonal) are represented in UMRE's advisory council, some supervisors protest that there has not been sufficient participation or communication between the two agencies.

> There is a need to polish certain instances [of participation] so that they are truly effective. Sometimes it is not enough to say that we are participating, that we want to participate. It is necessary that these spaces be created. The possibility is not always present. ... The will has been there, but the spaces are not instrumented so that we can actually share our opinions (Interview UGN12).

Ultimately, the inspectorate is wary of the overlap between UMRE's and their own functions. Supervisors stress that the national evaluation does not supersede their role in the education sector. "I believe [the UMRE evaluation] was a new thing for the educational system, but under no circumstance it precludes the other kind of evaluation that we have been conducting. They are complementary" (Interview UGN15). Some suggest that UMRE is an external agency that has unfairly arrogated their jurisdiction.

> The fact is that UMRE belongs to an organism that is called MECAEP and that is parallel to the normative system. It is alien to the Primary Education Council and to the [educational] system. Even though one may value some of the actions that they perform, we can't stop feeling this way. It is not an evaluation generated within the Primary Education Council. It comes out of an external organism. I believe this is one of the issues that produces great aggravation (Interview UGN10).

Others remark that UMRE has been unabashedly displacing them with an agenda of which they claimed to have no knowledge. "[UMRE] was coming above us. Sometimes we didn't even know what they were doing" (Interview UGN16). And yet others claim that UMRE oversteps the separation of responsibilities between the autonomous councils of the National Administration of Public Education (Interview UGN12).

> *Inspector 1.* Over the entire evaluative history in our country the ones that always performed a pedagogical review, a study, were the supervisory cadre and the Primary Education Council. Presently, that review is being done externally. We now wonder repeatedly, as inspectors, to what extent it is valid that somebody else comes along, with other possibilities, with other mechanisms, with more people, to do what we are doing. The measurement performed by UMRE is parallel to the functions of this deconcentrated authority (Interview UGN13).

> *Inspector 2.* The issue is that [UMRE and the inspectorate] each have their own lines of action. The inspectorate has a very clear agenda. But these lines of action get intercepted. Supposedly, UMRE ought to be an advisory or collaborative board in support of our activities. But if their actions are intercepting ours, or we are being displaced by UMRE, then that is where things are starting to become unwound (Interview UGN16).

In summary, supervisors object to the fact that UMRE is an external agency to the inspectorate with comparable functions. They resent that UMRE has had the ability to act independently, the authority to command the attention of educational establishments and the resources to implement directly remedial activities. To some extent, UMRE has come to embody a potential threat to the supervisory cadre. In a few schools, teachers even give credence to the rumor that the supervisory cadre will disappear or that it will be restructured. These criticisms not withstanding, the general consensus is that UMRE has been a positive asset and ought to continue the work that it has begun. "A system that does not evaluate itself cannot improve," remarks an inspector (Interview UGN17). According to the supervisory cadre, it is its organizational structure and relationship to the Primary Education Council that, in their eyes, begs to be redefined.

Principals and teachers

In hindsight, teachers and principals believe that the first national evaluation was an important experience. Private and public schools, as well as low- and high-income establishments concur that the UMRE assessment "was very useful, because it helped us to see where were our flaws, what we can do about them, and how we can change" (Interview UES2).

In its inception, teachers were suspicious of the UMRE test (Interview UGN9). Some expressed concern about whether student performance would be a means to appraise their own professional performance. Others feared that if their students did not attain high marks, they might be transferred to another grade (UMRE, 1996a). The Association of Teachers of Montevideo (ADEMU) expressed its rejection and opposition to UMRE. ADEMU protested that this was a test devised by an entity external to the Primary Education Council and supported by international donor agencies. "The economic expenditure that [the evaluation] supposes," the teachers' union announced in a newspaper communiqué, "does not conform to the austerity criteria that govern the education budget" (*El País*, 1995). The Uruguayan Federation of Teachers (FUM) also declared deep reservations towards the national assessment.

> In the second semester of 1996 and just prior to the measurement, the teachers' union picked up the debate [on the UMRE evaluation]. We reiterated certain existing reparations, about its expense and the degree of dependency to the World Bank's orientations. New elements of concern were also incorporated, like ... the possibility of using the

> results to categorize schools, to provide differentiated salaries to teachers according to test scores, to stigmatize a certain group of teachers or schools. Also, that it may favor the private sector in some way or other to the extent that it was predictable that public schools would have worse results than private ones. Another series of criticisms were directed to the pertinence of the instruments and the appropriateness of administering one instrument to measure processes in different social realities. Finally, there were concerns about the operational organization in itself, who was going to apply the tests, the access teachers would have, which guarantees existed about the formulation of the tests, the trustworthiness of correction criteria. The criticisms varied from highly ideological considerations, to reserve and distrust, to concerns about the everyday operations of the classroom. There was a wide scope of opinions (Interview UGN37).

Over time, these misgivings were assuaged. Although ADEMU remained defiant to the first national evaluation and encouraged educational establishments to forestall the entrance to exam proctors, teachers and principals collaborated with this governmental initiative. The Uruguayan Federation of Teachers recognizes that UMRE's "open attitude and desire to consult with the teachers' unions and technical-pedagogical assemblies" led to their participation in the Advisory Group and cooperation with the national test (Interview UGN37). An instructor from a rural area recounts that "at the beginning, teachers were not invested [in the evaluation], but during the past year, people started to talk positively about it" (Interview UES13). A representative from the Association of Private Education Establishments describes a similar experience:

> When UMRE appeared, we had a brick on each hand. I was ready to kill them. I had all my reasons against them ready. Little by little they convinced us. Now, after all that has happened and as we get more results, they convince us even more. It is OK that the test is obligatory. It has been a valuable experience (Interview UEM32).

The sense of trust and confidence garnered by UMRE among the teacher cadre can be attributed to four factors:

1. strict confidentiality of test results,
2. prompt devolution of student outcomes to school authorities,
3. contextualization of test scores by sociocultural background, and
4. abstention from holding teachers directly accountable for academic attainment.

> *Private school principal.* Teachers [initally] felt on the spot. There was talk ...that instructors who did not reach certain scores would be removed from office, that there was going to be a public ranking of schools, that this was an attempt to regulate teachers. The people from UMRE were quite clear in explaining what the objectives of the test were. But nobody believed them. Everybody feared that behind this there was something that somehow would harm teachers. ... It is now clear that they kept their word, that it was useful, that it helped us to review things, that two years later we are still working with the results (Interview UEM33).
>
> *Public school instructor.* Teachers feared that their school would be identified in some manner. And if the school was identified, so would their classroom. And from the classroom, the teacher [would be recognized] ... But the data were confidential. Only we got to know the scores. And the schools were later categorized according to their environment (Interview UEM16).

The national assessment has taken place within an education reform context that has espoused "teacher-friendly policies." "The appreciation of teacher professionalism and training" has been one of the four pillars of the reform (Rama,

1998). Real average teacher salaries have also risen progressively and consistently starting in 1993, after a period of decline between 1988 and 1992 (Domingo, 1998). (Note 10) This general setting might have contributed to generate a positive disposition among teachers towards the objectives of UMRE and a sense of trust that the evaluation had not been established to monitor their performance or increase their productivity.

Moreover, the attention of educators has not been focused on student achievement measures exclusively. The sociocultural data collected by UMRE were featured as an salient explanatory factor behind student performance. School background information has become a key justification to account for the level of academic achievement attained and an important consideration in the design of relevant remedial actions.

> The system of evaluation also conducted a family survey that took into consideration the role of the home in the educational process. We need to take into account that children only spend four hours at school, and twenty at home. The role of the family is fundamental in terms of the contributions that it can dispense to reaffirm educational processes (Interview UET21).

Teachers report that test scores were subject of repeated discussion and reflection sessions among school inspectors, principals, and the teacher cadre. The organization and participation in these initiatives was mandated by the central government.

> The following year, in 1997, when we came back to school, we were required to study the results of the UMRE evaluation, point by point, during our 'administrative days.' Then, we had to draw joint conclusions. It was an obligation to read them. [The order] came to the school in the form of an [official] act (Interview UES15).

The outcomes of the 6th grade assessment were the starting point of a process of pedagogical reflection for a wide range of public and private schools.

> *Medium-income school.* On the basis of the [exam results], we developed a plan for the following year. For instance, the discussion over problem resolution was very important for us in order to go deeper into this issue, to work more on reasoning. I don't know if this took us further away from the [official curricular] program, but ... Also, we've been working on the language [curriculum] in teacher meetings. ... In these sessions we analyzed some of the test items (Interview UET4).

> *Medium-low income school.* [UMRE] identified those competencies that experience the greatest problems. ... We studied the results and worked together with other teachers. We presented the findings in teacher meetings, and discussed the pros and the cons. We devised our [classroom] diagnostic tests at the beginning of the year on the basis of the test outcomes in order to give teachers the opportunity to continue working on these competencies (Interview UES7).

> *Low-income school.* The contents and approach [of the UMRE test] challenged a great deal of ideas that we had. When we saw the exam and what they were after, we came to realize that we were working wrong, that we were working differently, that we were behind, that we were traditional. ... The results and the design of the test (which was a very good proposition) led teachers to realize of everything that we lacked. From here on, we started to review everything, not because we did well, but because we could have done even much better (Interview UEM18).

> *Rural school.* [The UMRE test] does not evaluate for the sake of

evaluation, to just get some numbers back. It is meant to improve [our future practices], and to provide feedback. These are problem areas that require hard work and a different approach (Interview UET10).

Private school. I believe this was a very positive experience. It allows teachers to question if they are working well, along the lines they should be working, or if their approach is satisfactory (Interview UES21).

In some establishments, instructors aligned the course curricula according to the competencies measured by the UMRE test. Others describe that the evaluation triggered greater coordination between grade curricula. And in yet some others, they allude to the design of specific institutional projects to strengthen curricular objectives where students scored poorly.

Low-income school. I liked the approach of the [UMRE] test. It was an interesting proposition. The following year, we planned [our curriculum] on the basis of the approach forwarded by the test. We worked together with another sixth grade teacher on reasoning, geometry, numbers. We did all this basing ourselves on the UMRE test. Last year there wasn't an evaluation, but we administered the '96 exams at the end of the school year. We got a completely different result. In mathematics, it was very good. In language, it was low, but not as low as in the previous year. We even used the same methodology. You could not ask questions to a classmate or the teacher (Interview UEM3).

Rural school. I think that [the UMRE exam] was highly positive to shake teachers up a bit. It led us to question ourselves about many competencies and [curricular] areas that, perhaps, we were not developing well. Throughout the school cycle, students do not receive the same type of education. We might have missed a few steps. These concepts may not have been grasped at the right time and kids drag this handicap into the sixth grade. So the teacher covers the sixth grade curriculum, but oftentimes students do not have clear the concepts or processes necessary to sustain these new concepts. This is all very positive, so that we can all reflect. We are all responsible for specific areas. We have to make sure that students learn certain topics so that the teacher that follows can continue to build upon them (Interview UET11).

Low-income school: Principal. We observed that we needed to start all over again in language, particularly in reading comprehension. One of the factors that exerted incidence on this question was the lack of books at home. ... So we developed a project. *Instructor.* Yes, we developed a project that sought to overcome the current deficits. We called it "A Vegetable Garden to Learn." Through this project we are trying to address the problems detected in the evaluations. ... We find unsatisfactory or insufficient levels in competencies such as production of texts (... which comes to 52%), also algorithms (52%) and problem solving (48%). Those are the competencies with the lowest scores. Hence, we are trying to find solutions to those problems. At the same time, we see the need to continue working on discipline and the formation of good habits. The data from UMRE were particularly useful here. The study showed that we had a 47.6% of aggressiveness and misconduct, and lack of motivation or interest in a 29.9%. Through our little great project "A Vegetable Garden to Learn" we are trying to bring parents into the school and integrate them. Our school is from an unfavorable sociocultural context, and one of the problems that affects much of our functioning is that parents are not involved in student learning (Interviews UES9, UES10).

The impact of the UMRE test on academic practices can be appraised most overtly by how quickly it has become a standard for in-school evaluative practices.

Public and private schools that cater to children from diverse communities manifest that they have modeled their own student assessments after the UMRE test. Some establishments have photocopied the UMRE test and re- administered it. Others have prepared a different test with a comparable methodological approach.

> *Private school teacher.* Some teachers used the [UMRE] test again [the following year]. It was like a re- application. We've also used it as a model for other tests (Interview UES21).

> *Public school teacher.* I came to this school in 1997. I am the sixth grade teacher. Last year we administered something similar [to the UMRE test]. It was prepared here, together with other sixth grade teachers. We started to evaluate like UMRE. If the proposal is good, let's do it! I liked the narrative and argumentative text parts of the test in particular (Interview UES8).

Without entering into the discussion regarding the appropriateness or desirability of standardized evaluative practices in the classroom, it is apparent that UMRE's assessment experience reveals the influence nationwide examinations may exert in schooling practices, even in cases where these assessments do not involve *high stakes* testing. Uruguayan teachers adopted the evaluative approach proposed by UMRE despite of the fact that there were no incentive mechanisms or penalties openly associated with this test. It is also opportune to highlight that teachers did not experience this alignment an imposition of the central government or as a restriction to their pedagogical autonomy. They welcomed this methodology for finding it interesting or innovative.

Educators underscore that the type of evaluation proposed by UMRE epitomizes a novel pedagogical approach. Teachers find the emphasis on skill areas and problem solving particularly attractive. On the other hand, they recognize that they lack the know-how to implement it properly. That is, the methodological guidelines forwarded by UMRE are at best an initial referent; in order to be truly effective, they ought to be complemented with specific training.

> *Instructor 1.* The methodological guides say "this should not be like this," but they don't explain how we should do it. *Instructor 2.* There have been radical changes. We studied all our lives one way, under certain methodology. Suddenly, and especially in reading and writing, everything changes. *Instructor 1.* The explanations are very theoretical. Experts prepare these materials, but they remain up there, in theoretical issues. They are not very practical, or clear about how to apply them. *Instructor 2.* [They] first have to come to terms that we are not mathematicians or linguists (Interviews UES14, UES15).

The difficulties experienced in implementing change in classroom practices, according to teachers and principals, have centered around two broad predicaments: (a) lack of capacity, and (b) institutional organizational impediments. These obstacles afflict more acutely the public rather than the private sector, and low-income rather than high-income contexts.

Moreover, teachers lack the institutional space and time to master new techniques or ponder about educational practices. There are few opportunities for in-service training, team curricular planning and professional development. A notable exception was the seminar organized by UMRE for urban establishments from unfavorable sociocultural contexts.

> There are establishments that take into consideration the UMRE data, but there are also establishments that do not take advantage of [this information] not because they do not want to, but because they lack the institutional space for teachers to meet. There is no time for instructors to come together and reflect. It is all left in the hands of the good will of teachers to benefit from the results. ... This is an obstacle. There is an

> enormous quantity of information, but oftentimes it is wasted. It does not reach the teacher as it should (Interview UGN12).

Public establishments also undergo a frequent and dramatic staff turnover every few years. This has been a standard feature of the Uruguayan educational system. Educators are assigned permanent posts that they periodically vacate to fill in for temporary more desirable positions. This shift causes a ripple effect, encouraging another educator to leave her current post to fill in for that position now open. This permanent flux of school staff interrupts medium-term institutional processes as well as hinders educators from becoming intimately acquainted with local educational and social conditions.

> [In 1996,] I was the sixth grade teacher. It was my first year in the school. That year every teacher in the school was new. We had no knowledge of those kids. And neither did the principal, who had been assigned to the school the year before. It took us a year and a half to get to know the school integrally. The only original thing that remained in the school were the students (Interview UES18).

Educators have also professed some objections to UMRE's instruments and methodology. Classroom teachers, and especially those who work with low-income children, criticize the first national evaluation on three counts primarily. First, they object that UMRE depends upon the same instrument to evaluate disparate social realities. It is conceived as intrinsically "unfair" that children from underprivileged backgrounds must face the same exigencies as children that have access to plentiful resources. (Note 11) Second, they protest that exams were administered by outside proctors. The presence of an unknown person in the classroom allegedly distressed and distracted students.

> What I objected to was that the classroom teachers could not be the exam proctors. They did not trust us. The job of the proctors was only to distribute the tests, and we could have done that perfectly well. ... There was too much formality, and children are not used to it. ... And that had a negative impact. ... Children were neither at ease nor comfortable in that environment, and that was truly detrimental for them (Interview UET5).

Third, educators claim that unfamiliarity with a multiple choice methodology encouraged students to guess answers or select responses randomly.

In spite of these reservations, UMRE has managed to establish itself quite quickly within the Uruguayan educational landscape. This is a remarkable achievement provided that the evaluation system is barely a few years old. The words of a trade union leader capture this sentiment persuasively:

> I believe that at the [educational] system level, [UMRE] furnishes very valuable and interesting information. Although with some difficulties, it has been effectively incorporated into the school culture. The results are valued. The lack of discussion about the application of the third grade assessment immediately demonstrates that it has been incorporated into the school dynamic (Interview UGN37).

Teachers concur that participation in this experience has been beneficial. The UMRE test has, at worst, successfully fostered a dialogue about classroom practices and, in the best case scenarios, stimulated a renewal in pedagogic approaches.

> I speak sincerely. Sometimes, when teachers have many years of experience, we find that we must take on other activities outside school. The poor economic conditions oblige us to search for other activities so that we can live with dignity. Hence, suddenly we fossilize in certain aspects, certain methodologies. This test allowed us to see that we can evaluate in a different way. It has become a model. And it gave us

> bibliography so that we can continue along the path paved by UMRE (Interview UET21)

The first national evaluation has become a model on how to emphasize competencies rather than straight curricular contents. Many educators, in fact, argue that UMRE has taken the lead in educational matters, leaving the old official curricular designs to recede into the background and prompting teachers to challenge long-held assumptions.

> Our [curricular] program says Venn diagrams, it says operations, it says reasoning, it says application of knowledge, it says grammar, it says written expression, it says oral expression, it says reading. That is how our programs are currently structured. In the [UMRE] test, it said something else: mother tongue, reflection on language, text production. In the program, it says composition, it does not say written expression. Argumentative text is nowhere. In other words, the program is not what was evaluated. ... The program talks about sentence grammar, ... it talks about subject and predicate, but [UMRE] measured it as contextual grammar. ... We were convinced that we were teaching, but we had not realized that what we had in front was [expected of us too]. With UMRE, we came to realize that not everything that we did was right, that students were not quite responsible [for their shortcomings], that *we* needed to change behaviors (Interview UET7).

In summary, the assessment of educational quality in Uruguay went beyond a mere description of the conditions of schooling throughout the country. It was decidedly a call to action.

3. National assessment and the character of the Uruguayan nation-state

A. Assessment, rationality and State legitimacy

Assessment for rational decision-making.

The UMRE assessments have been designed as a recurrent diagnostic instrument of the characteristics of the Uruguayan education sector. "The evaluation of student learning ... is conceived as a systemic evaluation for feedback purposes" (UMRE, 1997a: 6). The main objective of UMRE is to supply educational actors—policy makers, school inspectors, principals and teachers—with relevant and updated information about student academic performance and the sociocultural variables that may condition it. This information will promote educational quality and equity through two channels. First, it identifies the strengths and shortcomings in education provision. Second, it sets the stage for school actors and government officials to take the necessary steps to correct deficiencies in the efficiency and distribution of educational services on the basis of systematically collected and objective data.

> What we endeavor is to produce information regarding ... which skills [students] have mastered and which ones they have not, what pedagogical and institutional strategies have succeeded to instill fundamental learnings in students from the neediest sectors and, finally, where it is still necessary to invest and provide technical assistance to attain a more democratic educational system that benefits all Uruguayan children without socioeconomic distinctions" (UMRE, 1996b: 1-2)

The national government has employed assessment outcomes to shape remediation policies and direct technical and economic resources to those segments of the population in greatest need. Student achievement measures and sociocultural context considerations have played a modest role in the allocation of didactic materials, technical assistance, and funds for school-based projects. The central State, however, has prioritized socioeconomic variables over strict performance

standards for redistributive purposes.

UMRE expects to bring about a renovation in pedagogical practices and classroom activities on the basis of the data it collects. Specifically, the assessment system propounds the following objectives:

> To make information available about [student] competency levels in areas considered to be fundamental; [and]
>
> To provide that information to teachers so that they can search for pedagogical alternatives that may revert situations prior to the exit of students from the primary educational system (UMRE, 1997a: 5).

Teachers and principals have been formally instructed to review the findings of the first national evaluation and devise compensatory strategies in response to them. The supervisory cadre has been closely involved in this process too, particularly in schools from unfavorable sociocultural contexts.

Assessment and State legitimacy

UMRE has consistently reported and analyzed student achievement outcomes in relation to socioeconomic measures. The first national report underscores the link between test results and background variables (UMRE, 1996g). The second national report is exclusively devoted to the impact of socioeconomic factors on academic performance (UMRE, 1997f). In other words, in Uruguay, the concepts of educational quality and equity are inextricably intertwined. The national evaluation system embodies another conduit for the central government to fulfill its obligation to reduce the gap between the privileged and underprivileged sectors of society.

> [I]t is considered that having information about fundamental skill levels is crucial to recuperate the democratizing role of education. The results obtained in the first national evaluation corroborate that strong inequalities in the quality of learning opportunities exist among students from social environments with great deficits. Although it is known that this is due to a multiplicity of factors, oftentimes external to the educational system, we assume our responsibility for the permanent improvement of the quality of learning. In socially disadvantaged sectors, the mediating function of the school becomes all the more necessary in order to contribute to the personal and social development of children (UMRE, 1997a: 6).

The contextualization of average test scores has become standard practice not just in official documentation, but in the collective mind of educators throughout the country as well. Educational establishments are keenly aware of their own location within the "socioacademic map" and have learned to interpret test results in relation to the social conditions in which the school is inserted.

> What is ultimately fundamental? To evaluate [student] linguistic and mathematical competencies, and to precise their family contexts. We have to see what the incidence of the [family] background is [on student achievement] (Interview UET14).

The identification of UMRE with the plight for educational equity has been instrumental for the legitimation of the State's evaluative activities. The collection of student achievement data has validated the reform initiatives of the national government by providing scientific proof of the erosion in the quality of educational services while furnishing rational-technical justifications for the pursuit of these compensatory measures. But perhaps more importantly, UMRE has bolstered the image of the central State as an interventionist agency supporting and tending for the neediest sectors of the population. As a teachers' union leader attests,

> [UMRE] ended up inspiring satisfaction. That is, it supplied schools with a depiction of their [academic] situation cross-referenced to sociocultural

> variables, repositioning results in terms of their contexts. This allows for a type of public stance that is congruous with the trade union's habitual position. Isn't it true? [It refers to] the degree of predetermination and conditioning faced by children as they enter the school. ... In short, there was a national test and there were results of that test that did not merit objections (Interview UGN37).

There are two additional factors that have ratified the validity of the assessment instrument and, *ergo*, the evaluation of educational quality as a legitimate State activity. First, the national evaluation apparatus has been construed as the fruit of a consensual process that has incorporated all of the actors in the educational system, including central government officials, regional and local school inspectors, teacher representatives from the Technical-Pedagogical Assemblies, trade union leaders, and private sector delegates. Second, the State has secured the support of educators by largely circumscribing their liability over test outcomes. "There is not going to be an index finger accusing anybody," declared Germán Rama, the National Director of the ANEP, upon the dissemination of test results (*El Observador*, 1996).

> Obviously, the deterioration [of the educational system] was not the unique or principal responsibility of teachers. **A multiplicity of factors external to the educational system has been in operation for this to occur:** the mass expansion of education, the deterioration of the quality of life of families, the retraction in educational investments during the military regime, etc. **However, it is necessary to recognize that there are variables internal to the system that affect the quality of student learning:** the pertinence of pedagogical strategies, the relevance of the curriculum, the modalities and expectations inherent in academic evaluations, the fact that schools from the poorest areas are the gateway to the teaching profession, among others (UMRE, 1996b: 1, bold in the original).

The circumscription of teacher liability was accomplished in two ways. First, by showcasing background variables as explanatory factors of academic attainment. "Student learning," UMRE (1997e: 2) attests, "is strongly stratified as a function of the sociocultural context within which each school operates." And secondly, by the central State acknowledging accountability over the conditions in schooling services. As established earlier, the national government accepts "*its* responsibility for the permanent improvement of the quality of learning" (UMRE, 1997a: 6, my italics).

The premise that the assessment of academic achievement legitimizes the central State potentially encompasses within itself a paradox. On one hand, evaluation endorses State action by making public its commitment and responsibility over educational processes and outcomes. On the other hand, the measurement of student learning implies a high risk: that poor test performance may provide irrevocable evidence of governmental inefficiency in educational service provision. Thus, if the central State is directly accountable for schooling processes and outcomes, doesn't evaluation jeopardize State legitimacy by calling attention to the deficiencies in schooling?

Sociological institutional theorists posit that assessment is primarily a *symbolic* activity (Meyer and Rowan, 1978). Its main objective, according to this paradigm, is not to produce results or provide relevant data for a diagnosis of the conditions of the education sector, but rather to *appear* that it does. That is, assessment strives to imbue the policy-making process with a guise of scientific rationality. The measurement of academic performance is foremost a legitimizing mechanism of State action by associating the policy-making process with scientific analysis.

Institutional sociologists underscore that attention to test scores may have a deleterious effect by uncovering inefficiencies within the educational system. Consequently, the relationship between assessment and legitimacy depends upon a *loose coupling* between evaluative processes and outcomes. In other words, assessment plays predominantly a figurative role, where the act of evaluating has greater salience than the findings it may uncover. This disjunction blurs the inconsistencies between educational goals and the existing conditions of schooling.

BEST COPY AVAILABLE

In summary, institutional sociologists profess that assessment systems prescribe officially acceptable standards of behavior and operation that uphold State action. On the other hand, these principles that educational establishments professedly embrace are in fact decoupled from the actual organization of schooling.

What do we observe in the Uruguayan case? The central government has reported aggregate test results from the UMRE evaluation at the national level. Student achievement data were not broken down by department or educational establishment. This practice differs significantly from evaluative experiences in other countries in the region that report testing outcomes by school or by region. Although withholding individual school data may indeed hide inconsistencies in educational service provision, it does not absolve the central government from liability over test outcomes. On the contrary, protecting individual school variability makes the central State the sole publicly accountable agent for educational quality. This strategy would appear to contradict the predictions of sociological institutional theorists. The Uruguayan government's approach to give ample dissemination to test results and advocate reflection over student outcomes, within a context where the central State has accepted responsibility for the quality of educational services, could give way to a crisis of legitimacy for the central government.

National test scores in the first national evaluation were, at best, substandard. Over 65% of students scored unsatisfactorily in mathematics and 43% performed poorly in language. (Note 12) Despite this inferior record, and contrary to common wisdom, UMRE did not delegitimate central State action. The central State, as predicted by sociological institutionalism, shifted the focus of attention from student outcomes to the role of sociocultural variables in academic achievement.

Assessment data fostered a national debate about the impact of socioeconomic forces in educational services. Evidence of the decay of the education sector was primarily a backdrop to champion governmental compensatory initiatives and vindicate the participation of the central State in social policies. The central government could afford to expose the deterioration in schooling because the root causes of the present educational landscape preceded the current administration. These had been already documented in detail in the student achievement studies conducted by the CEPAL in 1990 (Comisión Económica para América Latina y el Caribe, 1993; 1991; 1990).

Moreover, assessment data demonstrated that, controlling for sociocultural context, the performance of public sector schools is equivalent to their private sector counterparts.

> [W]hen we take into consideration the sociocultural context within which schools carry out their activities, results vary: public schools that operate in the most favorable contexts obtain results as good as private schools in the same contexts. At the other extreme, rural schools obtain results similar and sometimes even better than urban establishments from contexts equally unfavorable (UMRE, 1997f: 5).

Hence, UMRE asserted the value of public education and, consequently, of State-run educational service provision.

Then, if assessment does not jeopardize State legitimation, are evaluation practices in Uruguay an instance of a loosely coupled system as predicted by sociological institutionalism? That is, is the measurement of student achievement primarily a symbolic activity where evaluative processes are of greater consequence that their outcomes?

Interview and observational data collected for this research study suggest otherwise. In fact, school actors manifest that there is significant coincidence between State mandates around the UMRE evaluation and actual school behavior. In other words, there is evidence that central State action has successfully elicited organizational alignments. Teachers, principals and supervisors alike express a high level of familiarity with assessment policies. In most cases, they have largely complied with regulations to review and analyze test results. Furthermore, educators concur that this assessment has triggered reflection and some renovation in educational practices. As it has been documented in the section above, some schools have devised institutional projects in response to the findings of the evaluation.

Other educational establishments have modeled their classroom and evaluative activities after UMRE's appraisal, focusing for example on competencies rather than on curricular contents.

The degree of influence of the UMRE evaluation on educational practices stands out provided that this is a low-stakes test. There are no incentives tied to performance standards. Neither are educational establishments liable to the public or to the government for student scores. Similarly, a comparison between UMRE's appraisal and other school-based diagnostic evaluative exercises confirms that the former has had quite a distinct impact on classroom activities.

> *Low-income school instructor.* When we conducted our own evaluations, we tested concepts. The type of evaluation of UMRE, it makes you think and balance things out. It leads you to wonder what lies behind [a question]. It is evaluating the process itself. And it is providing feedback to our work. That is what we need to do ... We need to change (Interview UET15).

Two factors can account for this budding transformation in the classroom brought about by the national assessment. First, the evaluation was built and designed with the support and participation of the education community at- large. This process has fostered among educational actors a sense of appropriation and commitment to the work of UMRE. Second, UMRE accompanied its evaluative activities with in-service training workshops for teachers, principals and inspectors. Professional development has catalyzed the patronage and implementation of novel curricular and pedagogical propositions.

In summary, the Uruguayan central State is responsible and accountable for the conditions of the educational system. Assessment may potentially delegitimate State action by underscoring the weaknesses in the education sector. In spite of the shortcomings in schooling services exposed by UMRE, the national government did not suffer a crisis of legitimacy (Weiler, 1990). In fact, the central State was able to rally a wide basis of support behind this initiative. As sociological institutionalism predicts, the central State shifted the focus of public attention from testing outcomes to a comprehensive policy initiative addressing the socioeconomic wants that condition student learning. This displacement, however, did not necessarily decouple assessment from schooling practices. This is particularly striking given that the national evaluation was not designed a high stakes test for students, teachers or principals. The UMRE evaluation acted a conduit to channel the might of the State apparatus behind a pedagogical and curricular transformation.

B. Assessment and State ideology

Uruguay has a long-standing tradition of public support of social sector activities. It has the highest per capita spending on social sectors among Latin American countries. Social expenditures comprise approximately 50% of total government expenditures (World Bank, 1994). The State has been an ardent defender of public education and a champion of the conception of the *Estado docente*—the State as teacher (Fernández, 1997).

Uruguayan education reform program has leaned on two principles: (a) the pursuit and defense of basic social entitlements, and (b) the resolute participation of the State in the attainment of these entitlements through social promotion and redistributive policies. "The history of Uruguay shows that if you want to change qualitatively a social sector, it must originate from a strong State presence," remarks a high-ranking government official. "It is unimaginable to think of education reform without the State being an important protagonist" (Interview UGN7).

At an historical junction when the Keynesian Welfare State has been pronounced to be "in terminal decline" (Jessop, 1993: 34), the ANEP frames its vision for central State action in the education sector within this very paradigm. Renato Opertti, the National Coordinator for the Planning Area of the ANEP, portrays the current efforts to transform the educational system along this vein.

> The [education] Reform is rooted on a vindication of the Welfare State,

> in its objectives as well as in its contents; indirectly, it defies reform programs steered by the idea of an auditing and regulating State that "delegates" onto the market the direct provision of services (Opertti, 1997: 146).

Santiago González Cravino, another high-ranking government official, tempers this model of State action, while reaffirming the irrevocable duty of the national government to support the neediest sectors of the population.

> In order to attend to disadvantaged people, we need an Interventionist State. In order to favor and sustain the middle class, it is essential, sometimes, the intervention of the State. But the emphasis ought to lie in giving it a more positive and active role, using the private sector as a motivating instrument (González Cravino, 1995: 10).

The education reform program, an initiative born in the context of "budgetary limitations" and "commitments and conditions generated by international organizations," has been target of harsh criticisms from those that believe that the central State has relinquished its historic role.

> This "State" has had no incidence in overcoming the sociocultural deficiencies of increasing student cohorts. Neither have "compensatory" or "focalization" policies demonstrated any ability to surmount ... the true causes of pauperization, marginality and social exclusion (Pallares, 1998: 64).

President Sanguinetti, however, has staunchly defended his agenda for the transformation of the educational system as a "new form of humanist liberalism based precisely on the promotion of equity" (*El País*, 1997).

UMRE has evolved and operated within this framework of State-societal relations. Hence, the assessment system has sought to align its activities with a model of governmental action that promotes the production and distribution of social well-being.

> The Uruguayan education reform is statist in its defense of public education. The [UMRE] evaluation is very much linked to this. It is an attempt to promote social policies, to provide services. It is not symptomatic of a retracting [State] (Interview UGN3).

The national evaluation, as already documented earlier in this article, has stressed the utilization of student achievement information in support of remediation programs intended for disadvantaged communities. As test results have come to light, the central government has assumed responsibility for the conditions of schooling and voiced an institutional commitment to enact a policy agenda to address the shortcomings identified. In this sense, the national evaluation has proceeded in the spirit of *social accountability* and under the currency of social equalization.

The construction of UMRE in these terms has been a deliberate choice. The World Bank, who currently finances UMRE activities, had originally proposed an evaluation system based on a consumer accountability paradigm. The assessment would have operated under a different logic where parents, as consumers of educational services, rely on test results to select an educational establishment for their children (*Últimas Noticias*, 1996). When the administration of Germán Rama took charge of the ANEP in 1995, however, there was a change in strategy and UMRE was shaped after the assessment model that Rama had developed earlier at CEPAL (Comisión Económica para América Latina y el Caribe, 1991)

The association between the assessment system and the World Bank has inspired some mistrust regarding the credibility of the model of State-societal relations espoused by UMRE. In fact, this partnership has threatened to interfere with the legitimacy of the national appraisal. The Uruguayan Federation of Teachers and the Technical-Pedagogical assemblies have expressed opposition to the

evaluation of student achievement "because of its international perspective," associated with neoliberal policies that seek to reduce governmental intervention in the provision of social services (Interviews UGN6, UGN37).

> The [Central Board Council] has deteriorated the autonomy of the [Primary Education Council] by assigning functions to a parallel organization (MECAEP). [MECAEP] operates with resources conditioned by international loans and imposes EDUCATIONAL policies that do not respond to the needs forwarded by the NATIONAL TEACHER CADRE (Asamblea Nacional Técnico Docente, 1998: 30, caps in the original).

The national government, in turn, underscores its independence from the multilateral organization and reaffirms to the public opinion its defense of public intervention in the social sectors. "We are not dominated by [the World Bank]," asserts Claudio Williman, the vice president of the Central Board Council, "We are an underdeveloped country where State involvement is vital. ... Education is a competency of the State" (*El Diario*, 1996).

C. Assessment and State control

The Uruguayan educational system is structured in a greatly centralized and hierarchical fashion. All decisions—from administrative matters to curricular frameworks—are determined in Montevideo and uniformly enforced throughout the country. "Teachers in Uruguay behave like an army," remarks a government official. "If you give them an order, they will follow it" (Interview UGN3). There are extremely limited instances of organizational decentralization or institutional autonomy (Fernández, 1997).

World Bank report ascribes to this "extreme" concentration of power a profoundly deleterious effect.

> The highly centralized public primary education system hinders undertaking the required changes to achieve greater sectoral efficiency, equity, and quality. Centralization has restricted teachers and local managerial authority and initiative, reduced teacher-pupil interaction, discouraged personal growth and professional advancement, and limited the extent to which managerial staff and teacher opinions in pedagogical and administrative matters are solicited and recognized by those in charge of their workplace. On the other hand, centralization has overburdened policymakers and higher level staff with routine tasks and decisions, depriving them from having a more long-term strategic and prospective approach to the sector. The 19 [Departmental Inspectorates] are more concerned with transmitting centrally adopted policies and guidelines and collecting data on behalf of ANEP's central offices tan with enforcing activities to enhance the quality of education for which they are ill equipped and trained (World Bank, 1994: 11).

Uruguayan scholars concur that the educational system may benefit from greater flexibility and autonomy in its governance (Pallares, 1998; Fernández, 1997; Macedo, 1995). An initial step in this direction has been the disbursement of small grants to educational establishments for the implementation of school-based initiatives that can enhance educational quality (Uruguay—ANEP-CODICEN, 1998).

Undoubtedly, the national assessment supports the concentration of authority at the central level. As Hans Weiler (1993: 76) proposes, "evaluation is not merely the gathering and dissemination of information; it also has something to do with the authoritative interpretation of standards of knowledge and is endowed with a considerable amount of force, both real and symbolic." UMRE reinforces curricular mandates pronounced by the ANEP. It also fosters the alignment of school practices with centralized prescriptions. A government informant even claims that UMRE was an attempt from the central State to exert greater control over the flow of information on academic achievement after the release of the highly-critical CEPAL

studies (Interview UGN3).

On the other hand, the organization and implementation of the national evaluation defy this depiction of closed centralized control. UMRE has dedicated great effort to the incorporation of an ample array of voices and opinions into this process. It has steadily encouraged the systematic and continuous participation of all levels of civil society. The UMRE Advisory Group consists of representatives from the public and the private sectors, as well as central, departmental and local jurisdictions. Teachers, principals and supervisors have been repeatedly consulted on a wide variety of topics, from the design of the curricular matrix to be appraised to the development of test items.

UMRE's experience serves as a model of centralized governance sustained and enriched by democratic cooperation. The involvement of the Technical-Pedagogical assemblies and the teachers' union in the national assessment is living proof that even unpopular policies may garner the consent of reticent social actors in an environment that nurtures open and effective dialogue.

4. Concluding remarks

UMRE incarnates a model of social—as opposed to consumer—accountability where the central State must respond for the conditions of schooling. In this paradigm, the national government not only functions as a guarantor of educational quality and equity, but it also upholds its obligation as provider of educational services. The evaluation of student performance is an avenue to defend the role of public education as an equalizing social force and reaffirm the central government's support to the neediest sectors of the population.

However, student performance measures, as already expressed repeatedly, may potentially exert a destabilizing role by highlighting deficiencies in educational service provision. The central State averts the potential crisis of legitimation (Weiler, 1990) by shifting the character of assessment from the measurement of student outcomes to the remediation of the ills in student learning.

The conceptualization of education as a governmental responsibility has largely insulated the assessment process from finger-pointing or assigning blame. It is not teachers or schools that are being tested, but the educational system as a whole. This approach has generated the potential for educators to identify with and participate in evaluative activities. Democratic participation, in turn, buttresses the legitimacy of the assessment scheme.

UMRE has spurred the beginnings of a curricular and pedagogical transformation throughout the Uruguayan educational landscape. This is a promising first step for an evaluation system in its formative years. As new data are collected and UMRE consolidates its role within the education sector, the central State will face a new challenge: It will have to account for the effectiveness of its own policies in reducing existing inequalities.

Notes

1. I would like to thank the Unidad de Medición de Resultados Educativos for their support in this research project. I am indebted to Patricia Arregui, Martin Carnoy, John Meyer, Kathleen Morrison, Karen Mundy, Pedro Ravela and Michel Welmond for helpful comments and suggestions. All remaining errors are my own. This research project was supported by a summer fellowship from the Center for Latin American Studies at Stanford University and a Spencer Foundation Research Training grant. A Spanish version of this article has been published by the Working Group of Standards and Evaluation of GRADE-PREAL and can be accessed in Adobe Acrobat format at http://www.grade.org.pe/gtee-preal/docs/Benveniste.pdf
2. This analysis excludes tertiary education.
3. In 1998, the ANEP signed a loan agreement with the World Bank for an additional US$ 28 million for the second phase of the MECAEP project.
4. This section draws largely from personal interviews conducted with government officials involved in the design and implementation of the primary and secondary national assessment systems.

5. In the 1998 evaluation, teachers were invited to participate in the formulation of test items.
6. Correlation coefficients and standard errors were not provided by the source document.
7. A government informant explains the reasons behind secondary teachers' more contentious attitude in this manner:

> The secondary education teacher cadre is very different to primary school educators. The latter is a professionalized group. One hundred percent of [primary school] teachers obtain their degrees. They all went to normal institutes. They all have the title hanging somewhere at home. Hence, they have a positional culture that is more homogeneous. In secondary schooling, only 30% of the people teaching have specific preparation for being a 'professor.' There are university professionals, university students .... Thus, the heterogeneity is much greater. ... Secondly, secondary teachers have adopted a "let's see" attitude towards the education reform. Primary teachers were "calmer," more easy going, less opposition. That is why MECAEP, and more specifically UMRE, has been able to secure an active collaboration, inclusive of the teachers' unions and the ATD, the Technical-Pedagogical Assembly. In the case of secondary schools, the unions were more in opposition from the get go, more combative because the education reform was deeper. The ATD is also more politicized. The ATD leaders have emphasized their own ATD position over the stance of the [teachers'] union (Interview UGN3).

8. There is currently some uncertainty regarding the transfer of UMRE from the MECAEP project to the ANEP due to potential changes in the organizational and institutional structure of the evaluation system.
9. The characteristics of these evaluations vary from school to school. Thus, average student test scores are not comparable across schools.
10. Real average teacher salaries, however, are still slightly below their 1988 level nonetheless.
11. This stance contradicts another argument that points at the inherent inequity of holding different expectations for students from dissimilar sociocultural contexts, and particularly of holding lower expectations for children from lower-income backgrounds. The challenge would be not to "veil" the differences among social groups, but rather to introduce the necessary compensatory measures so that all students, regardless of their sociocultural context, can equally reach high achievement levels or national standards.
12. Unsatisfactory test performance was defined by UMRE as inferior to 60% of correct answers.

References

Áde Planeamiento de ANEP (1998). *Incidencia de los Factores Socioculturales en la Repetición Escolar*. Montevideo: Consejo Directivo Central, ANEP.

Asamblea Nacional Técnico Docente (1998). *Resoluciones*. Montevideo: mimeo.

Comisión Económica para América Latina y el Caribe (1990). *Enseñanza Primaria y Ciclo Básico de Educación Media en el Uruguay*. Montevideo: CEPAL.

Comisión Económica para América Latina y el Caribe (1991). *Que Aprenden y Quienes Aprenden en las Escuelas de Uruguay: Los Contextos Sociales e Institucionales de Éxitos y Fracasos*. Montevideo: CEPAL.

Comisión Económica para América Latina y el Caribe (1992). *¿Aprenden los Estudiantes? El Ciclo Básico de Educación Media*. Montevideo: CEPAL.

Comisión Económica para América Latina y el Caribe (1993). *Escuelas Productoras de Conocimientos en los Contextos Socioculturales Más Desfavorables*. Montevideo: CEPAL.

Comisión Económica para América Latina y el Caribe (1994). *Los Bachilleres Uruguayos. Quienes Son, Que Aprendieron y Que Opinan*. Montevideo: CEPAL.

Domingo, R. (1998). *Evolución Salario Docente Promedio*. Montevideo: Universidad de la República, mimeo.

El Diario (1996). *Estado No Puede Entregar la Educación a Manos Privadas*. Montevideo, November 6-7.

El Observador (1996). *Los Resultados No son Satisfactorios*. Montevideo, November 29.

El País (1995). *ADEMU Informa*. Montevideo, November 14.

El País (1997). *Sanguinetti Ratificó Apoyo para la Reforma Educativa*. Montevideo, July 30.

El País (1998). *Dinero para Universalizar la Enseñanza Preescolar*. Montevideo, October 6.

Fernández, T. (1997). "Educación post-vareliana: Evidencia para nuevas prioridades." *Revista Educación* 6: 63-78.

González Cravino, S. (1995). "La reforma de los servicios sociales del Estado: Una visión global." In Ministerio de Educación y Cultura, (Ed.) *La Gestión en la Educación, Política Educativa, Administración y Financiamiento*. Montevideo: Oficina de Planeamiento y Presupuesto, FAS, 6-10.

González Rissotto, R. (1997). *Sistema Educativo Nacional de Uruguay: 1993*. Montevideo: Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura.

Inspección Departamental de Montevideo (1998). *Evaluación*. Montevideo: ANEP-CEP- Inspección Departamental de Montevideo, mimeo.

Jessop, B. (1993). "Towards a Schumpeterian Workfare State? Preliminary remarks on Post-Fordist Political Economy." *Studies in Political Economy* 40(Spring): 7-39.

Macedo, B. (1995). "Descentralización de la Educación Básica: Las Experiencias de Chile y Uruguay." In U. C. d. U. D. A. Larrañaga, (Ed.) *Problemas y Desafíos en la Educación Básica*. Montevideo: Oficina de Planeamiento y Presupuesto, FAS, 27-29.

Meyer, J. and B. Rowan (1978). "The structure of educational organizations." In M. Meyer, (Ed.) *Environments and Organizations*. San Francisco: Jossey-Bass, 78-109.

Opertti, R. (1997). "La reforma educativa: Reinvindicación del Estado Benefactor." *Cuadernos del CLAEH* 22(78-79): 139-159.

Pallares, R. (1998). "Fundamentos y consideraciones para una política alternativa en la educación media." In F. V. Trías, (Ed.) *Reforma Educativa: Análisis Crítico y Propuestas*. Ediciones de la Banda Oriental.

Rama, G. (1998). *La Reforma Educativa en Uruguay*. Montevideo: Administración Nacional de Educación Pública, mimeo.

Ravela, P. (1997a). "Escuelas poductoras de conocimientos en los contextos socioculturales más desfavorables." In Organización de Estados Iberoamericanos, (Ed.) *Cumbre Iberoamericana: Memoria, Parte II. Aspectos Cualitativos y Cuantitativos en la Evaluación Educativa: Una Aproximación al Rendimiento*

Escolar. Buenos Aires: OEI, 99-125.

Ravela, P. (1997b). "La búsqueda de escuelas poductoras de conocimientos en el marco de la evaluación nacional de aprendizajes en Uruguay." In Organización de Estados Iberoamericanos, (Ed.) *Cumbre Iberoamericana: Memoria, Parte I. Aspectos Cualitativos y Cuantitativos en la Evaluación Educativa: Una Aproximación al Rendimiento Escolar*. Buenos Aires: OEI, 113-135.

Últimas Noticias (1996). *El Banco Mundial Sugiere que ANEP Reduzca sus Controles*. Montevideo, November 5.

UMRE (1996a). *Acta No. 8. Reunión Grupo de Consulta*. Montevideo: UMRE, mimeo.

UMRE (1996b). *Manual del Aplicador*. Montevideo: Administración Nacional de Educación Pública.

UMRE (1996c). *Material Informativo para Docentes: III. Manual de Interpretación Prueba de Lengua Materna*. Montevideo: Administración Nacional de Educación Pública.

UMRE (1996d). *Material Informativo para Docentes: III. Manual de Interpretación Prueba de Matemática*. Montevideo: Administración Nacional de Educación Pública.

UMRE (1996e). *Material Informativo para Maestros y Directores: I. Fundamentos*. Montevideo: Administración Nacional de Educación Pública.

UMRE (1996f). *Material Informativo para Maestros y Directores: II. Aspectos Organizativos*. Montevideo: Administración Nacional de Educación Pública.

UMRE (1996g). *Primer Informe de Difusión Pública de Resultados*. Montevideo: Administración Nacional de Educación Pública.

UMRE (1997a). *Fundamentos y Objetivos de la Evaluación Muestral de Aprendizajes en 3er Año Educación Primaria*. Montevideo: Administración Nacional de Educación Pública.

UMRE (1997b). *Lenguaje: Especificaciones y Sugerencias Didácticas*. Montevideo: Administración Nacional de Educación Pública.

UMRE (1997c). *Matemática: Especificaciones y Sugerencias Didácticas*. Montevideo: Administración Nacional de Educación Pública.

UMRE (1997d). *Primer Informe sobre Producción Escrita*. Montevideo: Administración Nacional de Educación Pública.

UMRE (1997e). *Programa de Mejoramiento de los Aprendizajes en las Escuelas Públicas Urbanas de Contextos Desfavorables*. Montevideo: Administración Nacional de Educación Pública.

UMRE (1997f). *Segundo Informe de Difusión Pública de Resultados*. Montevideo: Administración Nacional de Educación Pública.

UMRE (1997g). *Tercer Informe de Difusión Pública de Resultados*. Montevideo: Administración Nacional de Educación Pública.

UMRE (1998a). *Acta No. 2. Reunión Grupo de Consulta*. Montevideo: UMRE, mimeo.

UMRE (1998b). *Evaluación de Aprendizajes en Tercer Año de Educación Primaria*.

Boletín Informativo para Maestros. Montevideo: Administración Nacional de Educación Pública.

UMRE (1998c). *Uruguay--Unidad de Medición de Resultados Educativos*. Montevideo: mimeo.

Uruguay—Administración Nacional de Educación Pública (1996a). *Nota de Prensa. Primeros Resultados de la Evaluación Nacional de Aprendizajes en 6to Año de Enseñanza Primaria*. Montevideo: Consejo Directivo Central, mimeo.

Uruguay—Administración Nacional de Educación Pública (1996b). *Resolución, Marzo de 1996*. Montevideo: Consejo Directivo Central, mimeo.

Uruguay—ANEP-CEP-Inspección Técnica (1991a). *Circular No. 3*. Montevideo: ANEP-CEP- Inspección Técnica, mimeo.

Uruguay—ANEP-CEP-Inspección Técnica (1991b). *Circular No. 26*. Montevideo: ANEP-CEP- Inspección Técnica, mimeo.

Uruguay—ANEP-CEP-Inspección Técnica (1991c). *Circular No. 30*. Montevideo: ANEP-CEP- Inspección Técnica, mimeo.

Uruguay—ANEP-CODICEN (1998). *1998: La educación Uruguaya. Situación y Perspectivas*. Montevideo: Consejo Directivo Central, ANEP.

Uruguay—ANEP-MECAEP (1997). *Informe de Actividades No. 3 1997 1er Semestre*. Montevideo: ANEP- MECAEP.

Uruguay—Ministerio de Educación y Cultura (1996). *Desarrollo de la Educación: Informe Nacional de Uruguay*. Montevideo: Ministerio de Educación y Cultura, Dirección de Educación.

Uruguay—Ministerio de Educación y Cultura (1997). *Anuario Estadístico de Educación 1996*. Montevideo: Ministerio de Educación y Cultura, Dirección de Educación.

Weiler, H. (1990). "Decentralisation in educational governance: An exercise in contradiction?" In M. Granheim, M. Kogan and U. Lundgren, (Eds.) *Evaluation as Policymaking: Introducing Evaluation into a National Decentralised Educational System*. London: Jessica Kingsley Publishers, 42-65.

Weiler, H. (1993). "Control versus legitimation: The politics of ambivalence." In J. Hannaway and M. Carnoy, (Eds.) *Decentralization and School Improvement: Can We Fulfill the Promise?* San Francisco: Jossey-Bass Publishers, 55-83.

World Bank (1994). *Uruguay, Basic Education Quality Improvement Project*. Washington, DC: World Bank.

About the Author

Luis Benveniste
The World Bank, 1818 H Street, NW, Room MC8-414B
Washington, DC 20433

Tel.: 202.473.8495
Fax: 202.614.0935

Lbenveniste@Worldbank.Org
Email: luis_benveniste@post.harvard.edu

Luis Benveniste is an education specialist at the World Bank. His recent research work has focused on the politics of student achievement testing in Argentina, Chile and Uruguay. He has also co-authored articles on multigrade schooling and school accountability mechanisms in public and private schools. He has worked in a wide variety of international education projects in Latin America, Western Africa and East Asia.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Maria Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 597 times since July 17, 2000

Education Policy Analysis Archives

Volume 8 Number 33 July 17, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Economic Perspectives on Investments in Teacher Quality: Lessons Learned from Research on Productivity and Human Resource Development

Margaret L. Plecki
University of Washington

Abstract

This article reviews and critiques the ways in which researchers have used both productivity theory and human capital theory in efforts to measure the returns on investments in improving teacher quality. While studies utilizing these theories to measure investment returns provide useful insights, a critical need exists for research that advances our knowledge about the conceptual links between investments in teacher quality policies and improved student performance. The article also discusses several strategies for improving investigations regarding the returns on investments in improving teacher quality, including more refined measurement strategies, clearer conceptual frameworks, and a greater emphasis on resource re-allocation.

Investing in improving the quality of teachers and teaching is a central feature of many current education reform efforts at all levels of the policymaking system. Numerous calls for the improvement of teacher quality exist, and many states and local communities are targeting resources to ensure that all children have access to quality teachers. Many of the policy initiatives being considered require an increased level of investment in programs, training, and opportunities that support the ability of teachers to improve the level of student learning. Consequently, expectations are also increasing that the new investments will result in positive and enhanced outcomes for students.

Policymakers bear a responsibility for the equitable and productive management of resources as they address questions of how to best support the

improvement of the quality of teaching and learning. Difficult choices must be made regarding the distribution and use of a constrained set of resources targeted at improving teacher quality. Consequently, specific information about which improvement strategies hold promise can improve the understanding of the tensions and trade-offs that may exist under a particular set of educational conditions.

At the core of investments in the quality of teachers and teaching is some concept of teacher development. Either explicitly, or implicitly, policymakers presume that the resources they allocate purchase learning opportunities, offer incentives, and otherwise underwrite activities that—over time—develop the capabilities of teachers. These capabilities are further assumed to be the most immediate "cause" of student learning. Across the span of a teacher's career, these accumulating capabilities are likely to be associated with evidence of improved student performance.

This article reviews the contributions and the limitations of economic analyses of resource allocation policies aimed at improving teacher quality. Two analytic frameworks taken from the study of the economics of education are employed in this review: productivity theory and human capital theory. The article first summarizes results of various economic analyses of the productivity of resources, and discusses the strengths and limitations of this approach for informing questions about investments in teacher quality. Next, the aspects of human capital theory that are relevant to the issue of resource allocation for the development of teachers' capabilities and careers are presented. These aspects are considered in addressing two teacher policy arenas in which resource allocation is a critical feature: teacher compensation and teacher professional development. The article concludes with considerations for policymakers when faced with resource allocation decisions regarding policies aimed at improving teacher quality.

Inquiry about productivity

Let us first consider the premise that when policymakers decide how to best invest in strategies designed to support teacher development, they are faced with the issue of educational productivity—that is, what results (e.g., student achievement levels) are produced by investments in teacher development? Questions such as the following are key considerations in policy debates: What are the best approaches for getting the most for our educational dollar? How do we best support teachers in a climate of increased standards and expectations for student learning? How do we best reach the full spectrum of teachers and students in need of improvement? What do we know about existing efforts to improve teacher quality? The answers to these questions are complex and variable. The nature and the extent of the educational challenges differ in important ways at each level of the policymaking system (state, district, school, and classroom) and the specific conditions of students and teachers within each level of the system vary considerably. Each question emphasizes the need to better understand whether or not we are utilizing resources devoted to teacher development in the most efficient or equitable manner.

In order to wrestle with the notion of how productivity studies can inform teacher policy issues, we will briefly examine some of the existing research on productivity in education. A historical review of the literature indicates that there has been considerable debate in the research community about the manner in which increased spending on education may or may not be related to improved performance (Hanushek, 1989; Murnane, 1991; Hedges, Laine & Greenwald, 1994; Biddle, 1997: Ferguson & Ladd, 1996). However, this does not mean that inquiry regarding productivity does not have value. Instead, understanding the nature of the conceptual challenges involved in conducting such investigations of productivity may shed light on the strengths and weaknesses of any particular set of policy strategies. That is, facing the difficulties of specifying the exact nature of the costs and benefits to be derived from a set of policies can provide valuable insights that might be used in the process of selecting from competing demands for resources.

For the most part, studies of educational productivity have examined the relationship between the amount of money spent on various educational "inputs" and the levels of student achievement that are presumed to be associated with these inputs. These studies, typically referred to as education production function research,

derive much of their conceptual framework from the microeconomic theory of the firm (Benson, 1978). The production function model attempts to analyze the relationship between inputs and outputs. The goal of this inquiry is to investigate the changes in output (typically measured by student achievement test scores) associated with changes in the levels or mix of educational inputs (e.g., per-pupil expenditures, teacher characteristics, and teacher-student ratios, with some statistical controls for variations in student background and family characteristics). Production function research can also be viewed as an analytic frame in which cost/benefit analyses can be conducted.

Several significant conceptual and technical problems surface when attempting to apply a production function theory to educational productivity. Conceptually, the lack of agreement about the elements of a theoretically sound "theory of production" in education plagues the research in this area. In other words, unlike the microeconomic theory of the firm, the forces and conditions that comprise the human "equation" of student learning are neither obvious nor fully understood. The lack of agreement is understandable, given that education is characterized by interactive and developmental processes stretching across many years of schooling (Carroll, 1963; Mortimer et al., 1988). Given the lack of an agreed upon theory of educational production, it is little wonder that technical issues abound, such as the specification and measurement of proxies to best represent the important elements in the educational process. Hence, the choice of inputs and their metric specifications may rest on other than strong theoretical grounds. Production function researchers typically choose particular input or output measures because information is readily available, the variable has some policy relevance, or because the variable is intuitively plausible (Monk, 1990).

Conceptual and technical problems notwithstanding, researchers have repeatedly used production function theory and techniques to examine the way investments may have affected educational outcomes. While the results are mixed and in some dispute, they do offer insights into the relevance or impact of investments in teacher quality aimed at improving student learning.

A seminal article on the subject of educational productivity (Hanushek, 1981) claimed that after reviewing 130 studies of educational productivity, no consistent, positive, significant relationships could be uncovered between increased spending on education and improved student achievement. Subsequent reviews by the same author (Hanushek, 1986, 1989, 1991) yielded the same general result. These analyses have been central to a continuing policy debate about whether dollars matter in the quality or improvement of education. A re-examination of Hanushek's analysis of the literature, conducted by Hedges, Laine & Greenwald (1994), arrived at a different conclusion: when alternative procedures for aggregating the results of separate studies are used, certain input measures—among them, factors related to teacher quality—do have a significant relationship to student outcomes. These authors found that teacher education, ability, and experience, along with small schools and lower teacher-pupil ratios, are all positively associated with student achievement. The difference in results is due to the use of an alternative methodology for conducting the meta-analysis of the same literature. Others who have reviewed prior production function research (Ferguson & Ladd, 1996) claim that many of the earlier analyses did not critically sort out the methodologically weak studies from consideration, thus casting doubt on the validity of the conclusions being drawn.

Over the past two decades, there have been waves of productivity studies which have employed a more microanalytic approach using disaggregated data (Murnane, 1975; Summers & Wolfe, 1977; Thomas & Kemmerer, 1983; Brown & Saks, 1975; Rossmiller, 1986). These studies have focused on school and classroom levels, in contrast to the more typical studies or analyses which have used more global measures from macro- level databases. Findings from the microanalytic studies reveal a similar pattern of mixed results. However, several production function studies in this tradition have demonstrated positive relationships between teachers' ability levels (usually a measure of verbal aptitude) and student achievement (Ehrenberg & Brewer, 1995; Summers & Wolfe, 1977). Ferguson (1991) examined school districts in Texas and concluded that there are systematic relationships between educational inputs and student outcomes that he estimated to account for

between one quarter and one third of student achievement differences. Ferguson & Ladd (1996) examined Alabama schools and concluded that there is evidence that the input variables of teacher's test scores, the percentage of teachers with master's degrees, and small class size are positively associated with student test scores. The authors assert that the use of more methodologically sound analytic techniques (e.g., value- added specification) combined with a more disaggregated analysis can address some of the perplexing problems which have been associated with production function research. A recent multiple-method study by Darling-Hammond (2000), which examined relationships between teacher quality and student achievement, yielded somewhat different results from those of Ferguson & Ladd. Darling-Hammond examined state- level data from all 50 states and concluded that measures of teacher preparation and certification are correleated with student achievement measures. One of the study's specific findings was that state-level measures of the percent of fully certified teachers and a major in their academic field is a stronger positive correlate of student achievement than the percent of teachers with a master's degree.

Accompanying the ongoing search for empirical relationships between inputs and outputs are doubts about the utility of the production function literature. Some argue that even when significant relationships are found between input variables and student outcomes, these results do not have useful policy implications (Witte, 1990; Murnane, 1991). Others question the appropriateness of the specific variables being used and the limitations imposed by an almost exclusive focus on test scores as the measure of student outcomes (Smith, Scoll & Link, 1995). Furthermore, results from the production function research studies which do not uncover a significant relationship between increased spending and increased student outcomes collide with the widely-held, rather common-sense belief shared by many educators and policymakers that increasing the level of investment makes—or can make—an important difference. Some researchers assert that insufficient attention paid to how additional dollars have been spent on education inputs may explain the apparent lack of connections between dollars and outcomes. For example, in an analysis of school district spending in New York state (Lankford & Wyckoff, 1996) researchers found that a sizable portion of the increased resources were allocated to special education programs for the disabled. Given that student outcome measures for disabled students are often unavailable or excluded from aggregate data sets, it is likely that this aspect of increased spending is not accounted for in some of the production function research.

Alternatives to the input-output predictive model for assessing educational productivity, noted in the literature, may hold promise for capturing more precisely how resource investments targeted to the quality of teaching may translate into improvements in student learning. Barnett (1994) suggests that embedding production function and cost function studies in the theoretical model of private firms may not be appropriate for understanding how resources are allocated in public school systems. Alternatively, he suggests models which are derived from theories about the bureaucratic behavior of government institutions (Nikansen, 1971) may more appropriately explain how educational resource allocation decisions are made and what impact these resources have. In this alternative view, the unit cost of the school is determined by the available revenue, not by the most effective way to allocate revenue, and school administrators strive to maximize revenues and allocate resources to keep employees responsive and cooperative and maintain the school's reputation. Hughes, Moon & Barnett (1993) find that while resource allocation in schools is more closely linked to funding those factors presumed to be related to quality or general school goals (e.g., better equipment and facilities, newer texts, additional support staff), these factors may not be directly linked to improved educational outcomes. To discover more direct links between resources and outcomes, a line of inquiry in educational productivity research may be needed which elevates the importance of classroom-level analysis and complements the school-based studies (Monk, 1992; Rossmiller, 1986). Elmore (1994) offers the observation that traditional budgeting practices in schools and school districts are not centered on determining the actual costs of educational inputs, but rather focus on either adding or subtracting dollars from a baseline budget. He also notes that educators typically do not have any special training or background to assist them

with the complex problems embedded in budgeting and improving productivity. Odden & Clune (1995) discuss several factors related to low productivity, including a highly uneven distribution of resources across states, schools, districts, and students, unimaginative uses of dollars that do not translate into improved performance, and a focus on additional programs rather than results. The authors cite several areas where additional productivity research might be extended: research on increased course- taking at the secondary level, examination of organizational strategies which are associated with improved performance, and research on high-poverty schools.

The upshot of these lines of thinking and research to date is that we know less about the productive impact of policymakers' investments in teacher development than we might wish. To be sure, some analyses highlight certain teacher-related variables (teachers' verbal ability, education, and years of experience) that appear to bear some relationship to student learning. Other studies establish no clear or discernible relationships. The lack of connections and the mixed nature of results across studies may be due to the weaknesses in underlying theory or specification of measures. Or, these models have yet to represent adequately important variables intervening between the allocation of resources and their enactment in practice. By a similar argument, production function models take little account of the actual allocation and expenditure dynamics within public education bureaucracies, and hence we are unable to tell whether increased levels of resource investment overall were actually targeted to inputs of immediate relevance to improved classroom performance.

Inquiry into human resource development

The shortcomings of educational productivity research lead one to consider other lines of economic analysis that are built on a more explicit theory regarding the improvement of teachers' capacity for their work. Research on policies that seek to develop and reward the "human resource" of the teacher force is particularly relevant. Research regarding the effective, efficient, and equitable use of human resources is a critically important area to investigate when considering policy options that are intended to support improved teacher quality. The bulk of operating expenditures in education are allocated to pay for the cost of employing school personnel, with the largest portion of those expenditures allocated to classroom teachers. Arguably, the quality of education is ultimately dependent on the classroom teacher's ability to produce educational outcomes. Two specific policy strategies for supporting teacher development—teacher compensation and investments in ongoing teacher professional development—are conceptually linked to theories of human resource development. As a point of departure, we begin by outlining findings from research on human capital theory that are applicable to both teacher compensation and professional development, and have contemporary significance in examining investments made in these two teacher development policy strategies.

Human Capital Theory and the Development of Teachers

The examination of human resource development has been a central area of study in the economics of education. One of the long-standing theories of human resource development, human capital theory, views human beings as individuals who possess great potential which can only be fully realized by making investments in human development. As far back as 1776, the publication of Adam Smith's *Wealth of Nations* offered at least two insights into the nature of human capital that have applicability to the contemporary discussion of investing resources in teacher development. The first of these is the observation that labor inputs are not purely quantitative. Second, Smith observed that productivity is related to both "the quantity of the capital stock which is employed...and the particular way in which it is so employed." (Smith, 1776). These ideas suggest the importance of understanding both the "stock" and the "flow" of human resources (e.g., teacher's labor), as well as understanding the qualities of these resources. The evolution of human capital theory since Smith's time (Note 1) suggests that at least three elements are related to

the quality and productivity of human resources: the amount of human resources being employed, the quality of those human resources, and the way in which human resources interact in their employment.

These central ideas of human capital theory shed light on the thorny problem of measuring human resources and assessing their effects. The measurement of labor quality has been a subject of investigation by many who study the economics of education. Benson (1978) is one of several experts in the economics of education who has noted that we typically use proxies to judge the quality of labor inputs. He pointed out that education levels, degrees, and the acquisition of special credentials—the most common proxies for labor quality—are commonly used across all types of labor markets. Employers value education levels, degrees, and credentials because of a belief that these are acceptable proxies for valuable knowledge and skills that render the worker more productive in a particular type of labor market.

Proxies such as these have often been used to examine various policy strategies for improving teacher quality. We could reasonably assert that teachers who possesses higher levels of knowledge and skill in their craft will be associated with higher levels of productivity. While this assertion seems very obvious, the process of identifying teachers who possess higher levels of knowledge and skill is far from obvious. As is true in most professional labor markets, we search for reasonable proxies for the knowledge and skill of teachers. In particular, scholars—especially those engaged in productivity research—have traditionally focused on years of experience in teaching, degrees and credentials earned, and levels of education and/or training beyond certification, often known as continuing education credits. Each of these proxies is typically associated with some type of resource allocation policy.

By applying lessons learned from human capital theory, we can expect that these proxies are insufficient measures of teacher quality, and, consequently, investments aimed only at these proxies are likely to render variable results. The proxies focus too much attention on quantity, are only loosely connected to quality, and to a large extent, ignore the matter of the way in which the resource is configured in its employment. Thus, the conceptual basis for measuring the relation between the human resource inputs and the productivity of those inputs is quite weak.

The perspective provided by the application of human capital theory is useful when considering resource allocation strategies for improving teacher quality. For example, investments which produce higher levels of education, credentials, and/or training for teachers may result in increased productivity. However, the extent to which these investments pay off is dependent on the closeness of the conceptual link between the types of education and training purchased and the knowledge and skills needed *and used* in the classroom context. Keeping the perspective of human capital theory in mind, we now consider two types of investments in teacher quality: professional development and teacher compensation.

Investments in professional development

Research on investments in professional development has tended to address a different set of questions than productivity studies. Here, studies seek to answer two questions primarily: (1) who invests what in professional development? (2) what do these investments purchase? A more limited set of studies offer answers to a third question: how much and in what ways does professional development (and, by implication, investment in professional development) influence student learning? Virtually no studies address directly the question of the relation between investments of resources to support professional development and student learning measures.

Professional development for teachers has consisted of a myriad of activities and programs that are financed in a variety of ways from all levels of government. Several studies about the costs of staff development have been conducted (Moore & Hyde, 1981; Lytle, 1983; Stern, Gerritz & Little, 1989; Elmore, 1997; Education Commission of the States, 1997) but an analysis of the available research indicates that there is little generalizable information about the range of resources allocated for professional development (Orlich & Evans, 1990). Nonetheless, there are clear

modal patterns regarding what these resources buy. One study found that teachers are two to three times more likely to be participants in a district-provided staff development than enrolled in a college or university course (Little, 1989). The same study also calculated that more than four-fifths of state dollars for staff development were controlled by the local district. A study by the Education Commission of the States (1997) found that approximately three-fourths of school district resources designated for professional development are spent on teacher inservice days, conferences, and workshops.

Professional development activities have been dominated by a training- based delivery system, generally managed by school districts, which offers teachers a variety of workshops targeted on special projects or narrowly defined aspects of reform (Little, 1993). This type of packaged professional development is not well suited to current educational reform purposes and ignores the opportunities to learn that are part of the school organization (Hargreaves, 1990, 1993). Not surprisingly, scholars have increasingly noted the need to have professional development practices more crucially linked to the improvement of student performance (Darling-Hammond & McLaughlin, 1995).

The systemic reform initiatives during the past ten years have emphasized the importance of high standards for all students, a thinking-oriented curriculum, and performance-based student assessments linked to the standards (Resnick, 1993). Educational reform based on standards and performance-based assessment implies a focus on the development of new professional knowledge and skills which teachers will need to produce an elevated level of student outcomes. The particular set of required knowledge and skills would vary by the context and conditions of the individual school setting (Cohen, McLaughlin &Talbert, 1993). Efforts underway by the National Board for Professional Teaching Standards and the National Commission on Teaching and America's Future are two examples of the types of efforts underway to improve teacher recruitment, preparation, certification, continual development, and retention.

Some efforts have been made to calculate the costs of resources currently being devoted to the continuing education of teachers. Miller, Lord & Dorney's (1994) estimates range between 1.8% and 2.8% of the district's operating budget. The cost per regular classroom teacher ranged between $1,755 and $3,259. Their study was based on a series of intensive case studies in four districts located in different regions in the U.S., ranging in size from 9,500 to 125,000 students. The estimates are based on direct costs such as the salaries of district and school administrators, and substitute teachers as well as on the direct costs of materials and supplies. One detailed study of staff development in California (Little et al., 1987) estimated the investment in professional development to be almost two percent of total funding for education in that state. In a study of one New York school district, Elmore (1997) estimated that spending on professional development amounted to about three percent of the total budget. One long-standing observation has been that school districts with more than one percent of its budget allocated to professional development is an exception (Darling-Hammond, 1994; Houston & Freiberg, 1979). These studies do not consider, however, that most districts, somewhat due to the requirements of the bargained contracts with teachers, compensate teachers for staff development activities through an increase in salary, thus representing a "hidden" cost of traditionally-delivered staff development. For example, a study of spending on professional development in the Los Angeles Unified School District (Ross, 1994) found that the district expended $1,153 million in teacher salaries in 1991-92, and that 22% of this figure could be attributed to salary point credits that were earned because of courses or other approved professional development activities on the part of teachers. The analysis goes on to call several of the features of the salary point credit system into question and makes proposals for improving the current investment being made in teachers' professional development.

As the example of investing in professional development through salary increments implies, there is a pronounced difficulty in fully accounting for all the costs incurred. Professional development activities frequently are financed through a combination of revenue sources, including non-governmental sources, thereby complicating the cost accounting. Professional development experiences also might be associated with substantial contributions of volunteer time on the part of teachers

(Little et al., 1987). At the same time, teachers might accrue additional credits for professional development activities which advance them on the salary schedule, resulting in a long-term fiscal obligation to the district in the form of the resultant base salary increase. Finally, similar professional development activities might vary significantly in costs per teacher depending on the financing strategy which is employed. For example, one strategy for supporting teacher professional development which is increasing in popularity is the "early release" option in which students are released from school on some regular basis, thereby allowing time during regular school hours for teachers to engage in professional development. This option clearly is less costly for school districts, as it removes the additional costs of substitutes or additional hours worked by teachers. However, there is a significant opportunity cost borne by students in the form of reduced instructional time.

The studies of professional development costs briefly reviewed above concentrate on the more traditional forms of professional development delivery. However, significant changes have been taking place in recent years regarding the conceptualization of effective teacher professional development (Fullan, 1993; Little, 1993; Smylie, 1995, Johnson, 1990; Corcoran, 1995), resulting in significant re-thinking of how professional development is best provided (National Foundation for the Improvement of Education, 1996; Darling-Hammond & Ball, 1997). This re-conceptualization of professional development presents a number of conceptual and technical challenges for cost studies, (Note 2) including methods for assigning costs to professional development activities which are integrated into the instructional day and/or more informal interactions among teachers. Moreover, recent thinking about professional development raises questions about whether investments in conventional staff development are likely to contribute much to improving the quality of teaching.

Teacher compensation

Historically, teachers have been compensated for their efforts through a system which is based on an entry level salary. The base salary is then augmented by increments on an established salary schedule based primarily on years of teaching experience and levels of additional education (such as advanced degrees or credit for professional development activities). The level of teacher compensation is a perennial resource allocation question and is primarily determined by decisions about the salary schedule. While the argument can be made that raising compensation levels will assist in attracting and retaining quality teachers, the traditional form of teacher compensation, based on the two factors of years of experience and levels of education and training, does not provide the formula for producing the very best teachers. Consequently, research on teacher compensation has attempted to uncover the types of incentive system that are more closely linked to improved quality of teaching and student learning.

In the past two decades, a variety of reforms to the traditional system of teacher compensation have been attempted. During the early 1980s, merit pay was re-introduced as a policy alternative. In principle, merit pay individually rewards teachers based on the performance of their duties. Some merit pay plans provide for an individual financial bonus on a yearly basis, while other plans call for a permanent advancement on the salary schedule (Darling-Hammond & Berry, 1988). In many instances where they have been tried, merit pay systems have been abandoned, primarily due to internal dissension and problems determining who would receive the additional pay (Murnane & Cohen, 1986; Robinson, 1983). In addition to merit pay proposals, the idea of teacher career ladders has been put forth as another type of alternative compensation strategy, but programs based on this idea have met with a similar lack of success (Freiberg & Knight, 1991; Bellon et al., 1989; Southern Regional Education Board, 1994).

Why have the various attempts at altering teacher compensation borne so few fruitful results? One possible explanation is that the traditional salary structure provides for horizontal equity. That is, teachers are treated as equals on the salary schedule regardless of their gender, race, or teaching assignment (Protsik, 1996). This well-established practice provides for a uniformity of application across teachers that is resistant to change. Others assert that teachers are primarily

motivated by intrinsic rewards that result from the process of working as a teacher (Lortie, 1975; Conley & Levinson, 1993; Richardson, 1990) rather than changes in compensation rates. Firestone (1991) offers the view that research on merit pay has not sufficiently considered the relationship between money and teacher motivation. Firestone distinguishes between merit pay systems (which reward some teachers for doing essentially the same work better than other teachers) and job enlargement reforms (which provide additional compensation to teachers for doing different work) and argues that job enlargement is more closely linked to teachers' intrinsic motivations.

Another explanation is that prior reforms in compensation have focused on individually-based rewards rather than rewards for group performance. An alternative approach to teacher compensation suggested by Mohrman, Mohrman & Odden (1996) includes group-based performance rewards as well as skill- based and competency-based pay. The authors emphasize that the basis for determining the specific skills, competencies, and group rewards must be that the rewards support the central educational purposes of the school and are well suited to the type of organizational arrangements that define the particular site. Further work on the development of alternative designs for compensation systems that are more tightly connected to school improvement have been advanced by Odden & Kelley (1997). Finally, the work of the National Board for Professional Teaching Standards provides a basis for compensating teachers' knowledge and skills by demonstrating the achievement of higher levels of knowledge and expertise through the use of a rigorous professional review process.

Research also has been conducted regarding the alignment of compensation strategies with various education organizational designs. Kelley (1997) noted that historically teacher compensation has been viewed as separate from other aspects of reforming educational organizations. The author analyzes how compensation systems differ under four types of organizational models: scientific management, effective schools, content-driven, and high standards/high involvement and recommends that the design of teacher compensation systems should be better fitted to the type of organizational design which represents the school setting in which teachers work, including the organization's structure, values, and goals. There are states (e.g., Kentucky and South Carolina) and local school systems (e.g., Dallas, TX; Charlotte-Mecklenburg, NC; and a very recent pilot program in Denver, CO) which are in the process of implementing alternative compensation plans. Places where alternative compensation plans have been developed and implemented have relied on participation by educational administrators, teacher unions, and community members in the plan's design (Odden & Kelley, 1997).

Investments in teacher compensation, as in teacher professional development, are policies which have been commonly employed in efforts to improve teacher quality. Research on human resource development, particularly that which is derived from human capital theory, indicates that the proxies which have been used to capture important elements of teacher quality (e.g., verbal aptitude, degree earned, and years of experience) provide an incomplete picture of the factors which affect teaching quality. Most of the research to date on human resource development in education has focused on tracking the quantity of particular inputs that are presumed to be positively associated with teacher quality. A critical need exists for research which attempts to advance our knowledge about the conceptual links between investments in teacher professional development, teacher compensation, and improved teacher and student performance.

Implications for policy and research

This review of economic perspectives from human capital and productivity theories has implications for the design and implementation of investment policies targeted at improving teacher quality. In this section, we explore some of the possible policy implications in an effort to stimulate thinking and dialogue among educators, researchers, and policymakers.

How can we consider the knowledge gained from economic perspectives in its application to current policy debates about teacher quality? One set of observations about how we might characterize knowledge gained from economic research on

productivity and human capital and their implications for policy is provided below.

A significant challenge emerges from the lack of a solid conceptual framework for understanding the important elements in the education process. The lack of sophisticated models for the assessment of student learning needs, the application of teacher knowledge and skills in the instructional process, and the ways in which teachers enact a variety of resources to support instruction accounts for some of the existing shortcomings of econometric analyses of productivity. Many existing policy and resource allocation strategies for improving teacher quality are not theoretically linked to student outcomes. This lack of sufficient knowledge about how policies and resources are enacted by teachers to improve the quality of teaching and learning is precisely the reason why it is so problematic to design cost-benefit analyses of existing investments in teacher quality.

Alongside the conceptual challenges, and in part derived from them, econometric perspectives on the productivity of investments in teacher development face a multitude of measurement challenges. First, and perhaps, most importantly, difficulties exist in specifying the student outcomes to be assessed. While significant progress has been made in productivity research, primarily in microanalytic studies, we still face the question of how to improve on our measures of student learning. Test scores provide an insufficient measure of the content, number, and types of performances expected by the ambitious learning standards that the education reform efforts of this decade have promoted. Adding to the complexity is the extent to which the selected set of standards is universally applied (Monk & Rice, in press). Consequently, analyses of the extent to which specific investments have resulted in improved efficiency (that is, improved student learning according to the set of standards being addressed) are ultimately dependent on our ability to develop clearer, more appropriate outcome measures. In a similar vein, improvement also is needed in the proxies we use for teacher quality. The typical proxies such as years of experience, scores on standardized tests of verbal ability, degrees and credentials earned, and academic field are insufficient indicators of teacher quality. However, current work on developing and implementing teacher standards (such as the National Board for Professional Teaching Standards and the INTASC standards) holds promise for the improvement of measurements of teacher quality.

The lessons learned from human capital theory, reviewed earlier, suggest that the quantity of a resource, the quality of a resource, and the ways in which a resource is configured in its employment are all important aspects of assessing the resource's productive potential. When we view the economic research on the relationship between resources, productivity, and teacher quality, we find that tracking "investments" in teacher quality have been mostly limited to tracking proxies for the *quantity* of a given resource. While economic theory acknowledges the difference between the quantity and the quality of a given input, the research to date indicates that resource allocation strategies for improving teacher quality (1) overemphasize the effects of the quantity of resources, (2) give short shrift to the analysis of the effects of the quality of the resource, and (3) do little to illuminate the effects of re-configuring or reallocating resources—that is, does not help us get at the alternative uses of the same resources. Current economic models for examining the effectiveness of resource allocation practices targeted at teacher quality help articulate the challenges we must face, but are insufficient in their current state to provide the types of analyses that policymakers might find most useful.

In what ways might our conceptions of policy aimed at improving the quantity, quality, and reconfiguration of resources for teacher quality be improved? We might begin by first assuming that productivity can be improved through the re-allocation or re-configuration of existing resources. In other words, if we were to hold the overall quantities of resources constant, then we might focus more centrally on how the resources are allocated and used. There is a little research in this area, but recent work has pointed to the positive contributions and the efficiencies associated with redesigning resource allocation practices (Miles & Darling-Hammond, 1998; Miles, 1997; Odden & Busch, 1998). Resource re-allocation expands our traditional notions of how to bring resources to bear on the achievement of higher productivity. It also shifts the questions one asks, from those concerning the effects of incremental resource increases (a typical question in productivity research) to questions regarding the effects of alternative configurations of the same resource. In other

words, rather than seeking a new program from a new funding source, resources are viewed as available for redesign in order to develop a more productive way of managing existing resources. One of the most prominent resources to be re-configured is the allocation of time that teachers spend with students and with other educators.

From a policy standpoint, resource re-allocation challenges the typical manner by which new policies or initiatives are introduced by policymakers for implementation by educators. The press felt by policymakers to seek out solutions to problems faced in education often results in a response which includes the establishment of new guidelines, regulations, and/or opportunities, and may or may not be accompanied by the infusion of additional fiscal resources. That is, most education policies are not designed to be fiscally neutral. However, resource re-allocation assumes that there are no new dollars available for distribution. Rather, resources are shifted from the support of one program configuration or policy initiative to some other configuration or purpose. This implies that investment priorities change, resulting in the reduction or removal of goods or services that presumably were valued by some constituency. This shift is likely to encounter at least some resistance by those individuals or groups whose interests are perceived to be adversely affected by a particular re-allocation strategy. Consequently, policies which depend on resource re-allocation require a different approach than the traditional strategies of providing financial incentives for adopting particular policies or threats of loss of funding for failure to meet specific requirements.

There are multiple policy options that can influence teaching quality, each having implications for resource allocation or reallocation. Unless care is taken, however, investments in one policy may hinder the advancement of another, equally important aspect of teacher development. Let us consider the following example. One common and long-standing teacher compensation policy strategy has been focused on the goal of attracting and retaining higher quality teachers by raising salary levels. While human capital theory would indicate that this strategy has an evidentiary base, this policy might hinder the acceptance of other notions of compensation, such as skills-based pay. Another example taken from policies related to the provision of teacher professional development further illustrates the potential conflict among policy strategies. Traditional teacher compensation policies provide financial incentives for teachers who accrue additional continuing education credits. The acquisition of these credits is mostly within the purview of the individual teacher, and the type, amount, and quality of the offerings selected may or may not be an optimal match with the types of knowledge and skills which might be most effective in supporting the teacher's work with students. Additionally, the typical manner in which these continuing education credits are delivered often run counter to current notions of best practice in professional development. To further complicate matters, professional development opportunities are also connected to special revenue sources,(Note 3) each with its own set of guidelines and reporting requirements. Consequently, policymakers typically face a challenge when attempting to introduce new approaches to professional development as they will most likely face pressure to continue with existing forms of teacher compensation, add on new supports for the delivery of professional development, and ensure that activities which are undertaken meet the requirements of the various funding sources. Faced with this complexity, a crazy quilt approach to resource allocation for professional development often results. This mixed bag of resource allocation strategies does not take advantage of the potential opportunity for resource re-allocation fashioned through a more strategic approach. In short, without a comprehensive approach to policies which are aimed at improving teacher quality, it is unlikely that resources will be maximized.

Much work is being done throughout the nation to assist policymakers with the development of a comprehensive approach to addressing the improvement of teacher quality. The work of partner states who are collaborating with the National Commission on Teaching and America's Future is one such example of an effort to develop comprehensive policy strategies that support teacher quality. In order to maximize the effectiveness of this type of strategic approach, policymakers must also develop resource allocation policies which are responsive to and reflective of a comprehensive approach to investments in teacher quality.

In sum, economic perspectives can provide some useful insights in addressing the complex challenge of how resources can best be allocated for the improvement of teacher quality. Many questions regarding the effectiveness of resource allocation for this purpose remain. However, lessons learned from an economic perspective, particularly from human capital theory, indicate that we should be cautious of policy approaches which are simply additive. Instead, increased attention should be devoted to policies which focus on more finely tuned notions of teacher quality. Finally, initial work which investigates policies and practices which result in the re-configuration of existing resources ought to be significantly expanded.

Notes

1. For a contemporary review of the contributions made to the study of human capital theory, see Sweetland, S. (1996). Human capital theory: Foundations of a field of inquiry, *Review of Educational Research* 66(3), 341-359.
2. For a discussion of these cost implications, see Rice, J.K. (1999) "Recent trends in the theory and practice of teacher professional development: implications for cost," paper presented at the annual conference of the American Education Finance Association, March 18-20, 1999.
3. Examples of special revenue sources at the federal level which contain funding for professional development include Title 1, Part A (Basic and Concentrated Grants), Title II (the Eisenhower Professional Development Program), Title II (the Technology Literacy Challenge Fund), Title IV (Safe and Drug-Free Schools and Communities), Title VI (the Innovative Education Program Strategies fund), and Goals 2000:Educate America Act. Numerous special funding sources for professional development exist at state and local levels as well.

References

Barnett, W. S. (December 1994). Obstacles and opportunities: Some simple economics of school finance reform. *Educational Policy 8*, 4: 436-452.

Bellon, E.C., J. J. Bellon, M. A. Blank, D. J. G. Brian, and C. A. Kershaw (1989, March). Alternative incentive programs for school based reform. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Benson, C.S. (1978).*The economics of public education* (3rd ed.) Boston: Houghton Mifflin.

Biddle, B.J. (1997). Foolishness, dangerous nonsense, and real correlates of state differences in achievement. *Phi Delta Kappan* (September), 9-13.

Brown, B. W. & D. H. Saks (1975). The production and distribution of cognitive skills in schools. *Journal of Political Economy 83*, 3: 571-593.

Carroll, J. (1963). A model for school learning. *Teachers College Record 64*: 723-733.

Cohen, D. K., M. McLaughlin, & J. Talbert (Eds.) (1993). *Teaching for understanding: Challenges for policy and practice.* San Francisco: Jossey-Bass.

Conley, S. & R. Levinson (1993). Teacher work redesign and job satisfaction. *Educational Administration Quarterly 29*, 4: 453-478.

Corcoran, T.B. (1995). *Helping teachers teach well: Transforming professional development.* New Brunswick, NJ: Consortium for Policy Research in Education Policy Briefs.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review

of state policy evidence. *Education Policy Analysis Archive 8*, No. 1. Available online at http://epaa.asu.edu/epaa/v8n1/.

Darling-Hammond, L.(1994). The current status of teaching and teacher development in the United States. Background paper prepared for the National Commission on Teaching and America's Future.

Darling-Hammond, L. & B. Berry (1988). *The evolution of teacher policy*. Santa Monica, CA: RAND.

Darling-Hammond, L., & M. W. McLaughlin (1995, April). Policies that reform professional development in an era of reform. *Phi Delta Kappan 76*, 8: 597-604.

Darling-Hammond, L. & D. Ball (1997). *Teaching to high standards: What policymakers should know and be able to do*. A report to the National Education Goals Panel (http://www.negp.gov/Reports/highst.htm).

Education Commission of the States (1997). Investment in teacher professional development: A look at 16 districts. Denver, CO: Author.

Ehrenberg, R.G. & D. J. Brewer (1995). Did teacher's race and verbal ability matter in the 1960s? *Coleman* revisited. *Economics of Education Review 14*: 291-299.

Elmore, R. (1994). Thoughts on program equity: Productivity and incentives for performance in education. *Educational Policy 8*, 4: 453-459.

Elmore, R. (1997). *Investing in teacher learning: Staff development and instructional improvement in Community School District #2, New York City*. Washington, DC: National Commission on Teaching and America's Future, Consortium for Policy Research in Education.

Ferguson, R. F. & H.F. Ladd (1996). How and why money matters: An analysis of Alabama schools. In Helen Ladd (ed) *Holding Schools Accountable*. (pp. 265-298) Brookings Institution: Washington, DC.

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation 28*, 2: 465-497.

Firestone, W. (1991). Merit pay and job enlargement as reforms: Incentives, implementation, and teacher response. *Educational Evaluation and Policy Analysis 13*, 3: 269-288.

Freiberg, J. & S. Knight (1991). Career ladder programs as incentives for teachers. In S. C. Conley and B. S. Cooper (Eds.) *The School as a Work Environment: Implications for Reform*. Boston: Allyn and Bacon.

Fullan, M. (1993) *Change Forces: Probing the Depths of Educational Reform*. London: Falmer Press.

Hanushek, E.A. (1981) Throwing money at schools. *Journal of Policy Analysis and Management 1*, 1: 19-41.

Hanushek, E. A. (1991). When school finance "reform" may not be good policy. *Harvard Journal on Legislation 28*, 2: 423-456.

Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher 18*, 4: 45-65.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature 24*: 1141-1177.

Hargreaves, A. (1990). Teachers' work and the politics of time and space. *Qualitative Studies in Education 3*: 303-320.

Hargreaves, A. (1993). Individualism and individuality: Reinterpreting the teacher culture. In J. W. Little and M. W. McLaughlin (Eds.), *Teachers' work: Individuals, colleagues, and contexts*. New York: Teachers College Press.

Hedges, L.V., R.D. Laine, & R.Greenwald (1994, April). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher 23*, 3: 5-14.

Houston, R. W. & J. H. Freiberg (1979). Perpetual motion, blindman's bluff, and inservice education. *Journal of Teacher Education, 30*, 1: 7-9.

Hughes, J., C. G. Moon, & W. S. Barnett (1993, October). Revenue-driven costs: The case of resource allocation in public primary and secondary education. Paper presented at the annual meeting of the Atlantic Economic Society, Philadelphia, PA.

Johnson, S. M. (1990). Redesigning teachers' work. In R.F. Elmore & Associates, *Restructuring schools: The next generation of educational reform*. San Francisco: Jossey- Bass.

Kelley, C. (1997). Teacher compensation and organization. *Educational Evaluation and Policy Analysis 19*, 1: 15- 28.

Lankford, H. & Wyckoff, J. (1996). The allocation of resource to special education and regular instruction. In H.F. Ladd (Ed.), *Holding schools accountable* (pp. 221-262) Washington, DC: Brookings Institution.

Little, J.W., W.H. Gerritz, D.H. Stern, J.W. Guthrie, M.W. Kirst & D.D. Marsh (1987). *Staff development in California*. Policy Analysis for California Education (PACE) and Far West Laboratory for Educational Research and Development, (Policy paper #PC87-12-15, CPEC), San Francisco, CA.

Little, J.W. (1989). District policy choices and teacher's professional development opportunities. *Educational Evaluation and Policy Analysis 11*, 2: 165-179.

Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis 15*, 2: 129-151.

Lortie, D. C. (1975). *Schoolteacher*. Chicago: University of Chicago Press.

Lytle, J. H. (1983). Investment options for inservice teacher training. *Journal of Teacher Education 34*, 1: 28-31.

Miles, K.H. (1997, June) Rethinking the use of teacher resources. *School Business Affairs 63*, 6: 35- 40.

Miles, K.H. & L. Darling-Hammond (1998) Rethinking the allocation of teaching resources: Some lessons from high- performing schools. *Educational Evaluation and Policy Analysis 20*, 1: 9-29.

Miller, B., B. Lord, & J. Dorney (1994). *Staff development for teachers: A study of configurations and costs in four districts*. Education Development Center, Newtonville, MA.

Mohrman A. M., S. A. Mohrman, & A. R. Odden (1996). Aligning teacher compensation with systemic school reform: Skill- based and group-based performance rewards. *Educational Evaluation and Policy Analysis 18*, 1: 51-71.

Monk, D. H., (1990). *Educational finance: An economic approach*. New York:

McGraw Hill.

Monk, D.H. (1992). Education productivity research: an update and assessment of its role in education finance reform. *Educational Evaluation and Policy Analysis 14*: 307-332.

Monk, D.H. & J.K. Rice (in press). Modern education productivity research: Emerging implications for the financing of education. In *Selected Papers in School Finance*. Washington, DC: National Center for Education Statistics.

Moore, D., & A. Hyde (1981). *Making sense of staff development: An analysis of staff development programs and their costs in three urban school districts*. Chicago: Designs for Change.

Mortimer, P., P. Sammons, L.Stoll, D. Lewis, & R. Ecob (1988). *School matters* Berkeley, CA: University of California Press.

Murnane, R.J. (1991). Interpreting the evidence on "Does Money Matter." *Harvard Journal on Legislation* 28: 457-464.

Murnane, R. J. (1975). *The impact of school resources on the learning of inner city children.* Cambridge, MA: Ballinger.

Murnane, R.J. & D. K. Cohen (1986). Merit pay and the evaluation problem. *Harvard Educational Review 56*, 1: 1-17.

National Foundation for the Improvement of Education (1996) *Teachers take charge of their learning: Transforming professional development for student success.* Washington, DC: Author.

Nikansen, W. A. (1971). *Bureaucracy and representative government.* Chicago: Aldine Atherton.

Odden, A. & C.Busch (1998). *Financing schools for high performance: Strategies for improving the use of educational resources.* San Francisco, CA: Jossey-Bass.

Odden, A. & C.Kelley (1997). *Paying teachers for what they know and do: New and smarter compensation strategies to improve schools.* Thousand Oaks, CA: Corwin Press.

Odden, A. & W.Clune (1995). Improving educational productivity and school finance. *Educational Researcher 24*, 9: 6-10, 22.

Orlich, D. C. & A. Evans (1990). *Regression analysis: A novel way to examine staff development cost factors.* Unpublished manuscript. ERIC Reproduction Service Document No. ED 331 808.

Protsik, J. (1996). History of teacher pay and incentive reforms. *Journal of School Leadership 6*, 3: 265- 289.

Resnick, L.(1993). Standards, assessment, and educational quality. *Stanford Law and Policy Review 4*: 53- 59.

Richardson, V. (1990). Significant and worthwhile change in teaching practice. *Educational Researcher 19*, 7: 10- 18.

Robinson, G. (1983). *Paying teachers for performance and productivity: Learning from experience.* Arlington, VA: Educational Research Service.

Ross, R. (1994). *Effective teacher development through salary incentives (An exploratory analysis).* RAND, Institute on Education and Training.

Rossmiller, R.A. (1986). *Resource allocation in schools and classrooms: Final report.* Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Education Research, School of Education.

Smith, A.(1776). *An inquiry into the nature and causes of the wealth of nations.*

Smith, M.S. B.W. Scott, & J. Link (1995). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics 77*: 251-266.

Smylie, M.A. (1995) Teacher learning in the workplace: Implications for school reform. In T.R. Guskey & M. Huberman (Eds), *Professional development in education* (pp. 92- 113). New York: Teachers College Press.

Southern Regional Education Board (1994). *Reflecting on ten years of incentive programs: The 1993 SREB Career Ladder Clearinghouse Survey.* Atlanta, GA: Author.

Stern, D. S., W.Gerritz, & J.W. Little (1989). Making the most of the district's two (or five) cents: Accounting for investment in teachers' professional development. *Journal of Education Finance 14*, 4: 19-26.

Summers, A.& B.Wolfe (1977). Do schools make a difference? *American Economic Review 67*, 4: 639-652.

Thomas, J. A. & F. Kemmerer (1983). *Money, time, and learning.* Albany, NY State University of New York at Albany, School of Education.

Witte, J.F. (1990, August). Understanding high school achievement: After a decade of research, do we have any confident policy recommendations? Paper presented at the annual meeting of the American Political Science Association, San Francisco.

About the Author

Margaret L. Plecki
Educational Leadership and Policy Studies
University of Washington
Box 353600
Seattle, WA 98195-3600
Email:mplecki@u.washington.edu

Margaret Plecki is an Assistant Professor at the University of Washington, Area of Educational Leadership and Policy Studies. Her research and teaching focus on school finance, the economics of education, school improvement, and education policy. She earned her Ph.D. from the University of California, Berkeley.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografía e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

BEST COPY AVAILABLE

This article has been retrieved 334 times since July 17, 2000

prior vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 8 Number 34 July 17, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Implementation of the Kentucky Nongraded Primary Program

Patricia J. Kannapel
Lola Aagaard
Pamelia Coe
Cynthia A. Reeves

AEL, Inc.
Charleston, West Virginia (U.S.A.)

Abstract

We examine the development of the Kentucky nongraded primary program at the state level, and in six rural elementary schools from 1991 through 1998 (case studies of four of these schools are included in Appendix A). Data collected from our longitudinal qualitative study reveal that teachers changed their classrooms in response to the primary program mandate, and some positive outcomes occurred for students. Implementation was hampered, however, by rapid implementation timelines, failure to clearly articulate the purpose of the program and how it linked with a larger reform effort, and a firmly entrenched "graded" mindset. Currently, progress toward full implementation of a continuous progress model for primary students has stagnated. To revive the program, policymakers need to make program goals clear, demonstrate how its implementation will facilitate attainment of reform goals, and assist teachers in implementing the program as intended. (Note 1)

Introduction

The concept of nongraded schooling is not new. Nongraded, multi-age education has moved in and out of favor throughout the educational history of the United States. Yet, even though the notion of nongradedness often conjures up a

positive image of children moving at their own rate, of older students helping younger ones, and of younger students learning from older ones, nongraded schools and classrooms have failed to take hold in public schools in any large- scale or long-term way over the past several decades. Graded schools became the norm in urban school districts in the latter half of the 19th century, and in rural schools a short time later (Tyack, 1974), and have persisted to the present day. Tyack & Cuban (1995) suggest that, because the graded school arrived on the scene at a time when elementary education was rapidly expanding, and offered a standardized way to process large numbers of students, the organization of schools by grades became the generally accepted form of American public education. In this sense, gradedness might be thought of as one of the characteristics of the "real school," a concept proposed by Metz (1990) to signify a common script that American schools have come to follow, and that has come to be widely accepted by educators and parents alike. This article examines a recent attempt to stem the tide of gradedness: Kentucky's statewide effort to replace grades K-3 with a nongraded, continuous progress model.

Study Description

This report is based on findings from a longitudinal study of implementation of the Kentucky Education Reform Act (KERA) conducted by researchers from AEL, Inc. The research team studied state-level implementation, as well as implementation in four rural districts. AEL followed implementation in rural settings in Kentucky because most Kentucky school districts are rural, AEL had a rural focus at the time, and comprehensive reform in rural districts has been little reported or documented. The study districts were selected from a list of districts identified by various Kentucky stakeholders and policymakers as representative of "typical" Kentucky school districts: we asked that they identify districts that were neither at the forefront of reform, nor likely to subvert it. From 1991 through 1995, we studied the primary program along with other aspects of KERA implementation in all 15 elementary schools in the four districts. From 1996 through 2000, we narrowed our focus to six schools, and to a specific cohort of students within those schools: the class of 2006—a group whose entire schooling had been under KERA, and who were completing the primary program in 1996-97. This study sample of six schools included two schools in western Kentucky, two in central Kentucky, and two in eastern Kentucky. Four of the schools were located in towns, while two were in outlying communities or rural areas. Five were located in county districts; one was in a small, independent school district. When compared to urban and suburban schools, our study schools were relatively small, ranging in size from 80 students to 500 students. The percentage of students on free/reduced lunch has fluctuated throughout the study period, ranging from about 30-40 percent at the low end to 60-70 percent at the high end.

The study was qualitative in nature: we relied on interviews, observations, and review of documents to provide information. Across the years of the study, we observed over 180 hours in primary classrooms and conducted approximately 400 interviews with administrators, teachers, parents, primary students, and state officials. We also observed professional development sessions on the primary program. Documents analyzed included lesson plans, primary program action plans and annual evaluations, school transformation plans, school council minutes, school board minutes, and local newspapers. At the state level, we interviewed key officials who where instrumental in primary program implementation, regularly attended meetings of the Kentucky Board of Education, observed early professional development institutes on the primary program, and examined primary program implementation documents.

Our analysis has included extensive review and discussion of our field notes and key documents, as well as a discussion of preliminary findings with state officials, and with administrators and teachers in the local districts. This paper addresses the following questions:

1. What was the state and national context for Kentucky's nongraded primary program?

2. How was the program implemented at the state level?
3. What changes occurred in primary classrooms?
4. How did the primary program affect students?

Additionally, we have included in Appendix A short case histories of primary program implementation at four of the schools we have studied intensively.

The Context for Kentucky's Nongraded Primary

Kentucky's nongraded primary program (hereafter referred to as the "primary program") is but one component of a massive restructuring of the state's educational system. The Kentucky Education Reform Act, passed by the Kentucky General Assembly in the spring of 1990, came about as the result of a lawsuit filed by 66 of the state's poorest school districts charging that the state's system of financing public schools placed too much emphasis on local resources (Rose v. Council for Better Educ., 1989, p. 4). The Kentucky Supreme Court ruled in the summer of 1989 that the entire state school system was unconstitutional, and ordered the state legislature to restructure entirely the state's system of public schooling.

The Kentucky legislative leadership organized a task force, composed of legislators and representatives from then-Governor Wallace Wilkinson's office, to design the restructuring package. Subcommittees on curriculum, governance, and finance were created to work out the details of the reform. Each committee hired a national consultant to assist in developing its portion of the restructuring package. The consultant who designed the curriculum package, which contains the primary program, was David Hornbeck, then of Hogan and Hartson in Washington, D. C., but currently superintendent of the Philadelphia schools. Hornbeck, with substantive input from the Task Force and the Governor's office, designed a reform package that shifted the focus from teacher inputs to student results, required schools to ensure high levels of achievement for *all* students, and gave schools autonomy to decide how to help students achieve reform goals, but held them accountable for student performance as measured by a performance-based assessment instrument. This restructuring package strongly reflected an approach that would soon become known as "systemic reform" (Cohen, 1995; Fuhrman, Elmore, & Massell, 1993; Murphy, 1990; O'Day & Smith, 1993; Schwartz, 1991; Smith & O'Day, 1991).

The groundwork for this brand of restructuring had been laid by Governor Wallace Wilkinson in the two years prior to 1990. Wilkinson, with guidance from Education Secretary Jack Foster, developed an "outcomes-based" restructuring plan that called for school-based management, leadership and staff development, increased resources for instructional improvement, an outcome-based curriculum, performance standards, accountability, and a rewards program (Wilkinson, 1988a, 1988b).

As pointed out by Fuhrman, Elmore, & Massell (1993), the primary program—with its requirement that schools eliminate grades K-3—was a curious addition to a reform package that called for locally-designed instructional inputs. Former Kentucky Education Secretary Jack Foster, who served on the task force that designed KERA, explained inclusion of the primary program:

> Although not specifically proposing creation of a primary program, Governor Wilkinson contended in his reform proposal prior to the Supreme Court decision that it was time to alter the structure of the school to enable teachers to work more effectively with children who have different learning styles, aptitudes, or interests. Wilkinson contended that the traditional school leaves the educational needs of many children unmet because it is not flexible enough to meet their different learning needs.... A classroom in which everyone is studying the same thing at the same time is not one that can easily adapt to individual differences in either learning style or ability. With this as background, David Hornbeck, consultant to the curriculum committee of the Task Force on Education Reform, [recommended that grades K-3 be replaced with an ungraded model] (Foster, 1999, p. 70).

In addition to the push from the Governor's office, the decision to include a nongraded primary program as a starting point to a results-based restructuring plan is likely related to the fact that, at the time KERA was developed, nongraded instruction was making a resurgence as a "new" schooling structure (Anderson & Pavan, 1993). The recent movement toward nongraded instruction is a response to research in child development and the learning process, which suggests that nongradedness is an appropriate strategy for curbing ability tracking and grade retention, which have been shown to have harmful effects on children (Massachusetts Board of Education, 1990). Proponents of nongraded primary programs argue that they provide a developmentally appropriate way for teachers to deal with individual differences found among children at an age when they are psychologically vulnerable (National Association for the Education of Young Children, 1987; National Association of Elementary School Principals, 1990).

While nongraded programs are seldom cited as a feature of systemic reform, the emphasis in nongraded programs on tailoring instruction to individual needs so that all students can achieve is quite compatible with the systemic reform movement's emphasis on helping *all* children achieve rigorous academic standards. Unfortunately, this sort of link between the primary program component of KERA and the larger reform package was not made clear to Kentucky educators. In Hornbeck's final recommendations to the legislative task force, the primary program appears on page 65 of a 66-page document, is described in three sentences, and is not linked conceptually with the systemic-reform-like recommendations that precede it (Hornbeck, 1990).

Jack Foster acknowledged that the rationale for the primary program, and its link with the larger reform, was never made clear:

> We dropped that one in there very late... . We had no protocols, no models, we had no documentation, no references to literature, nothing. It just appeared. So it really left the Department of Education to do whatever they wanted. I was asked a couple of times to come over and interpret to them what we had in mind. Hornbeck was gone by now. I used my own philosophy as to the intent of that... So we got what we deserved on that one. You never want to lay something that significant into a piece of legislation without some sort of supporting documentation that people can use to get at the legislative intent. But there is nothing; there is nothing (personal communication, 9/17/99).

State-Level Implementation of the Primary Program

Radical change is a difficult and often messy process, an observation well-documented in the education change literature (see Fullan, 1996). The implementation of the primary program was no exception. The lack of clearly articulated legislative intent hampered primary program implementation from the outset. State officials involved in early implementation of the primary program, along with the first program description issued by the Kentucky Department of Education (KDE), reported that Department staff had to engage in extensive research to get at the intent of the primary program. The program description, entitled *The Wonder Years* (Kentucky Department of Education, 1991), states that staff examined all statutory provisions regarding the primary program; reviewed the provisions of KERA that impact the primary program; reviewed the curriculum committee recommendations; reviewed direction and clarification provided by David Hornbeck, other Task Force members, and legislative staff; reviewed literature and research on "nongradedness;" reviewed position statements of national organizations for the education of young children; attended conferences and heard national consultants; and visited schools with nongraded programs. From this research, the KDE identified seven critical attributes of the program, which focused around how primary classrooms should look, rather than what primary teachers should teach. The attributes were **developmentally appropriate educational practices, multi-age/multi-ability classrooms, continuous progress, authentic assessment, qualitative reporting methods, professional teamwork, and positive parent**

involvement. According to staff at KDE who were instrumental in developing the position statement, the critical attributes were meant to serve as a guide to schools as they developed their primary programs. The 1992 General Assembly, however, adopted the attributes into law.

The critical attributes quickly became the linchpin of the primary program, not only because they were now mandated, but because the attributes were virtually the only guidelines for reform implementation in the early years. The state's assessment contractor was developing the new performance assessment instrument, and the KDE was beginning to develop curriculum frameworks. But the primary program attributes were the first piece of guidance to fall into place, and it was around the attributes that professional development and primary program directives revolved.

Early implementation was further complicated by the implementation timelines. The original KERA legislation laid out no specific timelines for implementation. The 1991 program description suggested that implementation would occur over a three-year span beginning in 1992-93, but a former KDE official reported to us that KDE had envisioned full implementation occurring by 1996. This recommended gradual approach might have facilitated linkages between the primary program and the larger reform because curriculum supports could have been put in place to help primary teachers plan what they were to teach (KERA goals) before having to follow the state plan for how to teach it (critical attributes). In 1992, however, apparently in an effort to jump-start reform by getting the primary program in place, the legislature mandated beginning implementation in 1992-93, and full implementation by 1993-94.

The unintended effect of the new timeline, coupled with the critical attributes becoming statutory requirements, was that teachers were thrust into the overwhelming demands of multi-age classrooms before the state had provided the curriculum guidance required by KERA. State curriculum frameworks did not appear until 1993 (Kentucky Department of Education, 1993b), and the even more widely used *Core Content for Assessment* was not available until 1996 (Kentucky Department of Education, 1996a). Consequently, primary teachers fashioned a program that demonstrated implementation of the seven critical attributes, but the fundamental issues of what they were to teach and how the curriculum should align with KERA had not been worked out.

Another aspect of primary program implementation that became problematic was the issue of how to determine when students were ready for fourth grade. An interim process for determining successful completion of the primary program was adopted in December 1992 and is still in effect (Kentucky Department of Education, 1993a). There was some initial thinking that the interim regulation would be replaced by the Kentucky Early Learning Profile (KELP), which was developed by the state's assessment contractor. According to the KELP handbook (Kentucky Department of Education, 1994), this primary assessment tool was not intended to mirror the fourth-grade assessment, but was designed to provide students with opportunities that would lay the foundation for the fourth-grade assessment. The KELP was piloted during the 1992-93 school year and field tested in 1993-94. Training in use of the KELP was made available to primary teachers across the state in the summer of 1994: the summer following the year they were required to fully implement the primary program. Because of concerns about the amount of paperwork associated with the KELP, it was never made mandatory, but schools are expected to use a process similar to that spelled out in the "interim" regulation, or a "KELP-like" process for verifying successful completion of the primary program.

The KELP was not widely adopted across the state. Bridge (1995) reported that most teachers found the KELP so burdensome that they would discontinue using it if given the choice. The state Office of Education Accountability (OEA) reported in both its 1996 and 1997 annual reports that about one-third of schools were using the KELP, and that there was no monitoring of schools not using the KELP to determine if they were using an alternative that met the criteria for exiting the primary program. However, the KDE reported to the OEA in 1999 that, based on a survey returned by 94 percent of elementary schools, 75 percent of schools used one or more components of the KELP. Of this number, 44 percent used the KELP Learning Descriptions, which is the component that assesses students' continuous progress (Kentucky Department of Education, 1999).

The failure to link the primary program to the rest of KERA resulted in a perception among teachers that the primary program was out of sync with reform in grades 4-12. In our study schools teachers expressed concern about the adjustment primary students would face when they reached fourth grade, where behavioral and academic expectations would be more rigid. These teacher perceptions contrasted sharply with what we heard from state officials, who expressed hope that primary program practices would be so successful and well-received that they would work their way up through the grades as teachers, parents and students came to embrace and expect these sorts of practices. A key official at the state department of education commented to us in 1993:

> Now how responsive the rest of the system is to that group of children is going to be the next critical question. It has already been asked. Parents are saying, "What happens when my child leaves this wonderful program where they've become independent thinkers and they go into Miss Jones stringent fourth-grade classroom and they're not allowed to continue on?" Our response is, "If I were you, as a parent, I would really be at the door of that school principal or that school council insisting that the [intermediate grade] program change." That's where the dynamic of change can be. I don't think it should be mandated from here. I think it occurs because it's a good program and they want to continue it.

Former Education Secretary Jack Foster made similar comments:

> It was our hope that [the primary program] would be so successful that by the time [students] came out of the primary, we could convince other teachers up through the elementary school, and get the whole elementary school ungraded (personal communication, 9/17/99).

While state officials expressed the belief that the primary program would mirror the kinds of practices needed at all grade levels to help students achieve the higher-order skills emphasized in the KERA goals and expectations, the vast majority of training and support documents for the primary program did not link the program with KERA goals and expectations. In the primary, the focus was on eliminating student failure and on building student self-esteem and love of learning. This was to be accomplished through mandates as to how primary classrooms should operate (the critical attributes). In grades 4-12, by contrast, the focus was on preparation for the state assessment, which was the tool for judging whether students were making progress toward KERA goals.

Another major influence on primary program implementation was legislation that was meant to facilitate the primary program. Key members of the legislature believed that the focus on multi-aging had detracted from the broader purpose of the primary program. In 1994, the legislature passed a law that added flexibility for schools to determine, based on individual student needs, that multi-age/multi-ability grouping need not apply to every grouping situation throughout the day; and that permitted entry-level (or kindergarten) students to be grouped in self-contained classrooms if developmentally appropriate. Greater flexibility was added in 1996 legislation. These legislative acts relaxing the multiage, multiability requirement were viewed by some teachers as a signal that they no longer had to implement the one attribute that, to them, had become synonymous with the primary program. McIntyre and Kyle reported in 1997 that after multi-age grouping was made optional, fewer teachers were implementing the multi-age component, and that some teachers abandoned the primary program altogether; a phenomenon we also observed in our study schools.

Changes in Primary Classrooms

In the first two years of primary program implementation (1992–93 and 1993–94), primary teachers at our six study schools—in an attempt to implement the attributes—made changes in their approaches to instruction, assessment, grouping practices, reporting methods, working with other teachers, and working with parents.

While virtually all teachers tried new practices in their classrooms, some embraced the changes more enthusiastically than others. This sort of varied implementation was also reported in other studies conducted around the state (Kyle & McIntyre, 1993; Raths & Fanning, 1993; Raths, Katz & Fanning, 1992). Among our study sample, we studied one school where an enthusiastic and persuasive principal and an open-minded faculty combined their energies to make major changes in their approach to instruction and student grouping (see the case study of Orange County Elementary School in Appendix A). At two schools that had previously had high student achievement using traditional approaches, changes were approached with caution by nearly all teachers (see the Newtown Elementary School case study). At two other schools, the issue of how much change to make was divisive (see the case studies of Vanderbilt County Elementary School and Kessinger Elementary School).

While changes in primary classrooms were substantial and widespread initially, movement toward greater implementation of the primary program has stagnated in our study schools, as well as statewide (McIntyre & Kyle, 1997). Generally, primary teachers seem to have settled into an approach that is comfortable for them, whether it equates to full implementation or not. The reactions to and implementation of the primary program in the AEL rural study districts do not seem to involve distinctly rural issues, as similar findings were reported in reviews of other KERA research that included urban areas in the commonwealth (McIntyre & Kyle, 1997). One possible exception might be that none of these districts had tried a nongraded approach since a brief fling with it (when it was last popular) in the 1950s, whereas some of the more urban and suburban districts in the state had been experimenting with the practice for some time before KERA was passed (Kentucky Education Association/Appalachia Educational Laboratory, 1991).

Below, we describe more fully the changes that occurred—and the ones that persisted—under each of the critical attributes. We also consider the perceived disjunction between the primary program and reform in the intermediate grades.

Developmentally appropriate practices. With the new professional development money from KERA, virtually all primary teachers in the study schools received copious training and experimented with new instructional practices. Professional development was weighted most heavily toward developmentally appropriate instructional practices. Teachers reported being simultaneously overwhelmed and energized by what they were learning and doing. One teacher commented in 1992:

> I've attended a lot of workshops, I've attended a lot of seminars, I'm doing some things this summer. I'll be learning more about whole language for two weeks, and I've got a couple of other workshops I'm really interested in. I just finished training to be a math specialist. That was really rewarding. Everything that I have done and every workshop that I've gone to, I've learned a lot and I've tried to apply it in the classroom.

Of all the changes primary teachers attempted, changes in instructional practices were adopted most readily, and have persisted more than have changes in the areas of the other critical attributes, reportedly because teachers have had success with many of the new approaches. In a review of research on the primary program statewide, McIntyre & Kyle (1997) also reported that teachers found developmentally appropriate practices the easiest attribute to implement, continued to use varied instructional practices, and rated this attribute as the most important one in terms of student learning.

The most common practices we observed in the early years were use of hands-on and calendar activities to teach mathematics; thematic or interdisciplinary instruction; use of authentic literature, whole language, or literature-based instruction; journal or other writing activities; and flexible seating arrangements. Although the degree of implementation varied across schools and teachers, virtually all teachers experimented with these practices in the first two years of primary program implementation. In addition, about half of the teachers employed learning centers; a lesser proportion attempted cooperative learning activities. In general, teachers assigned less textbook work, drill, seat work, and rote memorization than in

the past—although these practices were in regular evidence at two of our study schools. Similar findings were reported statewide; Bridge (1995) found that teachers were using a variety of approaches and materials, attempting to integrate the curriculum through theme activities, and arranging the physical environment of their classrooms to facilitate primary program implementation.

As primary teachers tried new approaches, however, they found that developing thematic units, learning centers, and hands-on activities was labor-intensive and time-consuming. In addition, they worried that students would not acquire "basic skills" without the customary drill and practice. These concerns were echoed by intermediate teachers, who began to report almost immediately that primary students were coming to them lacking in basic skills. Thus, after the initial two years of classroom innovation, many primary teachers returned to more traditional practices such as using spelling books to teach spelling, drilling on math facts, and use of workbooks and worksheets to teach phonics. Some of the new instructional practices have persisted in our study classrooms, however, including more flexible seating arrangements, partner or group work, emphasis on process writing, use of authentic literature as part of the primary reading program, and greater use of hands-on activities. Practices that have mostly fallen to the wayside are learning centers (except in entry-level primary classrooms), cooperative learning activities, and broad use of themes or interdisciplinary instruction.

Multi-age/Multi-Ability Classrooms. Probably because the primary program had initially been referred to as the "nongraded primary," and because this was one of the most tangible attributes to be implemented, teachers equated the multi-age, multi-ability attribute most strongly with the primary program. While state officials retrospectively reported to us that this attribute was meant to serve as a tool to enable continuous progress, it was not presented that way in the state guidelines, nor in any professional development we observed. As a result, educators implemented multi-aging as an end in itself, and one that was difficult conceptually and logistically. Two schools initially attempted K-3 classrooms, pulling students into smaller groups (single or dual-age) for skills instruction. Three other schools grouped students into two- and three-age span groups, also breaking them into more homogeneous groups during the day for skills instruction. One school was more cautious, never experimenting with more than a dual-age classroom.

In response to the legislation that relaxed the multi-age requirement, by the 1996-97 school year, three of the six schools studied more intensively since 1996 had returned to single-age classrooms (although one of these has since opted to return to dual-age classrooms), two continued with dual-age classrooms because low enrollment forced split classes, and one school had a K-2, 3-4 arrangement. McIntyre & Kyle (1997) also reported that many schools statewide returned to single-age classrooms. The KDE reported in 1999 that the most common structure in the primary program was dual-age classrooms, with partial inclusion of five-year-olds; and that 21 percent of schools reported single-age groupings (Kentucky Department of Education, 1999).

At no school did we witness the envisioned elimination of "grade differentials." This finding correlates with other research around the state, where it was reported that multi- age/multi-ability grouping was one of the most controversial and difficult attributes for teachers, fewer teachers implemented the multi-age component over time, and teachers viewed the multi-age/multi-ability attribute as least important to student learning (McIntyre & Kyle, 1997; Raths, Katz, & Fanning, 1992). Similarly, a 1999 survey found that a majority of teachers, parents, and the general public did not believe that the graded structure should be eliminated in the first four years of schooling (Kentucky Institute for Education Research, 1999).

Throughout this time period, inclusion of kindergarten students was problematic at our study schools and across the state. Many educators and parents viewed kindergarten as a preparatory program, and did not believe young children should be mixed with older ones when they first began school. The issue was so divisive at one of our study schools that entry-level students were pulled in and out of the program several times during the 1993-94 school year as teachers struggled to reach consensus on integrating these students into the primary program. A parent of one of these students reflected on the experience:

> I felt like that it was a rocky start when he began here. His first year, they started out with multi-age, and some wanted multi-age and others didn't. So they were in that for a couple of weeks and then switched. In the first nine weeks, he had changed three times, teachers, grouping, etc. before they decided how to do it. As a parent, I was not very happy because he was young and immature and having all of that change constantly, not knowing where you are going or who your teacher is... [I have been] generally satisfied other than that beginning year. I just wish that there had been a decision made before school started as to how to do it.

When the 1994 General Assembly enacted legislation that permitted entry-level students to be grouped in self-contained classrooms if such grouping was developmentally appropriate for individual students, five of the six study schools studied intensively took this as a blanket endorsement for placing all entry-level students in self-contained classrooms.

Continuous progress. The state defines continuous progress as follows: "Continuous progress means that students will progress through the primary school program at their own rate without comparison to the rates of others or consideration of the number of years in school. Retention and promotion within the primary school program are not compatible with continuous progress" (Kentucky Department of Education, 1993a, p. 8). While this attribute appears central to the primary program philosophy, primary teachers in our study schools appeared to be more focused on implementing those attributes that had some concrete, visible manifestation: multi-age groups, new report cards, anecdotal records, parent orientation programs, common teacher planning time. Continuous progress was never articulated to us as a major goal of the primary program. Similarly, Bridge (1995) reported that fewer than half of the teachers she studied showed evidence that they were providing for the continuous progress of students through the primary program.

It appeared that the concept of gradedness was firmly entrenched at all levels of the system. Teachers, as well as parents and students, were never able to abandon the concept of gradedness and to think in terms of each student progressing continuously toward acquisition of KERA goals and expectations. Even within dual-age or multi-age classrooms, teachers often referred to students by grade level; or sometimes referred to the level of the task by grade, such as having "first grade spelling words" and "second-grade spelling words." Many schools attempted a change in terminology, so that kindergarten was referred to as P1, first grade as P2, etc. These new terms, however, served the same function as the grade designations--separating students by age. Principals told us that even the KDE required that enrollment information be provided by grade level.

Another difficulty teachers had with the notion of continuous progress had to do with retention. Teachers were told by state officials that the determination of whether students needed to spend a fifth year in the primary program should be made during the fourth year of primary. The rationale for this was that, if schools adopted a truly continuous progress model, then students would work continuously toward acquisition of KERA goals rather than having a determination made at some arbitrary point that they had not made adequate progress and thus, needed to repeat an entire year of instruction. Because the graded model and mentality had not been abandoned, however, the ban on retention created problems. One of our study schools ignored it entirely. Teachers at four schools did make an effort to allow students within their usually-dual-age classrooms to work at an appropriate level, but there was still a need to make a determination as to whether a child was ready to move on to the next dual-age classroom. For instance, where the primary program was configured into K/1st and 2nd/3rd grade classrooms, teachers felt a need to "retain" some students in the K/1st classrooms an extra year rather than send them on to the 2nd/3rd grade room. A principal, who was hired after the school council had voted to return to single-age classrooms in the primary program, described how she saw the single-age configuration at her school impeding continuous progress:

> We have single-age all the way through primary, self- contained. We have done a minimal amount of sliding students [from one level to the

> next to meet individual needs]. We had a child who was not happy and a behavior problem in kindergarten and I suggested moving him to first grade for 45 minutes daily in a skill area he was strong in. Little by little, that child was eased into first grade so he is there all the time. If we had had a multi-age situation, these things could be taken care of in the classroom without all this hullabaloo. It is not a naturally occurring thing that each child's need is met. We are meeting their needs but the curriculum is not set up to do it. We are having to reach out to make it happen.

It should be noted that some teachers had structures for allowing students to progress at their own rate in certain subject areas. At one school, teachers in a dual-age classroom used flexible grouping and regrouping for mathematics instruction, assessing and re-shuffling student groups at the end of each unit. The more common practice, however, was to use grouping practices in which students stayed with the same teacher most of the day and were placed in relatively stable ability groups for reading and math instruction. Even in schools where some teachers had worked out continuous progress within their own classrooms, the movement from one grade level to the next interrupted the smooth continuum of progress for children.

Authentic assessment. Authentic assessment practices attempted by most teachers in the early years included use of anecdotal records to record student progress and behavior as it occurred naturally, and accumulation of student work into portfolios of some type. At two of the six schools studied intensively, teachers, over time, continued to implement practices (such as engaging students in individual or group projects) that were better assessed with alternative instruments, such as scoring rubrics developed for specific assignments. One of these schools continued to use the KELP, mostly because it was a district requirement. At the remaining schools, use of anecdotal records and other authentic assessment techniques had nearly disappeared by the 1996-97 school year. As with multi-aging, teachers at these schools had implemented authentic assessment because it was required rather than as a tool to monitor students' continuous progress. Some teachers reported that they found it useful to share anecdotal records with parents at conferences but, for the most part, teachers were unclear how to manage or make use of these alternative assessment techniques.

Qualitative reporting. Traditional report cards with number/letter grades were replaced in all study schools with qualitative reporting, such as lists of broad skills or capabilities, accompanied by codes or narrative to indicate whether students were progressing or in need of further assistance. Teachers found these reporting systems cumbersome, however. They also reported that parents did not understand the qualitative progress reports. Many parents corroborated this story, reporting that letter grades gave them a better sense of how their children were progressing. As a result, by 1996-97, three of the six schools had replaced the qualitative progress report with a report card with number/letter grades, or some system for equating symbols on the report card with number/letter grades. And, as was the case with authentic assessment, traditional reporting methods were a comfortable fit with the more traditional practices preferred by teachers at these schools. At the one school that used the KELP, student progress was reported to parents in narrative, and was shared at conferences scheduled at regular intervals during the year. Teachers at this school reported that the KELP was time-consuming, but provided a great deal of information about student progress.

Professional teamwork. Primary teachers at all schools initially attempted some form of teaming, and tried to carve out time for common planning. Teaming often meant exchanging or mixing students for a portion of the day so that, for instance, one teacher taught to an advanced group while another taught lower ability students. At one school, however, primary teachers did teach together in a large, open classroom that facilitated communication and flexible grouping and regrouping of students. This sort of teamwork was still in evidence at that school in 1996-97. Over time, initial structures for common planning and teamwork either disappeared or became under-utilized at five of the six schools, as well as around the state (Bridge, 1995). However, primary teachers continued to communicate with one

another and work together more than in the past.

Positive parent involvement. The level of parent involvement has been highly varied among our study schools throughout the research period. Programs to acquaint parents with the primary program were held at all six schools the first year of implementation. Some schools instituted parent volunteer programs, and many primary teachers sent regular newsletters home to keep parents abreast of classroom activities. Initial efforts to get parents involved in the primary program have relaxed at all schools, but parent involvement efforts are generally higher now than they were pre-KERA.

Disjunction between primary program and intermediate grades. As mentioned previously, primary program implementation was hampered by the lack of clear linkages to the larger reform. This disconnect played out not only at the state level, where support materials and training for primary were developed separately from those for all grade levels, but also within local schools. Primary teachers were focused on the critical attributes, while teachers in the intermediate grades were focused on preparing students for the state assessment. Intermediate-grade teachers were themselves unclear on how to teach in ways that would help all of their students reach the demanding goals of the state assessment, but they did know that they had to help students develop portfolios and answer open-response questions, both key features of the state test. Because most elementary schools extend only up through fifth or sixth grade, and elementary students were administered the state assessment in grades four and five, the entire school was held accountable for these students' performance. The pressure of this accountability program led most intermediate-grade teachers to intensify the more traditional approaches rather than attempt new, untried, and unproven strategies in a high- stakes environment. Ideally, had the two groups of teachers come together with their concerns, primary teachers might have become more focused on KERA goals and expectations, and intermediate teachers might have looked to the primary to identify instructional practices that might help students acquire those goals. Instead, the two programs developed in relative isolation from one another. Primary teachers worked together to fashion programs that addressed the critical attributes, while intermediate teachers worked feverishly to prepare their students for the state assessment. As a result, it appeared that two separate reforms were underway in the study schools.

The split between the two programs was palpable, leading to resentment on both sides. Primary teachers were constantly given the message by intermediate grade teachers that the "cutesy" things they were doing in their classrooms were not preparing students for the rigorous expectations of fourth grade. Over time, rather than the primary program concept working its way up through the elementary school, pressure to prepare students for the state assessment program filtered down into the primary program. Primary teachers in the study schools were unsure how to incorporate rigorous content within the critical attributes of the primary program; and they had been given the message from intermediate teachers that the approaches they were using were NOT preparing students for the assessment. Therefore, instead of using the new approaches they had learned to teach to KERA goals, many primary teachers returned to the tried-and-true, scope-and-sequence curriculum materials to make sure they were covering all the content required to do well on the assessment.

Effects on Students

Studies of nongraded programs in other states and nations have generally shown that such programs do NOT negatively impact achievement, and sometimes have positive effects on non- cognitive measures such as improved student attitudes toward self, peers, and school (Lloyd, 1999; Miller, 1990; Pavan, 1992; Veenman, 1995). Determining achievement effects of Kentucky's primary program is difficult for at least three reasons: (1) the program was not fully implemented either in our study school or in most schools statewide (McIntyre & Kyle, 1997); (2) all Kentucky elementary schools were required to implement the primary program, so no control group of Kentucky students was available with which to compare achievement; and (3) there are no good baseline data with which to compare pre-KERA and post-KERA achievement. Most schools discontinued administering the CTBS for the

first few years after KERA was passed and when they resumed, a different version of the test was in place. With these provisos in mind, we will use the evidence that is available to conjecture about the effects of the changes that were implemented at the primary level.

Anecdotal evidence. As soon as the first group of primary students exited to fourth grade, we began to hear comparisons of them to previous fourth graders. Fourth- grade teachers reported that students coming out of the primary program were lacking basics skills, specifically in the areas of spelling and math facts. Some teachers also complained that students were unaccustomed to working alone because of being allowed to work with partners and help one another in the primary program. Another complaint was that, because primary teachers emphasized positive aspects of student work, students could not discern or did not care if they had done well or poorly on their work; for instance, believing that getting half of the answers correct on a test or exercise was good work.

To balance those complaints, parents and fourth grade teachers also told us that the exiting primary students were "better thinkers," asked more questions, and were better creative writers. Parents of randomly-selected students in the class of 2006 almost universally reported that their children enjoyed school, and had learned much more than the parents expected by the time the students reached fourth grade. Although some parents had initially been confused by the new system for reporting student progress and many still wished for letter grades, we did not see in any of our study districts a general uprising from parents against the primary program. By the time the class of 2006 had reached fourth grade, most of the parents we interviewed expressed satisfaction with the primary experience—although a few reported that some primary teachers had interpreted continuous progress to mean that children should be allowed to do only what they wished to do.

On a statewide survey conducted in 1999, school board members, principals, teachers, parents, and the general public were asked how well the primary program had worked to improve teaching and learning in local schools. Over 60 percent of school board members, educators, and parents serving on school councils believed the program had worked well. Over half of public school parents and the general public also believed the program had worked well; another 20-30 percent of these two groups reported that they did not know or were undecided. Less than one third of any group reported that the program had worked poorly (Kentucky Institute for Education Research, 1999).

Test scores. State assessment results suggest some positive outcomes of the primary program. Statewide, fourth-grade scores in all subject areas improved between 1993 and 1998, with the highest overall score and the greatest gains occurring in reading. NAEP scores have also improved at the fourth-grade level in reading and math, surpassing the national average in reading by 1998. On the CTBS/5 in 1999, exiting primary student scores had improved very slightly over the previous two years and were at or above the national average in all areas. While these scores alone may not be indicative of the primary program's effectiveness, given that our study and others cited previously indicate that many schools have not fully implemented the program, they suggest that at the very least, no harm has been done by the primary program.

McIntyre & Kyle (1997) reported that a study that compared student achievement on the state assessment to levels of primary implementation found no general pattern that linked the two (Hughes & Craig, 1994, as cited by McIntyre & Kyle). In our sample of six schools, three schools had consistently rising test scores—and relatively high scores—on the state assessment the first two accountability cycles (a period of four years). Of these three, two had maintained fairly traditional practices; the other was the one school that had most fully implemented the primary program. In the third cycle, however, one of the more traditional schools had declining scores, while the other had experienced a very small increase. Only the school that was most fully implementing the program continued to surpass the improvement goal set by the state. This school, where over 50 percent of the student body were from low-income families, also had the highest scores among our six study schools (see the Orange County Elementary School case study in Appendix A). While our study sample is too small to generalize these findings to the state, we might conjecture that schools implementing traditional

practices will reach a plateau on the state assessment, which is designed to measure higher order skills; and that more substantive changes are required if schools are to continue to improve on the state test. Further research is needed in this area.

Discussion

The above discussion illustrates the difficulties Kentucky experienced trying to move schools from a traditional graded approach to a continuous progress model. That schools should find it difficult to make this transition is hardly surprising, given that graded schooling has been a hallmark of formal education in this country for over 100 years. Studies of school reform have shown that graded instruction has been highly resistant to change over the years. Tyack and Tobin (1994) identify graded schools as part of a "grammar of schooling" that has remained remarkably stable over time. Similarly, Elmore (1996), Firestone, Mayrowetz & Fairman (1998), and Tyack and Cuban (1995), identify age and ability grouping as part of a core pattern of schooling that has historically proven highly resistant to change. Tyack and Tobin (1994) attribute the staying power of graded schooling (and other widely-accepted school structures) to the fact that this organizational form got in on the ground floor of organizational development of schools and thus, became institutionalized. They also note that inertia plays a role; and that familiar organizational structures such as graded schooling enable teachers to discharge their duties in predictable fashion: controlling student behavior, instructing heterogeneous populations, and sorting people for future roles in school and life. The historical record alone, then, suggests the monumental task that the Kentucky legislature undertook in attempting to replace grades K-3 with a nongraded structure. Our research, as well as other studies of Kentucky's primary program, adds Kentucky to the long list of places that have tried, somewhat unsuccessfully, to eliminate the graded structure of schooling.

What lessons might be learned from Kentucky's attempt at establishing a nongraded primary program? The first issue that must be considered is whether it is possible to mandate a change of this magnitude. National and international researchers who have studied and advocated for nongraded programs emphasize that nongradedness is a philosophy as much as a practice, and that only teachers with some commitment to the concept are likely to implement it with any success (Anderson, 1993; Goodlad & Anderson, 1987; Lloyd, 1999; Pavan, 1992).

In the face of such evidence, one wonders if states and localities might look at other ways to accomplish the goals of nongradedness. Lloyd (1999), who reviewed recent research on multi-age classes, poses this very question at the conclusion of his review: is the multi-age structure a necessary condition for delivery of developmentally appropriate curriculum, or would it be more fruitful to ensure that teachers of single-grade classrooms adopt the practices of good multi-age teachers, such as a focus on diversity/individual differences and continuous progress, differentiated instruction and developmentally appropriate curriculum, curriculum which can be engaged at different levels of complexity, flexible grouping, and collaborative learning?

In Kentucky, the vision for the entire reform was to create a system in which *all* students at all grade levels, through varied instructional approaches and continuous assessment of progress, would be helped to achieve challenging standards. While nongradedness seems a very rational means to accomplishing this goal, mandating such a program ran counter to the reform's overall philosophy of allowing schools to determine how to help students achieve KERA goals. In addition, research has demonstrated the intractability of the concept of graded instruction. Given that the desire in Kentucky and many other states and localities is to restructure educational systems so that all students can achieve at high levels without being stigmatized if they fail to do so in prescribed ways and on a prescribed schedule, resources might be better directed toward professional development and technical assistance on teaching challenging content to all students through diverse instructional strategies, rather than on mandating nongradedness for its own sake.

Yet, Lloyd (1999) asserts that the very fact that age-related assumptions about development are resistant to widespread change is a rationale for implementing nongraded programs. The multi-age structure itself is more likely to offer the

perceived benefits than are single-grade classrooms. In Kentucky, it was this sort of thinking that led to including the primary program in the reform package in the first place. This was a way to jump-start a reform that was meant to change teacher beliefs about who can learn, what they can learn, and how they can learn it.

While there have clearly been problems mandating this sort of sweeping change, we are unprepared to say that Kentucky's nongraded primary program should not have been attempted, or should be abandoned at this juncture. We have seen that instructional change aimed at meeting students' individual needs has been more widespread in the primary grades than at other levels of the system. Available achievement data shows that achievement for students who have been through the primary program has improved in some areas, while remaining stable in others. In addition, we have anecdotal evidence that the primary program has improved student motivation and attitudes toward schooling, as well as their creativity and thinking skills.

A great deal of time and energy has been expended in Kentucky on implementing both the primary program and the larger reform. Rather than disrupt the reform process and risk sending the message that the goals of the primary have been abandoned, the most prudent approach for Kentucky policymakers at this point is to work toward linking the primary program approach with the overall goals of KERA. The first step in this process would be to send clear, highly visible messages to schools that the primary program is still in place. Second, the overall goals of the primary program must be made clear. Fullan and Stiegelbauer (1991) argue that the crux of change involves the development of meaning in relation to a new program. In Kentucky, a basic problem that plagued implementation of the primary program from the beginning was that its meaning was unclear to teachers. In articulating the program's overall purpose, the link to overall KERA goals must be established. It should be made clear that the purpose of the primary program is to enable *all* students to progress continuously toward acquisition of KERA goals. Linkages need to be made between support systems and implementation documents such as the KELP, which helps establish whether primary students are ready to move on to the fourth grade, and the *Core Content for Assessment* (Kentucky Department of Education, 1996a), which defines the content on which fourth-graders will be tested.

Finally, Kentucky policymakers should accept (as they have been doing all along) variations on the primary program concept. The graded structure may never be entirely eliminated, but if implementation of the primary program leads teachers to move closer to a continuous progress model that enables all students to achieve the reform goals in ways that are appropriate to them, then the program will have been a success.

Note

This publication is based on work sponsored wholly or in part by the Office of Educational Research and Improvement, U. S. Department of Education, under contract number RJ96006001. Its contents do not necessarily reflect the views of OERI, the Department, or any other agency of the U. S. Government. This publication is based on work sponsored wholly or in part by the Office of Educational Research and Improvement, U. S. Department of Education, under contract number RJ96006001. Its contents do not necessarily reflect the views of OERI, the Department, or any other agency of the U. S. Government. AEL is an Equal Opportunity/Affirmative Action Employer. AEL's mission is to link the knowledge from research with the wisdom from practice to improve teaching and learning. AEL serves as the Regional Educational Laboratory for Kentucky, Tennessee, Virginia, and West Virginia. For these same four states, it operates both a Regional Technology in Education Consortium and the Eisenhower Regional Consortium for Mathematics and Science Education. In addition, it serves as the Region IV Comprehensive Technical Assistance Center and operates the ERIC Clearinghouse on Rural Education and Small Schools. Information about AEL projects, programs, and services is available by writing or calling AEL.
AEL, Inc.
Post Office Box 1348Charleston, West Virginia 25325-1348

304/347-0400
800/624-9120
304/347-0487 (Fax)
aelinfo@ael.org
http://www.ael.org

References

Anderson, R.H. (1993). The return of the nongraded classroom. *Principal, 72*(3), 9-12.

Anderson, R.H. & Pavan, B.N. (1993). *Nongradedness: Helping it to happen.* Lancaster, PA: Technomic.

Bridge, C. A. (1995). *The Implementation of Kentucky's Primary Program 1995: A progress report.* Lexington, KY: Institute on Education Reform, University of Kentucky.

Cohen, David K. (1995). What is the system in systemic reform? *Educational Researcher*, 24 (9), 11-17, 31.

Elmore, Richard F. (1996). Getting to scale with good educational practice. *Harvard Educational Review*, 66(1), 1-26.

Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis*, *20* (2), 95- 113.

Foster, Jack D. (1999). *Redesigning Public Education: The Kentucky Experience.* Lexington, Kentucky: Diversified Services, Inc.

Fuhrman, S. H., Elmore, R. F., & Massell, D. (1993). School reform in the United States: Putting it into context. In S. L. Jacobson & R. Berne (Eds.), *Reforming education: The emerging systemic approach*, pp. 3-27. Thousand Oaks, CA: Corwin Press.

Fullan, M.G. (February, 1996). Turning systemic thinking on its head. *Phi Delta Kappan*, 420-423.

Fullan, M. G. & Stiegelbauer, S. (1991). *The new meaning of educational change.* Second edition. New York: Teachers College Press.

Goodlad, J.I. & Anderson, R.H. (1987). *The nongraded elementary school.* New York: Teachers College Press.

Hornbeck, D. W. (1990, February 23). "Recommendations related to Curriculum (Adopted by the Task Force on Education Reform 2/23/90)." Frankfort, KY: Legislative Research Commission.

Hughes, K., & Craig, J. R. (1994). *Using performance assessment achievement data to evaluate a primary instructional program.* Paper presented at the annual meeting of the American Evaluation Association, Boston, MA.

Kentucky Department of Education (1991). *Kentucky's primary school: The wonder years.* Program Description I. Frankfort, KY: author.

Kentucky Department of Education (1993a). *State regulations and recommended best practices for Kentucky's primary program.* Frankfort, KY: Author.

Kentucky Department of Education (1993b). *Transformations: Kentucky's curriculum framework.* Frankfort, KY: Author.

Kentucky Department of Education (1994, June). *Kentucky Early Learning Profile: A part of the Kentucky primary system: Teacher's handbook.* Frankfort, KY: Author.

Kentucky Department of Education (1996a). *Core content for assessment, Version 1.0.* Frankfort, KY: Author.

Kentucky Department of Education (1999, August). Kentucky Department of Education, 1998-99 primary program: Response to data request from the Office of Education Accountability, August 10, 1999. Frankfort, KY: Author.

Kentucky Education Association/Appalachia Educational Laboratory (1991, April). *Ungraded primary programs: Steps toward developmentally appropriate instruction.* Charleston, WV: Appalachia Educational Laboratory.

Kentucky Institute for Education Research. (1999, August). *1999 Statewide education reform study, final report.* Bowling Green, KY: College of Education and Behavioral Sciences, Western Kentucky University.

Kyle, D. W. & McIntyre, E. (1993). *Developing and implementing non-graded primary programs: Experiences from the primary year.* Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA.

Lloyd, L. (1999). Multi-age classes and high ability students. *Review of Educational Research, 69* (2), 187-212.

Massachusetts Board of Education (1990). *Structuring schools for student success: A focus on grade retention.* Boston, MA: Author.

McIntyre, E. & Kyle, D. (1997). Primary program. In J. C. Lindle, J. M. Petrosko, & R. S. Pankratz (Eds.), *1996 review of research on the Kentucky Education Reform Act*, pp. 119-142. Frankfort, KY: Kentucky Institute for Education Research.

Metz, M. H. (1990). Real school: A universal drama amid disparate experience. In D. E. Mitchell & M. E. Goertz (Eds.), *Education politics for the new century: The twentieth anniversary yearbook of the Politics of Education Association*, pp. 75-91. London: The Falmer Press.

Miller, B.A. (1990). A review of the quantitative research on multigrade instruction. *Research in Rural Education*, 7(1), 1-8.

Murphy, J. (1990). The educational reform movement of the 1980s: A comprehensive analysis. In J. Murphy (Ed.), *The education reform movement of the 1980s: Perspectives and cases* (pp. 3- 55). Berkeley, CA: McCutchan Publishing Corporation.

National Association for the Education of Young Children (1987). 1987 NAEYC position statement on developmentally appropriate practice in the primary grades, serving 5- through 8-year olds. *Young Children*, 43(2), 64-68, 81-84.

National Association of Elementary School Principals (1990). *Early childhood education and the elementary school principal.* Alexandria, VA: Author.

O'Day, J. A. & Smith, M. S. (1993). Systemic reform and educational opportunity. In Fuhrman, S. H. (Ed.), *Designing coherent education policy: Improving the system*, 250-312. San Francisco, CA: Jossey-Bass Publishers.

Office of Education Accountability (1996). *Annual report.* Frankfort, KY: Kentucky General Assembly, Office of Education Accountability.

Office of Education Accountability (1997). *Annual report.* Frankfort, KY: Kentucky

General Assembly, Office of Education Accountability.

Pavan, B.N. (1992). The benefits of nongraded schools. *Educational Leadership, 50*(2), 22-25.

Raths, J. & Fanning, J. (1993). *Primary program reform in Kentucky revisited.* Report to the Prichard Committee. Lexington, KY: Prichard Committee for Academic Excellence.

Raths, J., Katz, L. & Fanning, J. (1992). *The status of primary school reform in Kentucky and its implications: Report to the Prichard Committee*. Lexington, KY: Prichard Committee for Academic Excellence.

Rose v. Council for Better Educ., 790 S. W. 2d 186, 60 Education Law Reporter 1289 (KY 1989).

Schwartz, H. (1991). Putting the pieces together: A systemic approach to educational reform. *Planning and Changing, 22* (3/4), 231-239.

Smith, M.S., & O'Day, J. (1991). Systemic school reform. In S.H. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing: The 1990 Politics of Education Association yearbook* (pp. 233- 267). Bristol, PA: Falmer.

Tyack, D. B. (1974). *The one best system: A history of American urban education.* Cambridge, MA: Harvard University Press.

Tyack, D., & Cuban, L. (1995). *Tinkering toward Utopia: A century of public school reform.* Cambridge, MA: Harvard University Press.

Tyack, D. and Tobin, W. (1994). The "grammar" of schooling: Why has it been so hard to change? *American Educational Research Journal, 31*(3), 453-480.

Veenman, S. (1995). Cognitive and noncognitive effects of multigrade and multi-age classes: A best-evidence synthesis. *Review of Educational Research*, 65, 319-381.

Wilkinson, W. G. (1988a) *A plan to restructure schools in Kentucky.* Frankfort, KY: Office of the Governor.

Wilkinson, W. G. (1988b) *Q & A: Improving Kentucky's schools: A conversation with Governor Wallace G. Wilkinson.* Frankfort, KY: Office of the Governor.

About the Authors

Patricia J. Kannapel, Research and Development Specialist for AEL, Inc. (Charleston, WV), has been co-director of AEL's ten-year study of the implementation of statewide reform four rural Kentucky school districts. She holds masters' degrees in education and anthropology from the University of Louisville and the University of Kentucky. Patty is currently a doctoral student in applied anthropology at the University of Kentucky.

Lola Aagaard was a researcher with AEL's study of the Kentucky Education Reform Act for eight years of its ten- year span. Her Ph.D. is in adult and community education (with an emphasis in research methods and data analysis) from the University of Oklahoma. Lola's eclectic academic history also includes degrees in nursing, biology, and public health.
Email: lboram@mis.net

Pamelia Coe has been principal investigator of AEL's ten- year study of the implementation of the Kentucky Education Reform Act. She obtained her masters' degree in cultural anthropology from Columbia University (NY) and her Ph.D. in foundations of education (specializing in anthropology of education) from Michigan

State University. Prior to becoming involved in educational research, Pam worked in community development with the American Friends Service Committee, mostly with American Indian groups.

Cynthia A. Reeves spent four years as a member of AEL's research team studying the implementation of statewide reform in four rural Kentucky school districts. She holds master's degrees in economic development and applied anthropology from the University of Kentucky. Cindy is currently working on her dissertation in applied anthropology at the University of Kentucky.

Appendix A
Case Studies of the Primary Program

Overview

These case studies illustrate the ways in which local factors influenced the implementation of the primary program, whether towards greater or lesser conformity with the mandate. The descriptions of these schools also portray the wide range of practices that are taking place under the "primary" umbrella in Kentucky. The schools profiled here are normal schools—neither the worst nor the best that Kentucky has to offer. Their responses to the primary program mandate ranged from grudging implementation of the least they thought they could get by with to enthusiastic acceptance and nearly full implementation.

Newtown Elementary School —"Tradition, Tradition!"

Overview. The local factor that most heavily influenced the development of the primary program at Newtown Elementary School (NES) was a longstanding tradition of excellence in education, as evidenced by some of the highest standardized test scores in the state and a college attendance rate of over 90 percent. This tradition reinforced teachers' deeply felt belief in the value of the rigorous traditional program the school provided. In addition, strong parental involvement and teachers' feelings of empowerment created a very positive school climate. When the school earned rewards after the first biennium of KIRIS testing, these factors were reinforced and there was even less incentive for change than there had been originally.

NES is located in a small town, which has had its own independent school district since the early years of the century. Newtown prides itself on raising enough local tax revenue to support a highly successful school system, whose students have outperformed those in any of the nearby rural county districts. Parents have traditionally been highly invested in their children's education, and middle class families from a number of nearby districts have paid tuition to send their children to the independent district.

History of the primary program. The principal who was at the school when the program was being developed encouraged teachers and parents to take leadership and gave them unstinted support. Planning for the primary program was accomplished mostly through the efforts of one or two enthusiastic teachers, who were interested in receiving additional training to implement the new program. Most of the faculty remained skeptical of the mandated changes.

The initial NES primary program plan specified three-year, multi-age classrooms, with a separate kindergarten program. Primary teachers had access to a broad spectrum of training opportunities, but not all availed themselves of the full range. Teachers and students were divided into multi-year primary families, with groups of teachers sharing students. Students studied reading and math in skill groups (largely single age) but were taught "themes" (usually science and social studies) in the multi-age setting. Teachers reported that it was difficult to keep the attention of and involve students across such a wide age range.

The first year of implementation, some teachers continued to use mostly traditional methods, but supplemented them with some new approaches, including centers, sustained silent reading, journal writing, and some hands-on math and science projects. Nearly all teachers rearranged their classrooms so that desks were in clusters or students seated around tables rather than in straight rows facing front. Many engaged in joint planning with one another. Some teachers shelved their

textbooks and taught thematically.

Teachers struggled with anecdotal records, but many began ensuring that primary students kept portfolios of work. (The content of the portfolios and the number of pieces of work varied from teacher to teacher.) Student progress was reported on a skills checklist with a narrative section rather than a traditional report card. Parents lamented the elimination of letter grades and reported that neither they nor their children could tell from the progress reports just how the students were doing.

The multi- year families at Newtown Elementary changed quickly to dual- age self-contained classrooms, and later they changed again to essentially single-age units. The dual-age rooms, in some cases, were taught as split classes with little mixing of the two age groups for instructional purposes. Joint planning decreased to cooperation among grade-level teachers with the exception of planning for periodic schoolwide themes.

Instruction remained largely traditional with a skills emphasis. Even so, teachers at higher grade levels reported that some primary students were advancing to the upper grades without the necessary proficiencies. Soon, even teachers who had enthusiastically embraced new methods returned to stressing skills either on their own or as a result of encouragement from others. Textbooks, worksheets, phonics workbooks, and spelling books were very much in evidence. Some teachers, especially at the third grade level, opted to give number or letter grades on student work.

These traditional approaches were reinforced when the KIRIS results began coming in: the school earned rewards in the first two accountability cycles. The success of the "tried and true" methods convinced school personnel that they were on the right track and should persevere. Most parents were very pleased with the school's approach; they had been uncomfortable with the year or two of cautious experimentation that followed the initial primary implementation.

Status of the primary program at the end of the 1996-97 school year. Newtown Elementary had retained some of the new strategies encouraged by the primary program. Teachers reported that primary students were writing more than in the past. Students worked in groups more than they did before KERA, according to the principal. Hands-on math and science have proven helpful and interesting for most teachers and students, although the extent to which these approaches were used varied by teacher. Teachers were conscious of the individual skill levels of students and tried to take them into account. Some teachers grouped students by skill level for reading or math instruction. Others gave whole class instruction in the basic subject areas but required less of students who had lower skill levels.

The school personnel seemed comfortable with their approach in the primary program, and there was no sense of movement toward more or less implementation. Throughout the school's implementation of KERA, the faculty was confident that NES students would be successful on the statewide assessment and that the school will continue to be recognized as one of the most academically rigorous and successful schools in the area.

Summary. NES was proud of its primary program before KERA was passed. The faculty has used the training made available as the KERA primary program was implemented to increase their repertoire of techniques and materials, and they have made some lasting changes, such as increasing the amount of writing done by primary students. But, for the most part, they have approached change with great caution. Their KIRIS scores—like their previous scores on standardized tests— have been high enough to convince them that their approach was correct and that their traditionally high academic standards will be maintained.

Kessinger Elementary School—"The Need for Leadership"

Overview. The factors that appeared to most strongly influence the evolution of the primary program at Kessinger Elementary were local ones: leadership, teacher beliefs, and school climate. Interestingly, many primary teachers at Kessinger appeared to grasp the intent of the primary program and to agree with the overall philosophy of allowing students to progress at their own rate through an instructional program geared to the needs of young learners. The primary program

might have been implemented in a consistent direction at Kessinger had the faculty been able to pull together toward a common vision. But the opportunity to do so was impeded by frequent changes in principals, as well as a longstanding lack of cohesiveness among the teachers. Differing philosophies among teachers that had been largely dormant pre-KERA—when teachers had the freedom to teach as they saw fit within their own classrooms—were brought to the forefront when the faculty was called upon to create a coherent primary program.

History of the primary program. Kessinger Elementary is located in a small, rural county where the economy is based largely on agriculture. In spite of an increase in the local tax rate and more state funding after KERA was passed, the district continues to struggle financially because of lack of industry and tourism in the county. There is a great deal of turnover in school and district leadership, in part because the district pays lower administrator salaries than surrounding districts. Kessinger has had five principals in the eight years since the passage of KERA.

When KERA passed, Kessinger teachers exhibited varying degrees of enthusiasm for the nongraded primary program. Generally, primary teachers were willing to give the program a try and planned to implement it as specified by state guidelines. Some teachers, however, found that the primary philosophy fit their own belief systems very well and were eager to begin implementation, while others were skeptical and wanted to proceed more slowly. These different viewpoints exacerbated existing tensions among the faculty. The principal was uncomfortable with the conflict that arose from trying to arrive at a common vision for the program. When differences of opinion surfaced at the first meeting to plan the primary program, the principal delayed the planning process to provide a cooling-off period. Instead, the controversy heated up.

By 1992–93, Kessinger teachers had been unable to agree on a primary configuration, so they implemented two different approaches. One team of teachers implemented a K–3 arrangement at one end of the hall, while another team implemented a dual–age arrangement (K–1, 1–2, and 2–3) at the other end. Neither team had common planning time with their colleagues, and teachers on both teams reported at mid–year that they were exhausted and frustrated from trying to implement new instructional programs without support or time to interact with their peers. Teachers on both teams tried different strategies for student grouping but were unable to settle on a strategy satisfactory to all. By the end of the year, teachers on the K-3 team began to differ among themselves, with some supporting the K–3 arrangement, others favoring a dual–age configuration, and others coming to believe that single–grading was desirable. There did not seem to be a strategy for teachers to meet and try to reach consensus on a unified approach.

In 1993–94, the frustration and confusion regarding the Kessinger primary program reached a peak. Teachers still had not agreed on the appropriate configuration, and a new source of conflict arose when some teachers began to push to exclude kindergarten students from the program. Teachers moved kindergarten in and out of the program during the school year, shifting students among teachers. A parent complained that her child changed classes four times during the year as the teachers wavered on kindergarten inclusion. Another parent described the primary program as "a mess," and reported that the two factions of primary teachers were constantly bickering. The teachers themselves contemplated having a "negotiator" from the state department come talk to them.

After the 1993-94 school year, the Kessinger principal opted to return to the classroom. The SBDM council hired a principal from outside the district who initiated and supported a move to dual–age classrooms with some ability grouping for skills. The primary configuration at Kessinger in 1994–95 was K–1, 1–2, and 2–3. Teachers kept their students in dual–age groups for a period of time each day, but students spent the bulk of the day in ability groups, mostly by grade. The disagreement over kindergarten inclusion in the primary program continued.

This second (since our study began) principal resigned for a better offer in another district at the end of 1994–95. The SBDM council, on a split vote with no principal yet on board, voted to switch to a single–grade configuration the following year. The move was supported by intermediate- grade teachers, as well as some parents. The council subsequently hired a new principal, who set out to support the program that was already in place. She divided Kessinger teachers into single–grade

teams and, for the first time, teams were given common planning time. Although teachers appeared to get along better, there were signs that factionalism continued. The principal reported that they were still "fighting the battle" in the school and with the community about what was expected of multi-age classrooms. A veteran faculty member reported that KERA had divided the school into "for" and "against" factions, and that teachers wasted a lot of time pulling in different directions and trying to win support for their views.

Status of the primary program at the end of the 1996-97 school year. At the end of 1995–96, the third principal resigned to return to her home county. A new principal was hired and set about to bring the primary program "into compliance" with state requirements in 1996-97. This fourth principal, however, came on too strong for some teachers and was unable to intervene successfully. She attributed the problems in the primary to the lack of continuity in leadership. She said she had tried to help with this, but conceded that "there are times when my vision impedes the process." At the end of the school year, she resigned because she did not feel she had sufficient support to be an effective leader.

The ongoing turmoil at Kessinger had considerably less detrimental effect on the primary program in particular and instruction in general than one might expect. In fact, Kessinger earned rewards in the second accountability cycle (1994-95 and 1995-96). By 1996-97, Kessinger primary teachers, as a group, did not seem to have been defeated by the conflict that had become a way of life at the school. Classroom observations at Kessinger revealed that very little instructional time was wasted, and that teachers were generally focused on helping students succeed. The majority of Kessinger primary teachers continued to implement many practices consistent with the primary philosophy. Many struggled within the single-grade structure to manage a continuous progress model in their classrooms or exchanged students with other teachers. For instance, at least two teachers within their own classrooms established individualized reading programs for students. Two teachers of different grade levels combined their classes three times a week to teach science, planning units together after school and on weekends.

Teachers who supported fuller implementation of the primary program were not vocal in their support, but seemed to have decided that the best way to manage the situation was to try to do what they thought best for students within their own classrooms or in conjunction with another, like-minded teacher. Teachers who opposed the primary program were more vocal. Generally, the KES teachers we interviewed and observed, whether they supported the primary concept or not, seemed to be conscientious and devoted to helping students learn. The two factions of teachers had simply been unable to arrive at a meeting of the minds with regard to the primary program. Those who opposed the program, including some parents, were more vocal and influential than supporters. The latter group continued to support the primary program and implement it to the best of their ability within a structure that was not conducive to the primary concept.

Summary. The Kessinger case illustrates how inconsistencies in leadership can seriously impede a school's progress, particularly in a school where a faculty that lacks cohesiveness is called on to make major programmatic and instructional changes. In the early stages of primary program implementation, teachers were mostly left on their own to work out their differences. At that time, most of the teachers were willing to at least give the program a try, although there were varying levels of enthusiasm. When things did not go well at first, teachers had only their own belief systems and past experience to fall back on in knowing what to do next. Those who had been skeptical about the program returned to practices with which they had been successful previously. Those who supported the philosophy forged on, thus widening the chasm between the two camps of teachers. By the time a principal was hired who understood and supported the primary program philosophy, the factions were well-entrenched and difficult to bring together. The constant change in leadership since that time has made the problem worse. By the time each new principal had begun to grasp the nature of the problem, the year was nearly over and then the principal moved on to another job. The situation will not be easily resolved under any circumstances, but there is a desperate need for continuity in leadership in order to get the primary program and the school on track.

The future of the primary program at Kessinger is uncertain. At the time of this

writing, the Kessinger SBDM council had hired a new principal, this time someone from within the district. The primary program has switched to a K, 1-2, 3 configuration in an attempt to bring the program into "compliance." It remains to be seen what role the fifth principal will play in shaping the direction of the primary program. Because she has several years of experience in the school district, she may have greater insight into the problems going in than have previous principals. Whether her familiarity with Kessinger and its teachers will be an asset or a liability depends not only on her ability to bring the faculty together, but on the teachers' own willingness to trust one another enough to ignore past differences and make another attempt at developing a common vision for students.

Vanderbilt County Elementary School—"Why Are We Doing This?"

Overview. Vanderbilt County Elementary School (VCES) illustrates, perhaps more than any school in our study, how the combination of state and local factors can influence primary program implementation. One of the most central factors at VCES was the lack of a shared philosophy among the faculty with regard to the primary program. The school had previously been traditional in its approach and had done well on standardized tests using this approach. KERA and a new principal arrived at the school nearly simultaneously, however, and it seemed that a new day had dawned at VCES. VCES teachers were initially willing to suspend disbelief and implement new programs and strategies at the principal's urging. Some primary teachers were enthused about the changes but many were skeptical, perhaps because of their previous success using more traditional methods. When the first round of KIRIS results was released and VCES had not met its threshold, the teachers began retreating from primary program implementation. As a result, a school that initially made many changes in its approach to primary instruction returned to a program that closely resembled pre–KERA practices.

History of the primary program. VCES is located in the county seat of a rural, agricultural community. The new principal, hired in 1991 by the newly-formed SBDM council, greatly supported the concepts embedded in KERA and set about to put the school on a new path. Early reports from teachers were mostly complimentary; they appreciated the principal's energy, enthusiasm, and aggressiveness in seeking resources and opportunities for them to get the training they needed to implement KERA.

The central office, too, was relatively pro–active in preparing teachers to implement the primary program, and several years of sound fiscal management enabled the district to provide substantial professional development to primary teachers. VCES teachers availed themselves of these opportunities more than teachers at other schools in the district, largely owing to the principal's encouragement, support, and initiative in locating additional time and resources for teacher training. Primary teachers were appreciative of the resources and training available to them, and most of them made many changes during initial implementation of the primary program.

At that time, the focus appeared to be heavily on implementation of the primary program critical attributes. VCES primary teachers changed their instructional and assessment approaches substantially, but did not express a strong sense of the overall purpose of the primary program. Many VCES teachers were especially skeptical of the multi-age requirement. The school was cautious in implementing a multi-age program, never going beyond a dual-age arrangement. During the first year of implementation, half of the primary teachers had dual-age classrooms all day, while the other half had dual-age groups for an hour daily. Kindergarten teachers incorporated their students into the program 90 minutes weekly. Teachers with full-day dual-age classrooms paired with another teacher for "skills grouping" in math and sometimes reading: the teachers grouped students according to their skill level, with one teacher taking the "high" group and another the lower group. Teachers were required by the principal that year to submit evidence of flexible grouping and regrouping of students. Teachers were provided with planning days and used these to collaborate with colleagues. Collaboration tended to be dual-grade rather than across the primary. Many teachers were systematic about keeping anecdotal records on students.

In 1993-94, VCES primary teachers configured their program with a variety of dual- grade arrangements: K-1, 1-2, and 2-3. In addition, two self-contained kindergarten rooms were in place for parents who preferred that option. Primary teachers generally felt that a wider age span would be too difficult to manage. Some teachers said they would prefer to return to a single–grade approach. Even with dual–age classrooms, VCES primary teachers reported that they did not keep the same students from one year to the next so that no teacher would have the same problem students each year. Primary teachers continued to use many of the new instructional approaches they had learned about.

In 1994–95, all VCES classrooms were configured as either K/1 or 2/3. Teachers worked in teams of two or three within their grade groups (teams were either K/1 or 2/3, but there was not a mix) to do skills grouping each morning for language arts and math instruction. The skills groups were largely single-grade groups, but some students crossed the grade boundary as needed. That same year, KIRIS results for the first biennium were released. Within the school district, other elementary schools that had not made as many changes as VCES scored high enough to earn rewards. VCES scores improved but the school did not meet its goal. Many teachers at VCES and throughout the district interpreted this as a sign that VCES had gone too far in throwing out tried–and–true methods. Teachers who had tried to follow the course the principal had set for the school began to question this course. The principal began to give teachers more freedom to find approaches with which they were comfortable.

The dual-age approach continued in 1995-96, but more and more teachers reported dissatisfaction with this arrangement; they expressed a desire to return to single-grade classrooms. Teachers began to incorporate some of the more traditional approaches back into their classrooms, such as using basal readers and teaching spelling and phonics as separate subjects. Teachers reported that they felt less pressure now to use only the newer methods, perhaps because the assessment results had given more credence to the argument that the new approaches were not effective. Teachers also began to back away from authentic assessment techniques. One of the changes teachers had made—collaboration with special teachers—increased in response to KIRIS results, as the school began to use Title I teachers as math and science specialists to help teachers plan hands-on activities in their classrooms.

Status of the primary program at the end of the 1996-97 school year. The VCES principal, who initially made a strong effort to get the primary program moving in a consistent direction, changed strategy after the first round of test scores were released. In 1996-97 when the primary teachers expressed a strong desire to return to a single-grade configuration, the principal insisted they clear this through the state department of education. When officials at the state department assured them that they could have single-grade homerooms with the understanding that students would be moved around during the day according to individual needs, the teachers moved to a single-grade arrangement without overt opposition from the principal. For the most part, VCES primary teachers appeared to have opted for a more traditional approach, placing students in single-grade classrooms and grouping them mostly by ability in relatively stable groups.

With the principal now giving the teachers more freedom in choosing instructional strategies, each primary teacher began implementing the program as she saw fit, resulting in approaches that varied from one classroom to the next. The majority of primary teachers expressed support for the single-grade approach, and several professed a belief that VCES teachers had thrown out too much initially and needed to return more to "the basics." Veteran primary teachers appeared to have reinstated the more traditional approaches. Younger teachers used more variety in their approaches, continuing to do some whole language, cooperative learning, hands–on activities, and centers.

Summary. The VCES case illustrates how an educational innovation can go awry when teachers do not see promising results after being obliged to make a change with which they do not agree and whose purpose they may not understand. VCES teachers were given ample professional development aimed at helping them implement the critical attributes, but they seemed to view the attributes as ends in themselves, rather than as means to an end. The principal, who seemed to grasp the

BEST COPY AVAILABLE

purpose of the primary program and felt implementation of the critical attributes was essential to achieving the goals of the program, hoped that the extensive professional development VCES teachers received would bring them on board in implementing the program. Whether this happened or not, however, the principal felt responsible for making sure the state–mandated primary program was implemented, and this was accomplished by a strong focus on process over content. As time went on and test results came in, however, the principal gave teachers more freedom in the classroom in the hope that, once they were comfortable that they were covering the necessary content, they would begin to incorporate strategies that enabled students with different learning styles to acquire the necessary knowledge and skills. It is too soon to tell what will become of the VCES primary program. In one sense, it might appear that KIRIS scores interrupted the reform process at VCES. However, if the principal and teachers can continue working toward an approach that successfully combines the teachers' expertise on what it takes to help students acquire basic skills with the principal's understanding of instructional strategies that enable all students to have success, then KIRIS results may have been just the impetus the school needed to get everyone moving in a common direction.

Orange County Elementary School—"Change and Change Again"

Overview. At Orange County Elementary School (OCES), local factors facilitated the development of the most fully fleshed out primary program implementation we observed. There was a strong principal, teachers who trusted the principal and accepted her leadership, and a district ethic of openness to educational improvements. During primary program implementation, the school moved into a new building designed to encourage flexible grouping and regrouping of students and professional teamwork among the faculty. School climate is positive, and the faculty is developing a common, child-centered vision. When the first KIRIS results were reported, the school had the largest gains of any elementary school in the district, and OCES earned rewards after the second biennium also. The faculty prided itself on what the school had been able to accomplish.

In spite of success on KIRIS while implementing a relatively innovative primary program, OCES educators became fearful that they could not continue improving without increasing the fit between the primary program and the KIRIS-driven upper elementary grades. Their solution was to combine third and fourth grades in a large open-space classroom. This combination resulted in a return to more traditional forms of instruction at the upper primary level, although continuous progress and other aspects of the primary program were still emphasized.

History of the primary program. OCES is located in a large, rural, eastern Kentucky county school district. A new principal, who provided vigorous leadership, came to the school shortly before KERA went into effect. Some of the faculty were initially leery of the new principal's strong advocacy of the nongraded primary program and research-based curriculum innovations, but the principal won their support by demonstrating respect for their professional opinions and decisions. From the beginning, teachers have been child- oriented; they are determined to make sure their students, mostly from non-advantaged backgrounds, have the opportunity to achieve at high levels. Leadership from the principal and an active school counselor have reinforced the focus on the whole child. The school has the feel of a large extended family, with cooks, instructional aides, and students, as well as teachers and administrators, taking responsibility for the student body.

The OCES primary committee, consisting of the principal, counselor, and all K-3 teachers, developed and implemented a plan in which children aged 5-9 worked together in multi-age home bases for several hours a day. Students worked on academic subjects in somewhat flexible skill groups for the balance of the day. Special education children were fully integrated into these families. The plan resulted in frequent movement in the halls as children moved from room to room in order to change skill groups. One primary family was able to use a different strategy, however. There was one large, open-space classroom that was able to accommodate four teachers and almost 100 children. This arrangement facilitated teacher collaboration and more flexible grouping and regrouping than was possible in the other families.

The primary teachers received a great deal of training in innovative curricula and strategies, especially during the planning year (1991-92) and the first year of

program implementation (1992-93). The primary teachers met as a group occasionally, and each family of teachers had common planning time scheduled daily, when they jointly planned interdisciplinary themes or units—usually taught during multi-age, multi-ability "theme time" in the afternoon, after the academic subjects had been covered.

Although the OCES primary teachers made a concerted effort to implement the critical attributes, they had difficulties that brought about an "implementation slump" during the third and fourth years of implementation. Even with common planning time, teachers never had enough time to do all they had to do, and they reported their personal life suffered. Parental participation, which was high during the first two years of the program, waned, and collaboration among the teachers in each family grew less intense. Teachers began using the common planning time for individual planning.

As primary students began entering fourth grade, the upper elementary teachers compared them with previous classes. They reported that the children were more creative and better at problem solving than previous classes, and less fearful of speaking in public, but that they were less disciplined and were often unwilling to sit quietly and work at their desks.

When the school moved to the new facility, most primary children were housed in large open rooms, as had proved so successful for one primary family during the first two years of the program. One family shared two smaller rooms. Another change for the primary was a district requirement that they use the full Kentucky Early Learning Profile (KELP) for recordkeeping and reporting to parents. While some teachers complained bitterly about the amount of time and paperwork required by KELP, they also said that it enabled them to know their students and understand their achievement better than they ever had before.

In 1996-97, the primary configuration was changed from K-3 families to two K-2 primary families and one large family combining Grades 3 and 4. There were five teachers and approximately 100 children in the classroom housing Grades 3 and 4. The rationale for this move was to ease the transition from the primary program to fourth grade in both academics and deportment.

The upper primary teachers responded to the pressure to prepare students for the academic rigors of KIRIS with a renewed emphasis on skills. They used basal readers and textbooks freely, following them closely in some cases and using them as resources in others. Instruction was less thematic, although science and social studies were still taught as units. Students did participate in a number of hands-on science projects.

The upper primary teachers incorporated continuous progress into basic skill areas. For a number of years every student in the school has taken a basic skills test each year to make sure that those skills were not being neglected. Beginning in 1996-97, the teachers in the third-fourth grade classroom assessed all students in both grades on math and reading skills and used the results—as well as their observation of student skills—to assign students to flexible skill groups. At the end of each unit or chapter, students were shifted to other groups or new groups were composed, based on student progress. Thus, in a skill group focused on multiplication, some students might be assigned to a group reviewing place value, while others were considered ready to move on to division. Reading groups were shuffled less frequently than math groups.

Status of the primary program at the end of the 1996-97 school year. The K-2 classrooms at OCES were still organized around the seven critical attributes of the primary program; however, the final year of primary was focused on preparing students to succeed on KIRIS. The program in upper primary incorporated continuous progress in the basic tool subjects, especially mathematics, as part of this strategy. It is likely that the OCES primary program will continue to change in response to local pressures, including those of KIRIS preparation, perhaps by holding the younger primary students to increased academic expectations.

Summary. OCES illustrates how local factors, including a felt need to improve local education, can lead a faculty to implement the nongraded primary program wholeheartedly and how their response to state factors (KIRIS preparation) can influence the direction of change. Orange County educators were committed to change because they wanted their students to achieve. Several factors came together

in a timely way to persuade teachers that the primary program was a step in the right direction. Subsequently, educators at the school came to believe that the disjunction between the primary program and the intermediate grades must be addressed if the school was to continue meeting its accountability goal. Their current solution to this problem seems to have pointed upper primary teachers toward a more traditional scope and sequence as they attempt to inject KIRIS content into their instruction.

The teachers have not, however, abandoned all the primary program innovations: they continue to employ some flexible grouping and regrouping, the KELP assessment/reporting program, frequent communication with parents, and hands-on and collaborative education as strategies for reaching their academic goals. Frequent testing as the basis for regrouping enables continuous progress in the basic tool subjects.

The OCES dilemma—how to teach rigorous, challenging content while using developmentally appropriate practices—is shared by other Kentucky schools struggling with simultaneous implementation of a continuous progress primary program and assessment-driven reform. The OCES primary program seems to be evolving in a rational and potentially positive direction. What the teachers need is assurance that it is possible to integrate a KIRIS content focus into the developmentally appropriate practices of the primary program, coupled with specific guidance in how to do that—then they would have the best of both worlds.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 517 times since July 26, 2000

prior vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 8 Number 35 July 26, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

State Standards, Socio-fiscal Context and Opportunity to Learn in New Jersey

William A. Firestone
Gregory Camilli
Michele Yurecko
Lora Monfils
David Mayrowetz

Rutgers University

Abstract

A survey of 245 New Jersey teachers provides a baseline for examining how the introduction of state standards and assessments affects the teaching of math and science in the 4th grade. These policies are promoting teaching of additional topics in both areas. The changes in the delivery of professional development have not yet been sufficient to lead to substantial changes in instructional practice. While inequities in access to material that characterized the state in the early 1990s have diminished, we find a pattern of inquiry-oriented science teaching more prevalent in wealthy districts and teaching to the test more prevalent in poorer ones. We also note some areas where middle-income districts appear disadvantaged.

A central goal of the standards movement has been to help all children learn challenging content (Smith & O'Day, 1991). Forty-four states have now adopted standards for student proficiency in the core academic areas, 41 states have aligned assessment with their math standards, and 25 have aligned assessment with their science standards (Quality Counts, 2000). While great attention is being paid to what students are learning, less scrutiny has been given to what they are taught. Yet, the former depends at least in part on the latter (Wiley & Yoon, 1995). For that reason, state standards are intended to provide guidance on what should be taught, as well as what students should learn (Smith, Fuhrman & O'Day, 1994).

The adoption of standards and assessments does not guarantee students access to instruction, especially for poor students. For that reason, people have begun to worry more about "opportunity to learn" (OTL) or "whether or not... students have had an opportunity to study a particular topic or learn how to solve a particular type of problem presented by a test" (Husen as cited in McDonnell, 1995, p. 306). Advocates for minorities have seen the reporting of OTL standards as a way of ensuring that poor and minority students are not disadvantaged inappropriately when standards are raised. As one observer noted, without OTL standards, "you don't know if the school if failing, or if students are failing" when test scores are low (Rothman, 1993, p. 21).

Both the federal and state governments have been much more willing to adopt student performance standards than OTL standards since the latter specify the government's obligation to deliver services to students (McDonnell, 1995). Moreover, the legal mandate for guaranteeing that OTL be provided is ambiguous, even though the issue arose in the early years of state testing. According to Millman and Green (1989, p. 356):

> The court decision in the *Debra P. vs. Turlington* (1981) case seems to have established the necessity that, at least for certification tests for high school graduation, the tested material must consist of content that is currently taught, that is, the student must have been provided adequate preparation and, thus, had a fair opportunity to learn the material.

Precise requirements of a fair opportunity to learn remain ambiguous.

Several decades of research have indicated how difficult it is to change teaching practice (McLaughlin, 1990; Cuban, 1993). Simply imposing standards by decree is not likely to modify teaching practice if teachers do not understand what is expected of them or have the resources to carry out a standards-based program of instruction. The situation can be especially challenging in mathematics and the sciences where elementary education teachers may lack the background knowledge to effectively teach more challenging content.

This article introduces a project designed to explore how state standards and related policies influence teaching practice. In May, 1996, New Jersey announced a new set of "core curriculum content standards" (NJSDE, 1996). These standards began to take practical reality for elementary school teachers when state assessments aligned with these standards were introduced in 1998. In the Spring of 1999, as the state administered its new fourth grade mathematics and science assessments for the second time (the first time for which results would actually be released publicly), we began a three-year study to examine how teachers in those grades teach mathematics and science. Using a state-wide representative survey, this article describes three dimensions of teaching practice: the content taught, access to and use of materials, and teaching to the test. In each area, we investigate what in being taught and how equitably practices are distributed among wealthy and poor districts. We also explore teachers' background knowledge and opportunities to learn about new practices. Our preliminary conclusions are that:

- The introduction of standards and assessments is broadening the range of topics taught in mathematics and science.
- A useful baseline measure for assessing teaching to the test can be developed.
- Opportunities remain limited for elementary teachers to learn the new knowledge required to improve their mathematics and science teaching.
- The inequities between wealthy and poor districts are complex and may be overstated, but there is clearly more teaching to the test in poor, urban districts and more hands-on science teaching in wealthier districts.

Before addressing these issues we describe the context for standards implementation in New Jersey and the research methods employed in the study.

The Policy Context

In the last decade educational policy in New Jersey has been driven by two related phenomena: school finance litigation and the development of standards and related assessments. Whereas financial resources can influence the distribution of OTL, legal battles surrounding the school finance issue also motivated the adoption of standards.

School Finance Litigation

Since school finance litigation began in New Jersey thirty years ago, there have been two court cases, eleven decisions, numerous school finance bills, and other laws and regulations (Goertz & Malik, 1999). The litigation and related legislation has focused on whether the state was obligated to provide all children therein a "thorough and efficient education." While these actions have had a number of implications for education in New Jersey, two are especially critical here: the definition of a thorough and efficient education, and the financial provisions to ensure that all children could receive one.

The court has been reluctant to define a thorough and efficient education except in the broadest terms:

> For those special needs districts [the approximately 30 poor urban districts identified by the court as inequitably served by the state], a thorough and efficient education—one that will enable their students to function effectively in the same society with their richer peers both as citizens and as competitors in the labor market—is an education that is the substantial equivalent of that afforded in the richer districts (*Abbott v. Burke*, 643 A.2d 575, 580 (1994)) (Abbott III)

Beyond stating that children in poor districts should get the same education as those in wealthy districts, this decision provided very little guidance; and the court continued its multi-year effort to urge the state department of education to specify criteria in more detail. This was accomplished in part in the Comprehensive Plan for

Educational Improvement and Financing (CEIFA), the school funding law of 1996, which defined a thorough education as one in which children succeeded in meeting the 56 outcomes specified in the Core Curriculum Content Standards. Thus, the standards became the criteria for educational effectiveness, and state tests administered in 4th, 8th, and 11th grade would operationalize those criteria. The court found that these standards and assessments were "the first real effort on the part of the legislative and executive branches to define and implement the educational opportunity required by the Constitution... and are facially adequate as a reasonable legislative definition of a thorough and efficient education" [*Abbott v. Burke*, 693A.2d 417, 428 (1997) (Abbott IV)].

This effort was not sufficient to clarify what constituted adequate educational funding for all children in the state. Thus, the court continued to use a two-part yardstick. First, the poorest districts in the state should spend essentially the same per capita as the wealthiest districts (Goertz & Malik, 1999). The state had developed a classification of districts (District Factor Group or DFG) based on a composite measure of community, social, and economic variables such as the educational and occupational background of the population, per-capita income of the district, and mobility. The DFGs were designated by letter with the poorest districts labeled "A" and the wealthiest labeled "J". Per-pupil spending in the special needs districts designated by the court was expected to match that of the highest DFG districts. As late as 1993-94, the 14% of districts were spending 22% more than the poorest although their collective tax rate was 43% lower (Firestone, Goertz & Natriello, 1997).

Second, in addition to equal base spending, the court required the state to support a series of supplemental programs for the poor urban, districts. Urban schools were expected to implement a whole school reform program model such as Success for All (Porter, 1999), extend early childhood education services to 3- and 4-year olds, and began programs to refurbish aging and decaying buildings. Since these programs could not be supported locally, they had to be underwritten by the state (Goertz & Malik,1999; Erlichson, Goertz, & Turnbull, 1999). By the 1999-2000 school year, the equal base funding provisions were in place and implementation of the special programs had begun although not without disputes about the local level of funding and district discretion in designing their whole-school reform and early childhood programs.

Equal basic funding is an important development, and extremely unusual in a state noted for inequities in education. In 1996 only two states had a greater dollar gap in spending between the fifth and 95th percentile districts than New Jersey (Quality Counts, 2000). However, the court remedies and new funding formula did not extend to all districts. Schools in DFGs as low as B and into the middle of the fiscal distribution were spending less per child than either the wealthiest or the poorest districts in the state.

Standards and Assessments

As a normative perspective, standards theory recommends that state standards become the criteria with which assessments are aligned. However, like many American states, New Jersey began with assessments rather than standards. Its first testing system, begun in the late 1970s, was designed to measure "minimum basic skills" as a means of maintaining the accountability of poor urban districts, who at that point were receiving a new infusion of state funds. Several revisions ensued, and by the early '90s the keystone of the state's testing system was the High School Proficiency Test (HSPT), administered in 11 th grade as a partial requirement for high school graduation. This test covered mathematics, reading, and writing at a more challenging level than the earliest test, but passing score was still set at a basic skills level. The HSPT was accompanied by an Early Warning Test (EWT), given in 8th grade to help schools identify children at risk of failing the graduation test. These tests had special significance to educators because patterns of low scores on these tests could become grounds for state takeover of a district. Districts were also expected to administer conventional achievement tests of their own choice at grades not tested by the state (Firestone et al., 1997).

During the 1990s as the standards movement took hold nationally, teams of content experts and teachers were formed within the state to write the core curriculum content standards in seven curricular areas as well as a set of cross-content workplace readiness standards. These efforts were heavily influenced by national standards documents in mathematics and science and became official in May, 1996 (NJSDE, 1996). The resulting standards for mathematics and science are listed in Appendix A. These core standards are accompanied by cumulative progress indicators for grades 4, 8, and 12. Separate documents provide curriculum frameworks to offer guidance to educators in implementing the standards.

The state is now phasing in 4th, 8th, and 11th grade tests that are intended to be aligned with the standards in each area. The degree of alignment to the standards is difficult to assess because—as in many states—strict confidentiality is maintained over operational test items. This creates difficulties for educators who wish to be given test results item by item in order to seek an easier method for aligning their instruction more closely with the assessments.

The current tests are an effort to move away from the basic skills or advanced basic skills orientation that characterized earlier state tests. The 4th grade mathematics tests include 32 closed-ended and five open-ended items; and the matrix for selecting items includes a dimension of "problem-solving skills" with categories like "procedural knowledge, conceptual understanding, and problem-solving skills" (NJSDE,1998, p. 6). The 4th grade science test is similarly organized. One sample open- ended item and one sample closed-ended item from the test specifications are included in Appendix A. The 4th grade mathematics and science tests were first administered in the spring of 1998, but because of technical problems scores were not released. The following year scores were released in the fall after the spring 1999 administration.

The introduction of new standards and assessments in mathematics and science should provide clarity regarding what is expected to be taught in each area, and ensure that these subjects receive consistent attention. Whether this attention takes the form of short-term "teaching to the test" or deeper changes in practice, and whether access to new forms of instruction is equally distributed in the state remains to be seen. Recent court and legislative actions may further stimulate access to new forms of instruction. We turn now to the survey designed to address these issues.

Study Sample

In the spring of 1999, we initiated a three-year study to examine teachers' response to the new testing program in the areas of mathematics and science. Data were collected from a statewide sample of 4th grade teachers. Just over 600 teachers were asked to respond to a complex set of instruments. After extensive telephone follow-ups and remailings, 245 teachers completed a telephone survey, 172 completed an additional mailed questionnaire, and 110 provided examples of mathematics and science lessons they taught, including materials given to students and more detailed reports on teacher and student activities conducted with those materials. (Note 1) The sample is highly representative with regard to district wealth as measured by DFG (See Table 1).

Past research suggests that successful change in teaching practice depends on opportunities for teachers to learn new practices required by the policy (Cohen & Barnes, 1993; Firestone et al., 1998). However, the kind of professional development that is most likely to lead to substantial change in practice continues to be rare (Loucks-Horsley, Hewson, Love, & Stiles 1997). In order to assess the effects of professional development, we sought to oversample schools that were known to engage in extensive professional development with respect to mathematics and science. The New Jersey State Systemic Initiative shared with us results of a survey identifying districts engaged in the most extensive professional development in those subjects. We attempted to ensure that 25% of our sample came from these districts. In fact 49 of the completed telephone interviews (20%) and 30 of the completed mailed questionnaires (17%) came from high professional development districts.

Table 1

Distribution of Responses by DFG

	District Factor Group						
	AB: (Poorest)	CD	DE	FG	GH	IJ: (Wealthiest)	Total
Interviews	71	29	32	24	35	54	245
Percent	29%	12%	13%	10%	14%	22%	100%
Questionnaires	49	21	23	14	25	40	172
Percent	28%	12%	13%	8%	15%	23%	100%
4th Grade Students in State (%) 30%	9%	15%	13%	13%	19%	100%	

In the following section we explore what content is being taught, teachers' access to materials, the extent of teaching to the test, self-reported knowledge about standards, and teachers' access to professional development.

Content Coverage

Standards and assessments are supposed to be able to influence the content taught to children. Smith (1991) and Corbett & and Wilson (1991) found that the introduction of minimum competency tests narrowed the range of subjects taught in a school to what was on the test. Firestone, Mayrowetz, & Fairman (1998) suggested that the introduction of more complex performance assessments can affect the presence and order of topics taught. There is reason to believe that the new standards and assessments are affecting content coverage in New Jersey. Fifteen percent of our sample said they were teaching more math and 14% said they were teaching more science. Noticeable changes are being made within each content area but these are different in mathematics and science.

Math Content

Traditionally, elementary mathematics has focused on basic arithmetic—addition and subtraction of whole numbers with some introduction of fractions and decimals and geometric shapes. New Jersey's Core Curriculum Content Standards expect the introduction of a wide range of content at the fourth grade level, including a broader range of geometric issues; the foundations of algebra; better understanding of measurement; an introduction to statistics, probability, and data analysis; and discrete mathematics (NJSDE, 1996). We wanted to access how teachers were using their time in mathematics and how that time use was changing. In order to avoid influencing respondents familiar with the standards terminology, we identified 17 topics that represented a mix of classic elements of the elementary mathematics curriculum and areas that were not likely to have been taught before the standards were introduced [Appendix C]. We then asked teachers how many lessons they taught each of the 17 topics, and whether they had increased or decreased the time allocated to each topic in the last three years—i.e., when the standards were being introduced and the ESPA was being given for initially.

Although we do not have a firm fix on how time was allocated to topics before the standards were introduced, it appears that the gap between conventional and newer topics is being reduced with teachers adding time to newer topics. Working with experts familiar with math teaching in the state, we identified three traditional topics: paper and pencil mathematical operations with whole numbers, adding and subtracting decimals via paper and pencil, and place value relationships (whole numbers, decimals); and three newer topics: open sentences, use of variables (strategies used to prepare students for algebra), probability, and dealing with data (collecting, organizing, analyzing, and displaying data). Most teachers reported that they spent many lessons on whole number operations: 96% spent eleven or more lessons a year on that topic. In addition, 58% devoted eleven or more lessons to

place value relationships, and 22% spent that much time on adding and subtracting decimals. Although fewer teachers devoted substantial time to the newer topics, 50% spent 11 or more lessons on dealing with data. Thirty three percent spent 11 or more lessons on open sentences, and 14% on probability.

Although the larger balance of teaching time was spent on older topics, most teachers reported *increasing* the amount of time the spent on the new topics (Figure 1). In general time spent on the older topics remained fairly constant, with the exception of whole number operations. A large portion of teachers (29%) reported decreasing time spent on whole number operations. Based on this evidence, it appears that newer topics are taking a more prominent place in the curriculum, but not necessarily replacing older topics.

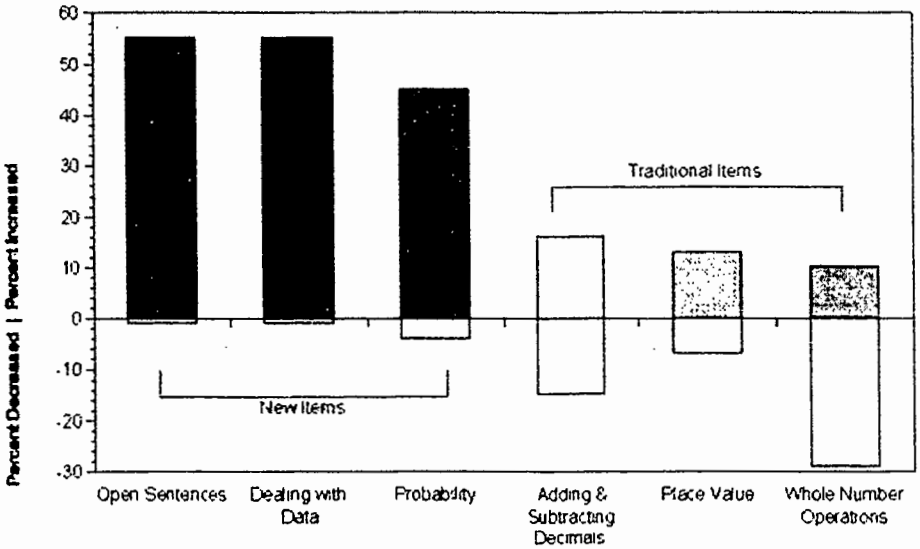


Figure 1. Percent Changes in Mathematics Items

We also explored whether the time allocated to topics was the same in wealthy and poor school districts. In 13 of the 17 topic areas there were no significant differences between DFGs. However, in four topics identified as new by our mathematics experts, we noted an interesting u-shaped pattern. Teachers in poor, urban districts and the wealthy districts spent more time on these topics than middle income districts (Table 2). An explanation for this pattern has not yet been found.

Table 2
Differences by DFG in Lessons Allocated to Math Topics
(Percent of teachers devoting 11 or more lessons to a topic, n = 151-154)

	District Factor Group			
	Abbott*	C-E	F-H	IJ
Probability	27%	12%	3%	19%
Patterns, functions	49%	16%	21%	36%
Open sentences	46%	29%	19%	41%
Discrete math	54%	25%	16%	36%

* District wealth is generally measured by DFG. The Abbott districts are all DFG A or B and have been designated by the state Supreme Court as those where spending must be equalized with wealthy districts in the state. The DFG metric runs from A (districts with large numbers of poor and generally at-risk children) to IJ with large numbers of children from wealthy families. Teachers from DFG-B districts that are not "Abbott districts" have been excluded from this comparison.

Science Content

As with mathematics topics, we explored the amount of time spent on topics that experts thought would have been part of the 4th grade curriculum before the state standards were introduced and topics that were probably introduced in response to the standards. It was more difficult to generate a focused list in science, but our experts identified four old topics: weather and climate; life systems; habitats, ecosystems and adaptation; and features of plants and animals. They also identified six new topics in three clusters: the process of doing science (investigative skills and using mathematics in science); chemistry [structure and properties of matter and states of matter (solid, liquid gas)]; and physics [forces, motion and energy and invisible forces (gravity, electricity, and magnetism)].

The difference between old and new topics was less marked in science than in mathematics. Seventy two percent of teachers reported that they spent eleven or more lessons on investigative skills, a new topic. However, the second and third most addressed topics were 2 old topics: life systems (54% spending eleven or more lessons);and habitats, ecosystems, and adaptation (49%). After that, distinctions are difficult to make. Topics in the 30% range include the two remaining old topics (features of plants and animals (32%); and weather and climate (39%)), and three new ones (using mathematics (34%); solids, liquids and gases (33%); and gravity, electricity, and magnetism (35%)). The remaining two new topics were taught extensively by fewer teachers. Only 21% reported teaching the structure of matter in eleven or more lessons and 23% spent that much time on forces and energy.

In general more teachers reported increasing the amount of time they spent on science topics than on mathematics topics (Figure 2). The biggest increases were in the new topics related to the process of doing science (investigative skills with 69% reporting increases in the amount of time spent, and using mathematics in science where increases were mentioned by 42% of teachers). The smallest increases were in physics-related topics (forces and motion; and gravity, electricity, and magnetism) each with about 25% of teachers reporting increases. About a third of the teachers reported increases in time spent on chemistry- related topics, weather, and most of the biology-related topics. Plants and animals were the exception with only 25% reporting increases. Unlike the mathematics area, there were no meaningful differences between DFGs in treatment of science topics.

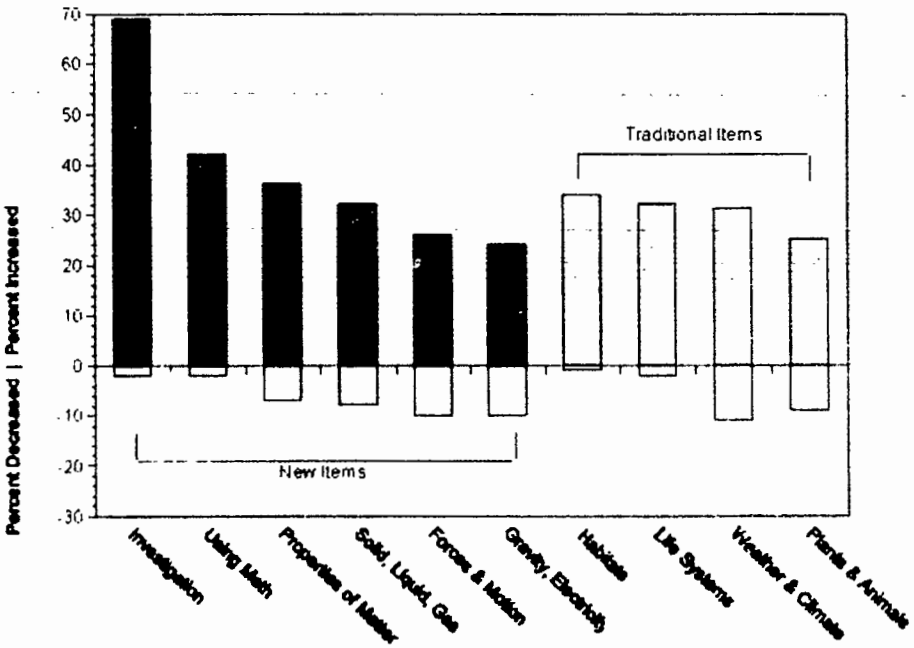


Figure 2. Percent Changes in Science Items

Access to Materials

BEST COPY AVAILABLE

New Jersey's Core Curriculum Content Standards place an increased emphasis on a more active role for students to take in learning mathematics and science. The mathematics standards require students to "develop an ability to pose and solve mathematical problems,... develop reasoning ability and... become self reliant independent mathematical thinkers; [and] regularly and routinely use calculators, computers, manipulatives, and other mathematical tools to enhance mathematical thinking, understanding, and power" (New Jersey State Department of Education, 1996, p. 4-9). The science standards require that students "develop problem-solving, decision-making, and inquiry skills, reflected by formulating usable questions and hypotheses, planning experiments, conducting systematic observations, interpreting and analyzing data, drawing conclusions and communicating results" (New Jersey State Department of Education, 1996, p. 5-3). These changes are in keeping with national standards which require more problem solving in mathematics and hands-on inquiry in science. At the same time they place greater demands on districts to provide additional materials—mathematical manipulatives, calculators and computers, the wherewithal for scientific experiments—beyond the basic textbooks that have been so typical of American teaching (Cuban, 1993). In fact, some textbooks include alternatives like science kits or math manipulatives.

Access to teaching equipment and supplies has historically been unequal, favoring wealthy districts. In the early 1990s, teachers in poor, urban districts reported less access to both textbooks and computers than their peers in wealthy districts. For a period of time following the passage of the Quality Education Act (QEA) which increased funding to urban districts for a short time in the early 1990s, there was some indication that poor districts were working hard to bridge the gap between themselves and wealthier districts. However, they have not been successful (Firestone et al., 1997).

The current study indicates that access to materials may be improving in poor districts. Across DFGs teachers reported having enough materials for most purposes, especially for teaching mathematics. Ninety-five percent of the teachers surveyed reported having enough math textbooks for every child to have one. (Note 2) Ninety-four percent reported having enough manipulatives for children to share, and 97% reported enough calculators for every child. The situation is nearly as good in science where 77% of the teachers reported having enough textbooks for every child, 76% reported enough science kits either for every child or for children to share, and 85% reported enough measurement and observation tools to share.

Use tends to lag behind access. Seventy eight percent of teachers report using their math texts almost every day, (Note 3) 66% use manipulatives once or twice a week, and 53% use calculators once or twice a week. The pattern in science is somewhat different. While 36% report using a textbook everyday, 40% report using it once or twice a week. Sixty-five percent report using science kits at least once a week, and 38% report using measurement and observation tools that often.

We did not identify any inequities in access to mathematics materials, supported by the high percentage of teachers who reported having enough math textbooks for every child. The situation in science is more complicated because teachers in poor, urban districts appear to emphasize the use of textbooks, while those in the wealthier districts balance textbooks with the use of science kits and other materials (Figure 3). Almost all the teachers in the Abbott districts and mid-wealth districts say they have enough science textbooks for every child and more than four fifths use them weekly. However, less than half the teachers in the wealthy districts have enough textbooks for every child and use them weekly. A third of the teachers in wealthy districts have enough kits for every child and two thirds use them weekly.

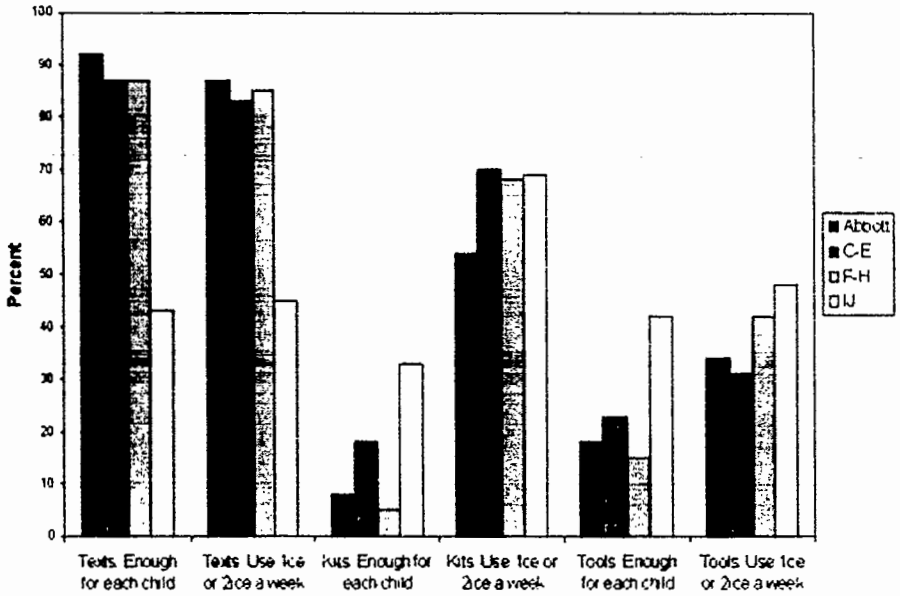


Figure 3. Access To and Use Of Science Materials

Kits are much less accessible in the poor and mid-wealth districts. Still about half the teachers in urban districts report using them weekly and use in the mid-wealth districts is comparable to that in the wealthy districts. The pattern of access to tools for observation and measurement parallels that to access to kits with substantially more teachers reporting having enough for every child in the wealthiest districts. There is a gradual trend of increasing use as one moves from the Abbott to the wealthiest districts. The reasons for these differences are not clear. However, the fact that most teachers in the state report little change in their access to materials suggests that this pattern reflects a difference in philosophy about how to teach science more than recent changes in funding.

Teaching to the Test

One of the greatest concerns with standards- and assessment-based reform has been that this strategy might lead to teaching to the test and its concomitant negative effects such as narrowing the curriculum; constricting instruction time; increasing the amount of drill while undermining efforts to promote higher order thinking skills; and increasing stress for teachers and students (Corbett & Wilson, 1991; Smith, 1991). There is also a fear that teaching to the test will undermine the validity of test results by artificially inflating test scores (Mehrens, 1998). There has been some question about whether these are inevitable effects of high- stakes accountability-oriented tests. Some have suggested that changes in test format should include more performance- oriented items and test items assessing more than mere retention of facts and computation skills might lead to tests worth teaching to and encourage teaching that promoted more conjecture, exploration, and active participation in learning (Baron & Wolf, 1996; Rothman, 1995).

To explore the distribution of teaching to the test in the state, we developed a seven-item scale with a mixture of items that seemed to reflect some of the feared negative effects of this practice and others construed as positive. The scale had an alpha coefficient of .71. Specific items included:

1. Teach test staking mechanics like filling in bubbles, how to put your name on the test, or how to pace yourself during the test.
2. Motivate students to make their best effort on the ESPA, such as suggesting they prepare by getting a good night's sleep or encouraging them to try hard.
3. Have students use rubrics to grade each other's work.
4. Teach the regular curriculum using performance-based exercises similar to the ESPA.

5. Teach test-besting skills like methods for turning story problems into arithmetic calculations or how much to write after an open-ended math item.
6. Use commercial test-preparation materials like "Scoring High" and "Measuring Up on the ESPA."
7. Give practice tests with items similar to those on the ESPA.

We asked teachers how often they performed these activities (on a scale of 1-4) all year long and the month before the ESPA was given. (Note 4) Figure 4 shows two patterns in teachers' reported teaching to the test. First, as might be expected, there is a small increase in activity during the month before the test compared to the entire year (scale mean of 2.50 for the whole year versus 2.86 for the month before the test). Second, there is a distinct pattern of teachers in the Abbott districts reporting more teaching to the test than teachers in the wealthiest districts. Teachers in the mid- wealth districts fell somewhere in between. Thus, the emphasis on test preparation as a separate activity were concentrated in the districts that most need help in improving student learning.

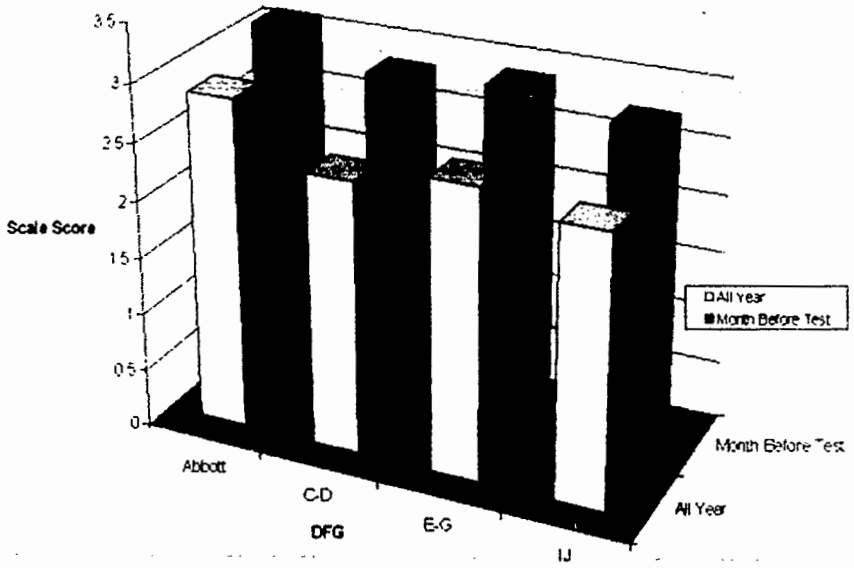


Figure 4. Teaching to the Test

Familiarity with Standards

We asked teachers to report how familiar they are with state and national standards in mathematics and science. Teachers' familiarity with state standards could contribute to changes in content taught, although central office staff who understand state standards and assessments can unilaterally change district curriculum. The national standards movement in science, and especially in mathematics precedes New Jersey's efforts by several years; and some districts were using those national standards to guide changes before state standards were adopted or tests were implemented.

Teachers were much more familiar with state than national standards. Fifty-seven percent said they understood the state's mathematics standards well, (Note 5) and 53% say they are understand the science standards well. In contrast, only 28% said that they understood the national mathematics standards well and 16 said they understood the national science standards well. Even if teachers overestimated their understanding of the standards, the state's effort has increased attention to standards-based teaching here.

For the most part, understanding of standards is equally distributed across wealthy and poor districts. The one exception is the national mathematics standards

where there is a complicated pattern of differences between districts (Table 3). Generally, more teachers in the wealthy districts believed that they understandd the national standards well. However, it is not true that most teachers in the Abbott districts have limited familiarity with the national math standards. The largest concentration having moderate familiarity is in the Abbott districts while the almost two thirds of the CE teachers have only limited familiarity with the national standards. One possibility is that the wealthy districts have sought to adopt the national standards for a long time. Growing familiarity in the Abbott districts may reflect a mix of three factors: a side effect of the attention to standards in general from the adoption of state standards, the special pressures placed on the Abbott districts by the state as a by-product of the series of court cases and large amount of state money going to those districts (Firestone & Nagle, 1995), and the additional funds coming from CEIFA after the Abbott IV decision.

Table 3
Understanding of National Mathematics Standards by DFG
(Percent of Teachers, n = 158)

	District Factor Group			
	Abbott	C-E	F-H	IJ
Limited*	37%	63%	32%	33%
Moderate**	47%	21%	29%	28%
Extensive***	16%	16%	40%	39%

* Awareness only and read through once or twice.
** Understand somewhat (can implement parts in class)
*** Understand well (can implement fully in class) and expert (could lead workshop)

Professional Development

Past research on policy implementation in a variety of fields suggests that regardless of changes in incentives and punishments, teachers will not change their practice until they have learned how to perform the new tasks expected of them (Berman 1986, Cohen and Barnes, 1993). Firestone and colleagues (1998) suggest that one reason state-administered performance-based assessment has had limited impact on teaching is because teachers have had limited opportunities to learn the new content and pedagogy required by the new assessments.

Teachers reported on several dimensions of their professional development experience. Regarding the source of professional development, most learning opportunities for teachers came directly from the district. Sixty seven percent of teachers reported that some time in their district-provided professional development days in the last year had been devoted to mathematics or science. In the last year, 40% had mentored student teachers or first year teachers, 41% had served on district curriculum development or textbook selection committees, and 21% had served as lead or specialist teachers helping other experienced teachers in their district. All of these are learning experiences even though they may involve helping others.

Relatively few teachers had opportunities to develop new knowledge by interacting with experts from outside the district. Eighteen percent had taken a college course in math, science, or math or science education in the last year. Twenty two percent had participated in one the programs for improving math and science teaching supported by the National Science Foundation through its State and Local Systemic Initiatives or the US Department of Education through its Eisenhower grants to institutions of higher education. Given elementary teacher's reputation for aversion to mathematics and science, these numbers are fairly reasonable. However, since the objective is to achieve statewide high quality mathematics and science teaching, it seems quite unlikely that teachers' understanding of effective practice will grow quickly unless more avail themselves of these opportunities.

One recurring criticism of professional development is that it is usually provided through one-shot workshops where teachers receive limited and often inapplicable

information with little or no follow up to help in using what they are supposed to have learned. That seems to have been the case among New Jersey's fourth grade teachers (Table 4). Only about one fifth of the teachers reported having more than two days of professional development on either content and instruction in science and math. Slightly fewer received more than two days of professional development on strategies to help students score high in math or science. It is somewhat encouraging that teachers received about as much professional development on the underlying content and instructional issues as they did on strategies to raise test scores. On the other hand, only one in 20 received more than two days on using assessment results. It is particularly disconcerting that teachers received so little support in using assessment results to improve instruction, although this may be because the state had not yet reported any ESPA results to schools when this survey was conducted.

Not only is professional development limited, so is follow up. Between 20% and 30% of the teachers report being visited later by a trainer. Follow up by principals is more common, but principals are often less well informed about the content of professional development. Their follow up may show concern and signal that the material covered is important, but substantive assistance is likely to be less than that coming from an expert. Nevertheless, between one third and one half the teachers found the professional development they received to be very useful. This may be in part a reflection of the growing demand for help in this area.

Table 4
Time in Professional Development
(Percent Reporting Various Categories)

	More than 2 days PD in year	Follow-up by trainer	Follow-up by principal	PD is very useful
Content and instruction in science	22%	25%	22%	44%
Content and instruction in math	20%	25%	26%	48%
Using assessment results	6%	21%	35%	30%
Strategies to score high in math	19%	29%	33%	48%
Strategies to score high in science	14%	22%	29%	41%

Where New Jersey teachers received more professional development, they found it more useful. The correlation between the amount of time spent in professional development and its perceived utility were .66 for content and instruction in science, .63 for content and instruction in mathematics, and .61 for using assessment results. They were lower for strategies for scoring high in math and science (.44 and .40, respectively). These findings suggest that extensive professional development efforts will be most helpful when helping teachers better understand the underlying material in a subject and effective strategies for helping students learn it. Longer time investments may also pay off for helping teachers to use assessment strategies to improve practice. Comparable concentrations are probably not as necessary to give teachers strategies to raise test scores.

Discussion

While there are limitations to what can be learned about changes in teaching practice from one administration of a survey that focuses on elementary school mathematics and science, the data presented here suggest some tentative conclusions

and raise questions about two issues: ongoing changes in practice, and differences between wealthy and poor districts.

Statewide, it appears that the topics taught as part of the 4th grade curriculum are changing. This may have implications for elementary curriculum in general. In mathematics, what had been an unremitting diet of whole number facts is being leavened with other topics like probability and dealing with data. Generally, more science is being taught, and the small sampling of biology and meteorology is being expanded. There is a large increase in attention to the process of scientific investigation, some increase in attention to the introduction of chemistry and at least a smattering of attention to physics-related topics. These changes help prepare children to use mathematics as part of their adult life and give them an introduction to a broader range of science topics.

The simple addition of topics may be a mixed blessing, however. One criticism of mathematics teaching in the past has been that too many topics are taught at too little depth (Schmidt, McNight, & Raizen, 1997). The addition of new topics to the state standards could exacerbate such shallow coverage. The quality and depth of coverage is difficult to assess with surveys; hopefully, direct observation in classrooms, which is currently underway, will help address this issue. It will also be useful to collect longitudinal data on coverage of content areas to verify that the changes we believe are happening are in fact taking place. Teachers are also becoming more familiar with the state standards, and believe they are more familiar with state than with national standards. We suspect that the extent of their familiarity is overstated. Again, we hope to learn more from direct observation.

On the equity front, the picture is mixed. The good news is that some of the obvious inequities in access to materials that were prevalent at the beginning of the decade appear to be fading. However, there are hints that two pedagogies may be developing in the state: one for children in districts serving the poor, and another for districts serving the wealthy. Pedagogy in the poor districts may come to be dominated by conventional, textbook-oriented teaching and teaching to the test, while wealthier districts seem to be moving towards more exploratory, active modes of learning that are less dependent on textbooks and less driven by state tests. If so, the reasons are likely to have less to do with differences in funding and more with heavier pressures to comply with state expectations in urban districts and the challenges that come with teaching poorer children (Natriello, Pallas & McDill, 1990).

There is also the issue of those districts in the middle of the DFG distribution. These more working-class districts are not as well funded as either the Abbott districts or the wealthy districts. There are some indications that teachers in the Abbott districts are moving faster than those in the poorer of the mid-wealth districts to embrace the standards and introduce new topics to the curriculum. How strong this trend is, whether it will continue, and what its implications are for teaching practice and student achievement remain to be explored through further surveys and direct observation in classrooms.

Notes

This article was presented as a paper at the Annual Meeting of the American Educational Research Association in New Orleans, LA, April, 2000. We wish to thank Warren Crown, Roberta Schorr, John Shafransky, Sharon Sherman, and Carol Stearns for their assistance. This research was supported by a grant from the National Science Foundation. The opinions expressed are those of the authors. Neither the Foundation nor Rutgers University is responsible for them.

1. The teacher work samples are not used in this report.
2. The choices offered teachers were none, one or two to demonstrate in class, enough for children to share, and enough for every child to have one.
3. The options were almost every day, once or twice a week, once or twice a month, once or twice a semester, and never.
4. Respondents were asked to report on a 4-point scale where 1 was "almost never" and 4 was "almost always."
5. The actual choices were "Awareness only, read through once or twice,

understand somewhat (can implement parts in class), understand well (can implement fully in class), and expert (could lead workshop)." The responses reported are for the last two combined.

References

Baron, J.B. & Wolf, D.P. (1996). *Performance-based student assessment: challenges and possibilities*. Chicago, IL: University of Chicago.

Berman, P.E. (1986). From compliance to learning: Implementing legally-induced reform. In D. Kirp & D. Jensen (Eds.), *School days, rule days*. (pp. 46-62). Philadelphia, PA: Falmer.

Cohen, D.K., and Barnes, C.A. (1993). Pedagogy and policy. in D.K. Cohen, M.W. McLaughlin, & J.E. Talbert (Eds.) *Teaching for understanding*. (pp. 207-239) San Francisco: Jossey Bass.

Corbett, H.D. & Wilson, B.L. (1991). *Testing, reform, and rebellion*. Norwood, NJ: Ablex.

Cuban, L. (1993). *How teachers taught: Constancy and change in American classrooms, 1890-1980*. (2nd ed.) New York: Teachers College Press.

Erlichson, B.A., Goertz, M. & Turnbull, B. J. (1999) *Implementing whole school reform in New Jersey: year one in the first cohort schools*. New Brunswick, NJ: Rutgers, the State University of New Jersey.

Firestone, W. A., & Nagle, B. (1995). Differential regulation: Clever customization or unequal interference? *Educational and Evaluation and Policy Analysis*, 17(1), 97-112.

Firestone, W.A., Goertz, M.E., & Natriello, G.J. (1997) *From cashbox to classroom: The struggle for fiscal reform and educational change in New Jersey*. New York: Teachers College Press.

Firestone, W.A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis*, *20*(2)95-113.

Goertz, M. & Malik, E. (1999) In search of excellence for all: the courts and New Jersey school finance reform. *Journal of Education Finance, 25*(Summer)5- 32.

Loucks-Horsley, S., Hewson, P.W., Love, N., & Stiles, K.E. (1997) *Designing professional development for teachers of mathematics and science*. Thousand Oaks, CA: Corwin Press.

McDonnell, L. M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis*, *17*(3), 305-22.

Mehrens, W. A. (1998). Consequences of Assessment: What is Evidence? [Online]. *Educational Policy Analysis Archives*, *6*(13). Available: http://epaa.asu.edu/epaa/v6n13.html

Millman, J. & Green J. (1989). The specification and development of tests of achievement and ability. In R. L. Linn (Ed.), *Educational measurement* (pp. 335-366). New York: Macmillan.

Natriello, G., McDill, E.L. & Pallas, A.M. (1990). *Schooling disadvantaged children: Racing against catastrophe*. New York: Teachers College Press.

New Jersey State Department of Education. (1996). *Core curriculum standards*.

Trenton, NJ: Author.

Porter, A. (1999). Creating a system of school process indicators. *Education Evaluation and Policy Analysis*, *13*(1), 13-29.

Quality Counts 2000. (2000). *Education Week*, *XIX*(18).

Raymond Arthur Abbott, et al. v. Fred G. Burke, Commissioner of Education, et al., 643 A.2d 575, 580 (1994) (Abbott III)

Raymond Arthur Abbott, et. al. v. Fred G. Burke, Commissioner of Education, et. al., 693A.2d 417, 428 (1997) (Abbott IV).

Rothman, R. (1995). *Measuring up: standards, assessment, and school reform*. San Francisco, CA: Jossey-Bass.

Schmidt, W. H., McNight, C. C., Raizen, S. A. (1997) *A splintered vision: an investigation of U.S. science & mathematics education*. New York, NY: Kluwer Academic Publishers

Smith, M. & O'Day, J. (1991). Systemic school reform. In S. Fuhrman & B. Malen (Eds.) *The politics of curriculum and testing.* (pp. 233-67). Bristol, PA: Falmer Press.

Smith, M.L. (1991). Put to the test: The effects of external testing on students. Educational Researcher, 20(5), 8-12.

Smith, M.S., Fuhrman, S.H. & O'Day, J (1994) National curriculum standards: Are they desirable and feasible? In Elmore, R.F. & Fuhrman, S.H. (Eds.) *The governance of curriculum* (pp. 12-29). Alexandria, VA; ASCD.

Wiley, D.E. & Yoon, B. (1995). Teacher reports on opportunity to learn: Analyses of the 1993 California Learning Assessment System (CLAS). *Educational Evaluation and Policy Analysis*, *17*(3), 355-70.

About the Authors

William A. Firestone
Center for Educational Policy Analysis
Rutgers University

Email: wilfires@rci.rutgers.edu

William A. Firestone is Professor of Educational Policy; Chair of the Department of Educational Theory, Policy, and Administration and Director of the Center for Educational Policy Analysis. His research on the effects of both testing and professional development on teachers has appeared in the American Educational Research Journal, Educational Evaluation and Policy Analysis, and Kappan. His most recent book is *From Cashbox to Classroom: School Finance Reform and Educational Change in New Jersey* (with Margaret E. Goertz and Gary Natriello).

Gregory Camilli
Rutgers University

Email: camilli@rci.rutgers.edu

Gregory Camilli is Professor, Department of Educational Psychology, at the Rutgers Graduate School of Education. His areas of research interest include psychometric issues in educational policy, meta- analysis, and differential item functioning. Dr. Camilli is a member of the editorial Boards of *Educational Measurement: Issues and Practice*, *Educational Policy Analysis Archives*, and *Educational Review*. He is a

regular reviewer for *Applied Measurement in Education, Journal of Educational Measurement, Psychometrika,* and *Psychological Methods*, among others. As a member of the Technical Advisory Committee of the New Jersey Basic Skills Assessment Council, he provides expertise on testing and measurement issues to the state's assessment program.

Michelle Yurecko
Center for Educational Policy Analysis
Rutgers University

Michelle Yurecko, a Ph.D. candidate in Educational Psychology and a research associate at the Center for Educational Policy Analysis. She is a statistician and focuses on educational testing and measurement.

Lora Monfils
Center for Educational Policy Analysis
Rutgers University

Lora Monfils, a doctoral candidate in Educational Statistics and Measurement, is a research associate at the Center for Educational Policy Analysis and a mathematics educator. Her research interests concern large-scale assessment and modeling differential educational outcomes.

David Mayrowetz
Center for Educational Policy Analysis
Rutgers University

David Mayrowetz is a doctoral candidate in the Department of Educational Theory, Policy and Administration, Rutgers University. His interests include policy implementation, inclusion of students with disabilities, and assessment reform. He is the co-author, with William Firestone of "Rethinking "High Stakes:" Lessons from the US and England and Wales" (Teachers College Record, forthcoming) and with Carol Weinstein, of "Sources of Leadership for Inclusive Education: Creating Schools for All Children" (*Educational Administration Quarterly,* September 1999). He will be joining the Policy Studies faculty of the University of Illinois at Chicago in January 2001.

Appendix A
New Jersey's Core Curriculum Content Standards

Mathematics:

1. All students will develop the ability to pose and solve mathematical problems in mathematics, other disciplines, and every day experiences.
2. All students will communicate mathematically through written, oral, symbolic, and visual forms of expression.
3. All students will connect mathematics to other learning by understanding the interrelationships of mathematical ideas and the roles that mathematics and mathematical modeling play in other disciplines and in life.
4. All students will develop reasoning ability and will become self-reliant, independent mathematical thinkers.
5. All students will regularly and routinely use calculators, computers, manipulatives, and other mathematical tools to enhance mathematical thinking, understanding and power.
6. All students will develop number sense and an ability to represent numbers in a variety of forms and use numbers in diverse situations.
7. All students will develop spatial sense and an ability to represent geometric properties and relationships to solve problems in mathematics and in everyday life.

8. All students will understand, select, and apply various methods of performing numerical operations.
9. All students will develop an understanding of and will use measurement to describe and analyze phenomena.
10. All students will use a variety of estimation strategies and recognize situations in which estimation is appropriate.
11. All students will develop an understanding of patterns, relationships, and functions and will use them to represent and explain real-world phenomena.
12. All students will develop an understanding of statistics and probability and will use them to describe sets of data, model situations, and support appropriate inferences and arguments.
13. All students will develop an understanding of algebraic concepts and processes and will use them to represent and analyze relationships among variable quantities and to solve problems.
14. All students will apply the concepts and methods of discrete mathematics to model and explore a variety of practical situations.
15. All students will develop an understanding of the conceptual building blocks of calculus and will use them to model and analyze natural phenomena.
16. All students will demonstrate high levels of mathematical thought through experiences which extend beyond traditional computation, algebra, and geometry.

Science:

1. All students will learn to identify systems of interacting components and understand how their interactions combine to produce the overall behavior of the system.
2. All students will develop problem-solving, decision- making and inquiry skills, reflected by formulating usable questions and hypotheses, planning experiments, conducting systematic observations, interpreting and analyzing data, drawing conclusions, and communicating results.
3. All students will develop an understanding of how people of various cultures have contributed to the advancement of science and technology, and how major discoveries and events have advanced science and technology.
4. All students will develop an understanding of technology as an application of scientific principles.
5. All students will integrate mathematics as a tool for problem-solving in science, and as a means of expressing and/or modeling scientific theories.
6. All students will gain an understanding of the structure, characteristics, and basic needs of organisms.
7. All students will investigate the diversity of life.
8. All students will gain an understanding of the structure and behavior of matter.
9. All students will gain an understanding of natural laws as they apply to motion, forces, and energy transformations.
10. All students will gain an understanding of the structure, dynamics, and geophysical systems of the earth.
11. All students will gain an understanding of the origin, evolution, and structure of the universe.
12. All students will develop an understanding of the environment as a system of interdependent components affected by human activity and natural phenomena.

Appendix B
Content Area Topics From The Teacher Survey

Mathematics:

1. Paper and pencil mathematical operations with whole numbers (adding, subtracting, multiplying & dividing)

2. Doing mental math operations with whole numbers (adding, subtracting, multiplying & dividing)
3. Estimation (magnitude, results of computation, measurement)
4. Place value relationships (whole numbers, decimals)
5. Adding and subtracting decimals via paper and pencil
6. Identification of geometric figures
7. Area and Perimeter
8. Fraction Concepts (Fractions as parts of a whole, equivalency)
9. Operations with Fractions (addition, subtraction)
10. Measurement (customary, metric)
11. Probability
12. "Dealing with data" (collecting, organizing, analyzing and displaying data)
13. Statistics
14. Graphing
15. Patterns, functions
16. Open sentences, use of variables
17. "Discrete math" (Combinations, puzzles, optimization, classification, algorithms, networks, tree diagrams)

Science:

1. Understanding natural and man-made systems (recognizing systems, identifying parts)
2. Investigative skills (observing, classifying, dealing with data)
3. Using mathematics (measurement, estimating, counting)
4. Nature and history of science & scientists
5. Selecting and using tools
6. Needs of living things/Life systems
7. Habitats, ecosystems, & adaptation
8. Features and classifications of plants and animals
9. Structure and physical properties of matter
10. States of Matter: Solid, liquid, gas (heating and cooling)
11. Forces, motion & energy
12. Invisible forces (gravity, electricity & magnetism)
13. Earth Materials: Rocks, soil, fossils
14. Weather and climate
15. Earth, moon, sun system
16. Stars and galaxies
17. Humans and the environment

Appendix C
Sample ESPA Items

Traditional Mathematics Item:

Find the exact answer: 110 + 70

1. 18
2. 81
3. 180
4. 810

Newer Mathematics Item:

Mr. Jones gave each of the students in his class a one-ounce box of raisins. When the students opened the boxes and counted the raisins, they found different amounts. The tally sheet below shows their results.

Number of Raisins	Tally	Frequency
10	|	1
11	||	2
12	|||	3
13	|||||	5
14	|||	3
15	||	2

Construct a bar graph to represent the students' findings on the grid in your answer booklet. Be sure to label your graph completely.

Traditional Science Item:

Which thing does a living duck do that a toy duck does not do?

1. Floats on water
2. Breathes air
3. Makes a sound
4. Sits still

Newer Science Item:

Victor has two glasses. One glass is filled with ice cubes and the other is filled with water. Give three ways the ice and water are different.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 181 times since July 30, 2000

Education Policy Analysis Archives

Volume 8 Number 36 July 30, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Taxing the Establishment Clause: The Revolutionary Decision of the Arizona Supreme Court in *Kotterman v. Killian*

Kevin G. Welner
University of Colorado, Boulder

Abstract
This article explores the nature and implications of a 1999 decision of the Arizona Supreme Court, upholding the constitutionality of a state tax credit statute. The statute offers a $500 tax credit to taxpayers who donate money to non-profit organizations which, in turn, donate the money in grants to students in order to help defray the costs of attending private and parochial schools. The author concludes that the Arizona decision elevates cleverness in devising a statutory scheme above the substance of long-established constitutional doctrine.

This article is one of four on the Arizona Tax Credit Law:

- Moses: Hidden Considerations of Justice
- Wilson: Effects on Funding Equity
- Rud: Moral Considerations

Diverse Beliefs Within a Unitary System

Democracy, according to Chubb and Moe (1990), undermines American schooling. These authors point to a purported bureaucratic subversion of educational goals and efficiency. The market—individual choices by students and parents—would in their view drive more efficient and higher-quality schools.

While Chubb and Moe (1990) support their arguments by positively comparing Catholic schools to public schools, their book gives short-shrift to issues concerning how a market-based educational system might implicate issues of religious liberty.

Ultimately, in discussing which existing private schools should be included among those eligible to participate in the government-funded market of schools, they parenthetically offer their "own preference ... to include religious schools ..., as long as their sectarian functions can be kept clearly separate from their educational functions" (p. 219).

But the issue of religious schools is inextricably intertwined with the market-based model. Simply put, many families would—all other factors being equal—choose a sectarian education for their children. Not surprisingly, then, the recent trend toward market-based educational policies, such as public-school choice, charters, magnets, and vouchers, has prompted a new series of disputes concerning how to best balance the conflicting religious protections in the First Amendment.

Federal courts have perpetually struggled to address the tension between the First Amendment's "establishment clause" (forbidding laws "respecting an establishment of religion") and its "free exercise clause" (forbidding laws "prohibiting the free exercise thereof"). These two clauses can press courts to act in diametrically opposed directions. Such conflicting pressures are evident, for instance, in the governmental practice of exempting church property from taxation. If one's perspective is that this policy is a preference for religious institutions over secular institutions, then these exemption laws violate the establishment clause's dictate against government benefits for religion. However, if one's perspective is that the power to tax is the power to destroy, then these laws merely fulfill the free exercise dictate against burdening religious freedom.

Given this tension and the importance of the perspective of the policy-maker, government neutrality toward religion is an aspiration —a goal to strive for but one that is not realistically attainable. Stephan Carter (1993) gives the example of an Alabama law allowing schools to mandate a one-minute period of time, before the school day begins, for "meditation or voluntary prayer." This law was held by the U.S. Supreme Court to violate the establishment clause (*Wallace v. Jaffree*, 472 U.S. 38 (1985)) because it created a coercive environment promoting student prayer. Carter writes:

> And what are the likely classroom dynamics? I have nothing on which to base an empirical judgment, but I can hazard an educated guess. Many students *will* pray—we can take that as given—but if the effect on the dissenter of silent prayer during a moment when all students are silent is as coercive as the majority feared, then the Court is probably wrong to suggest that, in the absence of the moment of silence, nothing prevents those students who want to pray from doing so. After all, if the knowledge that many of one's classmates are praying during the moment of silence produces pressure to pray (and the Court may be right), then surely the knowledge that many of one's classmates are *not* praying as the school day opens will produce pressure *not* to pray. There is, in short, no neutral position (p. 191).

Faced with this dilemma, vouchers offer an attractive alternative. Instead of trying to fit all schools to all children, vouchers allow each child to select an appropriate school. This is particularly salient in the area of religious teaching, since the establishment clause clearly prohibits public schools from providing the religious education that many parents want for their children. Vouchers offer a loophole, allowing the government to assist all parents in funding their children's education, even if those parents' educational decisions are driven by religious beliefs.

But vouchers themselves are constitutionally suspect. As discussed in greater detail below, courts have placed substantial restrictions on state and local voucher plans, the more daring of which clearly run afoul of the establishment clause (as applied to the states through the due process clause of the Fourteenth Amendment).

The Arizona Law

Accordingly, given the legal instability of vouchers, Arizona's state government in 1997 passed legislation creating a non-voucher avenue of accomplishing the same

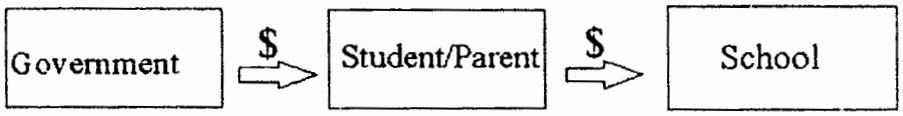
goals—allowing a state tax credit of up to $500 for donations to school tuition organizations (STOs), which would then allocate voucher-like grants to students. In full, the statute reads as follows:

> A. For taxable years beginning from and after December 31, 1997, a credit is allowed against the taxes imposed by this title for the amount of voluntary cash contributions made by the taxpayer during the taxable year to a school tuition organization, but not exceeding five hundred dollars in any taxable year. The five hundred dollar limitation also applies to taxpayers who elect to file a joint return for the taxable year. A husband and wife who file separate returns for a taxable year in which they could have filed a joint return may each claim only one-half of the tax credit that would have been allowed for a joint return.
>
> B. If the allowable tax credit exceeds the taxes otherwise due under this title on the claimant's income, or if there are no taxes due under this title, the taxpayer may carry the amount of the claim not used to offset the taxes under this title forward for not more than five consecutive taxable years' income tax liability.
>
> C. The credit allowed by this section is in lieu of any deduction pursuant to §170 of the internal revenue code and taken for state tax purposes.
>
> D. The tax credit is not allowed if the taxpayer designates the taxpayer's donation to the school tuition organization for the direct benefit of any dependent of the taxpayer.
>
> E. For purposes of this section:
>
> > 1. "Qualified school" means a nongovernmental primary or secondary school in this state that does not discriminate on the basis of race, color, sex, handicap, familial status or national origin and that satisfies the requirements prescribed by law for private schools in this state on January 1, 1997.
> >
> > 2. "School tuition organization" means a charitable organization in this state that is exempt from federal taxation under §501(c)(3) of the internal revenue code and that allocates at least ninety percent of its annual revenue for educational scholarships or tuition grants to children to allow them to attend any qualified school of their parents' choice. In addition, to qualify as a school tuition organization the charitable organization shall provide educational scholarships or tuition grants to students without limiting availability to only students of one school.

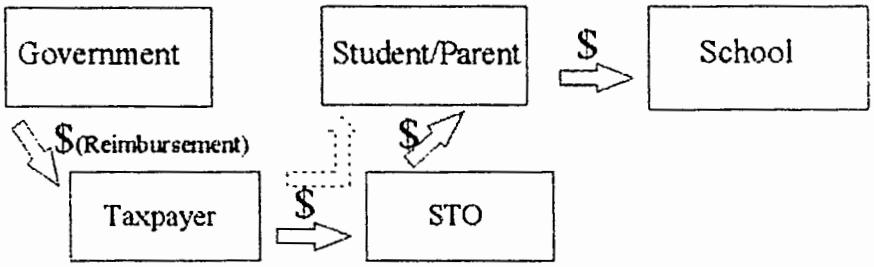
A.R.S. § 43-1089 (footnotes omitted).

In short, the mechanism created by the state of Arizona tells those who owe state taxes that they may reallocate that money from the state general fund to a scholarship-granting organization. (Note, however, that while the statute calls these grants "scholarships," they are not necessarily tied to either need or merit. (Note 1)) Whereas voucher plans entail granting state-allocated funds to schools through the private decisions of parents, the Arizona plan inserts two intermediate steps into the process. First, the grants are issued by privately-created, non-profit School Tuition Organizations (STOs), rather than directly by the government. Second, state allocation is achieved through a dollar-for-dollar tax credit given to donating taxpayers. The following flow charts illustrate the added steps:

Vouchers



Arizona Tax Credit



This Arizona system results in the government still footing the bill for all the scholarships—through directly foregone revenues (essentially reimbursing the taxpayer). But control over the funding is taken from the government and given to two other parties: (a) individual taxpayers, who can decide to which STOs they will allocate the funds, and who can earmark the funds to anyone who is not a dependent; and (b) individual STOs, which can decide the grant recipients for any non-earmarked funds. The following table outlines differences between vouchers and the Arizona tax credit.

	Vouchers	Arizona Tax Credit
Funding Ultimately From:	Government	Government
Funding Allocation Decisions Made by:	Government Officials	Private Non-Profit Organizations and Donating Taxpayers
Grants Made by:	Government	Private Non-Profit Organizations
State Money Directly Allocated to:	Schools through Parents	Self-Selected Taxpayers
Level of Regulation:	Moderate	Low

The legal challenge to the Arizona tax credit law argued that this mechanism has the same practical effect as a direct grant of general fund money in the form of vouchers. Legally, the transformation from voucher to tax credit constitutes, the argument goes, *a distinction without a difference*. Consider the statement of John Huppenthal, the Republican chair of the Arizona Senate's Education Committee, who is a longtime voucher supporter: "This has turned into something so close to vouchers you almost can't tell the difference" (Bland, 2000, A22). Or, as stated by Trent Franks, the former Arizona legislator and activist who came up with the tax credit idea, "Why do we need vouchers at this point?" (Bland, 2000, A22).

The *Kotterman* Decision & Dissent

The Arizona Supreme Court's majority opinion (the court's five justices split on a 3-2 vote) rejected challenges based on the state constitution as well as the U.S. Constitution (*Kotterman v. Killian*, 972 P.2d 606 (1999)). Below, I briefly address the arguments and decision concerning the state provisions; I then focus on the

establishment clause claims. For a more complete discussion of the Arizona constitutional issues, please see Professor Paul Bender's foreword to the *Arizona State Law Journal*, Volume 32, Number 1.

Arizona Constitution

The Arizona constitution provides that "no public money . . . shall be applied to any religious worship, exercise, or instruction or to the support of any religious establishment" (Article II, §12). It also prohibits any "tax . . . in aid of any . . . private or sectarian school . . ." (Article IX, §10). The court majority rejected arguments based on these provisions, holding instead that (a) the tax credit scheme does not give "public money," nor does it levy any "tax;" and (b) tax credits are no different from tax deductions, which have long been allowed for charitable contributions to religious institutions.

The majority's assertion that the credit does not implicate "public money" hinges on a rather formalistic definition of the term. The opinion points out that "no money ever enters the state's control or is deposited in the state treasury or other accounts under the management or possession of governmental agencies or public officials" (972 P.2d at 618). Because the state never gains actual possession or immediate control over the funds involved, these tax credits were held to not constitute public money.

The dissent calls this a "dangerous doctrine that permits the state to divert money otherwise due the state treasury and apply it to uses forbidden by the state's constitution" (972 P.2d at 640). Certainly, the state exercises a substantial degree of effective control over this money, and this control arises out of the state's power to tax. (Without the tax, the state could not direct taxpayer donations to STOs.) This aspect of the dissent relies on a line of scholarship that explains how tax credits are analogous to government expenditures. This "tax expenditure" doctrine looks at the practical effect of the credits and determines that they are the equivalent of direct government grants (both are charges made against the state treasury).

The majority rejects the tax expenditure approach, which it argues assumes "that the tax return's purpose is to return state money to taxpayers" (972 P.2d at 618):

> For us to agree that a tax credit constitutes public money would require a finding that state ownership springs into existence at the point where taxable income is first determined, if not before. The tax on that amount would then instantly become public money. We believe that such a conclusion is both artificial and premature. It is far more reasonable to say that funds remain in the taxpayer's ownership at least until final calculation of the amount actually owed to the government, and upon which the state has a legal claim. (972 P.2d at 618, footnotes omitted.)

The majority also defends the tax credit based on an analogy to tax *deductions* for charitable contributions to religious institutions, the constitutionality of which have never been seriously questioned. "If credits constitute public funds," the court argues, "then so must other established tax policy equivalents like deductions and exemptions" (972 P.2d at 618). In response, the dissent points to "very significant differences between valid tax benefits and the Arizona tax credit" (972 P.2d at 642). The latter, the dissent asserts, "is not an inducement to charitable giving; there is no philanthropy at all because the credit provided is dollar-for-dollar" (972 P.2d at 642). Because a taxpayer's $500 donation is rebated in full as a credit against the tax that otherwise would be paid to the state, the dissent views the donation more as an allocation of state money than of private money. "Unlike neutral deductions [available for all charitable giving], the credit is not the state's passive approval of taxpayers' general support of charitable institutions" (972 P.2d at 643).

To illustrate, the dissent explains the effective difference between a tax credit and a tax deduction.

> A couple with an income of $60,000 per year sending $500 to an STO would receive a tax credit of $500 and would thus save $500 in taxes. The "contribution" would cost them nothing. The same couple, contributing to almost any other qualified philanthropic cause, would

receive a deduction from gross income. To reduce their state taxes by $500, that couple would need to contribute approximately $13,000. (972 P.2d at 643, n. 18.)

For the majority, however, this purported difference is simply a matter of degree—they see no principled basis to distinguish between the two types of benefits, given that they both amount to a reduction in amounts otherwise owed to the treasury.

Establishment Clause of U.S. Constitution

To determine whether the Arizona tax credit statute violates the federal establishment clause, the *Kotterman* court applies a three-pronged test set forth in a long-standing (although often-attacked) Supreme Court precedent: *Lemon v. Kurtzman*, 403 U.S. 602 (1971). Pursuant to this *Lemon* test, a law does not violate the establishment clause if (1) it serves a secular purpose; (2) its principal or primary effect neither advances nor inhibits religion; and (3) it does not foster an excessive government entanglement with religion. The second (primary effect) prong was the main point of dispute in the *Kotterman* case. This prong requires that a statute be "neutral on its face and in its application" and not have the "primary effect" of advancing the sectarian aims of nonpublic schools. (See *Mueller v. Allen*, 463 U.S. 388, 392 (1983); see also *Committee for Pub. Educ. & Religious Liberty v. Nyquist*, 413 U.S. 756, 788 (1973).)

These latter cases, *Mueller* and *Nyquist*, are the touchstones for the majority and the dissent, respectively. The majority argues that the Arizona law is analogous to the Minnesota law upheld in *Mueller*; the dissent argues that the Arizona law is analogous to the New York law held unconstitutional in *Nyquist*. Below, I discuss those two cases, followed by a brief discussion of a second case relied upon by the *Kotterman* majority, *Jackson v. Benson*, 218 Wis.2d 835, *cert. denied*, 119 S.Ct. 467 (1998).

Mueller v. Allen

In *Mueller v. Allen*, 463 U.S. 388 (1983), the U.S. Supreme Court upheld a Minnesota tax deduction for school expenses incurred on behalf of children attending elementary or secondary schools. The antecedent to this Minnesota law was originally passed in 1955. That law allowed parents to claim a tax deduction of up to $200. For public school students, these expenses included textbooks and transportation expenses. For private school students, these expenses also included tuition. Among the subsequent amendments to this law were occasional increases in the maximum deduction per child (e.g., in 1976, the maximum for elementary school expenses was raised to $500, with $700 allowed per child for secondary school expenses). (Note 2)

The *Mueller* Court held that these deductions benefitting parents of parochial school children did not violate the establishment clause. Applying the *Lemon* test, the *Mueller* Court held that the programs had the secular purposes of ensuring that Minnesota's citizenry is well-educated and that private and parochial schools' financial health remains sound. Further, the *Mueller* Court held that these deductions did not primarily advance the sectarian aims of parochial schools and did not excessively entangle the state in religion. As the *Kotterman* court notes, the *Mueller* Court focused heavily on distinct characteristics of the Minnesota law: (a) it was open to all parents incurring educational expenses, including those whose children attend public school; and (b) the funds did not go directly to the private schools but rather reached those schools as a result of the numerous private choices of individual parents.

In discussing this primary effect prong of *Lemon* as applied to the Arizona statute, the *Kotterman* majority draws parallels to these latter two characteristics—openness to all parents and private parental choices. Arguing that the Arizona benefits are open to all parents, the majority points to companion language in the Arizona code, which allows taxpayers to claim up to a $200 tax credit for contributions to their neighborhood public school's extracurricular activities (§ 43-1089.01). Arguing that public funds do not go directly to the private schools, the majority contends that the "primary beneficiaries of this credit are

taxpayers who contribute to the STOs, parents who might otherwise be deprived of an opportunity to make meaningful decisions about their children's educations, and the students themselves" (972 P.2d at 616). While acknowledging "ripple effects ... viewed through a wide-angle lens, radiat[ing] to infinity," the majority concludes that "[p]rivate and sectarian schools are at best only incidental beneficiaries of this tax credit, a neutral result that we believe is attenuated enough to satisfy *Mueller* and the most recent Establishment Clause decisions" (972 P.2d at 616). (Note 3)

The dissent, however, distinguishes *Mueller* as follows:

> Under the provision upheld in *Mueller*, religious schools benefitted only as a result of true choice made among a wide selection of alternatives, both public and private [citation omitted]. Under the Arizona plan, there is no real choice—one may contribute up to $500 to support private schools or pay the same amount to the Arizona Department of Revenue. In reality, this is not a choice but government action designed to induce taxpayers to direct financial support to predominantly religious schools. (972 P.2d at 629.)

The Arizona tax credit, the dissent also notes, "is available only to those who choose to support private, predominantly religious schools. Those who wish to contribute to public schools are allowed only a $200 credit, and their contributions can be used only to reimburse fees paid for extracurricular activities" (972 P.2d at 628, citation omitted).

In response to the majority's contention that public schools do not need the same benefits, since public school students do not pay tuition, the dissent points to "deficiencies of state financing of public schools and the underfinanced and unfilled educational missions of those schools [citations omitted]. If we are to consider equality or neutrality of the two credits, we must bear in mind that public schools, like private schools, need assistance to perform their educational mission" (972 P.2d at 626). Provisions, the dissent asserts, "could have been made for a tax credit for contributions supporting the educational mission of the public school system" (972 P.2d at 626).

Committee for Pub. Educ. & Religious Liberty v. Nyquist

Instead of *Mueller*, the dissent argues, the controlling precedent for the *Kotterman* case is *Committee for Pub. Educ. & Religious Liberty v. Nyquist*, 413 U.S. 756 (1973). The U.S. Supreme Court in *Nyquist* struck down a New York law providing (a) tuition grants to low- income families (vouchers redeemable only at private schools) and (b) tax deductions for tuition payments, varying by income level. The law provided no benefits aimed at families with children in public schools. Noting that the private schools in New York were predominantly religious, the *Nyquist* Court stated that grants "offered as an incentive to parents to send their children to sectarian schools by making unrestricted cash payments to them [violate] the Establishment Clause ... whether or not the actual dollars given eventually find their way into the sectarian institutions. Whether the grant is labeled a reimbursement, a reward, or a subsidy, its substantive impact is still the same." 413 U.S. at 786.

The *Kotterman* dissent characterizes the Arizona law as similarly providing benefits aimed only at private school contributions, doing so in an unregulated manner likely to lead to abuse:

> Because Arizona's tax credit statute does not require that grant use be restricted to the secular aspects of education, the STOs' grants to private schools may be used in any manner the recipient school wishes. Nor does the statute prevent an STO from directing all of its grant money to schools that restrict enrollment or education to adherents of a particular religion or sect. Moreover, there is no limit on the dollar amount the STO can give to a school on behalf of a student. Thus, an STO could pool several contributions and then pay the full tuition for any student, group of students, or for that matter, all students in any group of schools

of a single religious faith. (972 P.2d at 630.)

In contrast, the majority perceives "safeguards built into the statute," such as "the way in which an STO is limited, the range of choices reserved to taxpayers, parents, and children," and the system's "neutrality" (972 P.2d at 620). These safeguards, the court reasons, result in an attenuation of any benefits received by religious schools and "ensure that the benefits accruing from this tax credit fall generally to taxpayers making the donation, to families receiving assistance in sending children to schools of their choice, and to the students themselves," rather than to those schools (972 P.2d at 620).

The dissent, in comparing *Mueller* and *Nyquist*, first notes that *Nyquist*, like *Mueller*, involved a scheme whereby the state funds went initially to parents and then to schools of the parents' choosing. The dissent then focuses on a point of contrast:

> In *Mueller*, the Court upheld a Minnesota law allowing a deduction, in part because it was "available for educational expenses incurred by all parents including those whose children attend public schools." Making the benefit available to this neutral and "broad class" is an "important index of secular effect." The Court said the Establishment Clause does "not encompass the sort of attenuated financial benefit . . . that eventually flows to parochial schools from the neutrally available tax benefit at issue . . . ." Indeed, the *Mueller* Court described *Nyquist's* unconstitutional, nonneutral, private school program in words directly applicable to the Arizona: "thinly disguised 'tax benefits,' actually amounting to tuition grants, to the parents of children attending private schools," the majority of which were sectarian" (972 P.2d at 627-28, citations omitted).

The dissent then notes that at least seventy-two percent of private schools in Arizona are sectarian, and it concludes that the Arizona law "is everything *Nyquist* held unconstitutional—a direct stipend that has the primary effect of advancing religion by tuition grants to religious schools" (972 P.2d at 628).

In contrast, the majority's primary basis for distinguishing *Nyquist* focuses on the "broad class of citizens" to whom the Arizona tax credit is available (972 P.2d at 613). That is, while the New York benefits were available only to parents who sent their children to private school, the Arizona benefits are available to all taxpayers. The Arizona credit is not limited only to parents, let alone just those parents of private school students. "Thus, Arizona's class of beneficiaries is even broader than that found acceptable in *Mueller*, and clearly achieves a greater level of neutrality" (972 P.2d at 613).

Jackson v. Benson

As briefly mentioned above, the *Kotterman* majority supplemented its reliance upon *Mueller* with a discussion of the recent opinion of the Wisconsin Supreme Court in *Jackson v. Benson*, 218 Wis.2d 835, 578 N.W.2d 602 (1998). The Wisconsin court upheld the constitutionality of a voucher plan directed at low-income students in the Milwaukee Public Schools (MPS). (Note 4)

The Milwaukee Parental Choice Program (MPCP), which began in 1989, includes the following provisions: (1) students may use the voucher at the private or parochial school of their choice; (2) the amount of the voucher is the lesser of two numbers: the private or parochial school's operating and debt service cost per pupil or the state's per-pupil aid to the MPS (about $4,900); (3) students qualify for vouchers if their family income is not greater than 1.75 times the poverty level and if they meet certain enrollment requirements (e.g., during the previous school year, they were enrolled either in the MPS, in a private school in Milwaukee, in grades K-3 in a private school outside of Milwaukee, or were not enrolled in any formal school); (4) Wisconsin sends a check directly to the school but made out to the parents, who endorsed it over to the educational institution; (5) participating schools must notify Wisconsin of their intention to participate in the program, comply with certain laws and meet at least one of four legislatively-established performance

standards; and (6) no more than 15% of the school district's enrollment may attend participating schools in any school year. As of the 1998-1999 school year, 6,194 students were participating in the program, far below the ceiling of approximately 15,000 students technically allowed to participate.

In 1998, the Wisconsin Supreme Court upheld the program as constitutional (*Jackson v. Benson*, 218 Wis.2d 835, 578 N.W.2d 602 (1998)), and the U.S. Supreme Court denied a *writ of certiorari* (119 S.Ct. 467 (1998)), meaning that the state supreme court opinion was allowed to stand, but without the U.S. Supreme Court's approval (or rejection). (Note 5) While a Wisconsin state opinion is not binding precedent upon an Arizona court, such an opinion may prove persuasive. The *Kotterman* majority found the following language particularly persuasive:

> In our assessment, the importance of our inquiry here is not to ascertain the path upon which public funds travel under the amended program, but rather to determine who ultimately chooses that path. As with the programs in *Mueller* and *Witters*, not one cent flows from the State to a sectarian private school under the amended MPCP except as a result of the necessary and intervening choices of individual parents. *Jackson v. Benson*, 578 N.W.2d at 618.

In Arizona, the *Kotterman* majority reasons, the decision-making process preceding the scholarship allocation is "completely devoid of state intervention or direction" (972 P.2d at 614):

> Arizona's statute provides multiple layers of private choice. Important decisions are made by two distinct sets of beneficiaries—taxpayers taking the credit and parents applying for scholarship aid in sending their children to tuition-charging institutions. The donor/taxpayer determines whether to make a contribution, its amount, and the recipient STO. The taxpayer cannot restrict the gift for the benefit of his or her own child. A.R.S. § 43-1089(D). Parents independently select a school and apply to an STO of their choice for a scholarship. Every STO must allow its scholarship recipients to "attend any qualified school of their parents' choice," and may not limit grants to students of only one such institution. A.R.S. § 43-1089 (E) (2) (emphasis added). Thus, schools are no more than indirect recipients of taxpayer contributions, with the final destination of these funds being determined by individual parents. (972 P.2d at 614).

For its part, the *Kotterman* dissent found the *Jackson* opinion to be of little persuasive value. The Wisconsin statute, the dissent notes, includes an "opt-out" provision, pursuant to which students may be excused from the religious aspects of schooling at sectarian institutions. Similarly, Wisconsin requires schools receiving grants to admit applicants without regard to the applicants' religious or nonreligious preference. (Note 6) Wisconsin's statute also explicitly limits state support to private institutions' educational (as opposed to religious) programs. Finally, unlike Arizona's system, which is weighted in favor of wealthier taxpayers and provides no incentives for STOs to consider wealth as a scholarship criterion, the Wisconsin program is designed to provide greater choice to low-income families:

> Arizona's statute . . . contains no religious instruction opt-out provision, appears to permit religious discrimination, permits funding of religious observance, and makes the tax credit available to all taxpayers, those who have children in school and those who do not, the rich and the poor. Further, our statute makes no limitation on the amount of funding a school can receive from an STO for a particular student. Wisconsin, in short, has made some attempt, successful or not, to limit the use of state subsidies for religious instruction and ceremony. Arizona's program, on the other hand, will inevitably and primarily benefit religious observance and instruction. (972 P.2d at 631).

Circumventing the Constitution?

The Arizona tax credit law provides a government subsidy for taxpayers who wish to support religious activities. The state supreme court upheld the law even though more direct support by the government is constitutionally forbidden. The circumventing nature of the law is pointed out by the *Kotterman* dissent, which warns that it allows private and religious STOs to provide scholarships to current private school parents, essentially turning the donated money into tuition rebates. "Further," the dissent adds, while the law prohibits the STOs "from making grants to 'only students of one school,' the statute does not prevent an STO from directing all of its grant money to a group of schools that restrict enrollment or education to a particular religion or sect" (972 P.2d at p. 626). That is, "nothing forbids an STO from limiting its grants or scholarships to students who adhere to a particular religion and will participate in the required religious observance" (972 P.2d at p. 626). This enables the formation of STOs devoted to the supports of a particular religious belief.

In fact, groups like the "Arizona Christian School Tuition Organization" (ACSTO) have formed in order to target donors interested in supporting scholarships to schools with particular beliefs (in this case, evangelical Christianity). In its first year (1998), the ACSTO raised over a half-million dollars, second in the state only to the Catholic Tuition Organization of Phoenix (CTOP), the STO formed by the Roman Catholic Diocese, which raised more than $837,000 (Schnaiberg, 1999; Center for Market-Based Education and the Goldwater Institute, 2000). In 1999, these amounts increased dramatically, to over $2.8 million for the ACSTO and almost $4.7 million for the CTOP (Bland, 2000). Overall, $1.8 million was raised in 1998 by a total of fifteen tuition organizations (Center for Market-Based Education and the Goldwater Institute, 2000), and over $13.3 million was raised in 1999 by a total of twenty-nine STOs (Bland, 2000).

Even though the STOs cannot control parents' school choices, they can target parents based on their knowledge of those parents' inclinations. The president of the ACSTO, when asked if the group had ever had a parent not choose a Christian school, responded that this had never happened: "I don't know what we'll do when we see that," he said. "The people coming to us know who we are and that we're interested in giving scholarships to kids to go to these schools" (Schnaiberg, 1999).

Moreover, parents have found a huge loophole in the legislation, which prohibits donors only from earmarking money for their own dependents. According to article in the Arizona Republic, "parents are writing $500 checks for their friends' kids and asking them to do the same for theirs" (Bland, 2000, A22). The newspaper identified one fund for which 96% of all donations were earmarked for specific private school students.

The troubling nature of this scheme does not escape the attention of the *Kotterman* dissent. It points out that the majority opinion's reasoning leaves no principled reason why the limit could not be increased far beyond $500, to pay the full cost of private, sectarian education. (Note 7) Accordingly, the dissent attacks the tax credit as "directed so that it supports only the specific educational institutions the Arizona Constitution prohibits the state from supporting—predominantly religious schools":

> By reimbursing its taxpayers on a dollar-for-dollar basis the state excuses them from paying part of their taxes, but only if the taxpayers send their money to schools that are private and predominantly religious, where the money may be used to support religious instruction and observance. If the state and federal religion clauses permit this, what will they prohibit? Evidently the court's answer is that nothing short of direct legislative appropriation for religious institutions is prohibited. If that answer stands, this state and every other will be able to use the taxing po wer to direct unrestricted aid to support religious instruction and observance, thus destroying any pretense of separation of church and state. (972 P.2d at 645).

The Slippery Slope

The rationale of the *Kotterman* majority would seem to allow for the positive check-off system presently included on federal 1040 forms to fund the Presidential Campaign Fund. Using all the present Arizona STOs (see Bland, 2000, A22), I can envision something like the following potentially appearing on Arizona tax forms:

Would you like up to $500 of your tax payment to be diverted to a School Tuition Organization?

Yes	No	Note: checking "yes" will not increase your taxes owed or reduce your refund. The amount allocated will simply be deducted from the funds that would otherwise go to the state general fund.

If yes, please allocate per your wishes among the following STOs:

Arizona Adventist Scholarships, Inc	$
Arizona Christian School Tuition Organization, Inc.	$
Arizona Episcopal Schools Foundation	$
Arizona Scholarship Fund	$
Arizona Independent Schools Scholarship Foundation	$
Arizona Native Scholastic and Enrichment Resources	$
Arizona Private Education Scholarship Fund	$
Arizona School Choice Trust	$
Brophy Community Foundation	$
Catholic Tuition Organization of the Diocese of Phoenix	$
Catholic Tuition Organization of the Diocese of Tuscon	$
Christian Scholarship Fund of Arizona	$
Educare Scholarship Fund	$
Florence Englehardt/Pappas Foundation	$
Foundation for Montessori Scholarships	$
High Education for Lutherans Program (HELP) Foundation, Inc.	$
Institute for Better Education	$
Jewish Community Day School Scholarship Fund	$
Lutheran Education Foundation	$
Maranath Christian Co_Op Tuition Fund	$
Montessori School Tuition Organization	$
Northern Arizona Christian School Scholarship Fund	$
Patagonia Scholarship Fund	$
Prescott Christian School Scholarship Foundation	$
School Tuition Association of Yuma	$
Schools with a Heart Foundation	$
Southern Arizona Foundation for Education	$
VVBC Christian Education Fund	$
Walter T Beamis Scholarship Foundation	$

Such a form, of course, could become unwieldy as more STOs are created and if the form also included every public school (in connection with the $200 option for

extracurricular activities). But the logistics are hardly insurmountable.

Mitchell v. Helms

Many scholars who follow the U.S. Supreme Court are guessing that a case arising out of Ohio (see the Appendix to this paper for a description of this case, called *Simmons-Harris v. Zelman*, 72 F. Supp. 2d 834, N.D. Ohio, 1999), will eventually end up before the Court. If this happens, much of the uncertainty surrounding the constitutionality of vouchers (and, perhaps, voucher-like alternatives) might finally be resolved. For the immediate future, however, onlookers are closely reading the Justices' three opinions *Mitchell v. Helms*, 120 S. Ct. 2530 (2000), decided on June 28, 2000.

The Mitchell case arose out of a challenge to Chapter 2 of Title I of the Elementary and Secondary Education Act of 1965, which allows state education agencies to distribute "secular, neutral, and nonideological services, materials, and equipment" to students who are enrolled in private nonprofit elementary and secondary schools. In 1994, Congress enacted the "Improving America's Schools Act," which provides for loans of, among other things, taxpayer-funded computers to parochial schools (20 U.S.C. § 7301-73). This legislation was challenged by parents in Louisiana, and a federal appeals court agreed that the provision of computers violated the establishment clause (*Helms v. Picard*, 151 F. 3d 347, 5th Cir. 1998).

Although the facts of Helms are not directly connected to vouchers, the Supreme Court's deliberations were watched closely by those concerned about vouchers' constitutionality. Voucher supporters filed an amicus brief, urging the court to use the case to pave the way for vouchers to pass First Amendment muster. Justice Thomas obliged, writing an opinion clearly implying vouchers' constitutionality; but he was able to get only three other Justices to join in his opinion. (Thomas' opinion goes so far as to equate a refusal to aid religious schools with hostility toward religion.) Two concurring justices refused to go along with this judicial activism, issuing a much narrower opinion. (In addition, three Justices dissented.)

The concurrence, written by Justice O'Conner, upheld the law on the narrow ground that it does not define recipients by reference to religion, instead using neutral and secular criteria to allocate aid to students enrolled in religious and secular schools alike. O'Conner pointed out that, like the law challenged in *Agostini v. Felton*, 521 U.S. 203 (1997), Chapter 2 allocates aid on the basis of neutral, secular criteria; it is supplementary to, and does not supplant, non-federal funds. She concludes, "no Chapter 2 funds reach the coffers of religious schools; the aid is secular; evidence of actual diversion is *de minimis*; and the program includes adequate safeguards" (p. 133 of the Court's slip opinion).

Because O'Conner's opinion represents the "swing votes" on the present Court, it sets forth the governing law for the moment. Whether the Thomas position is eventually joined by the one additional vote needed to constitute a majority will likely depend upon who is appointed to the Court by the nation's next President.

Federal Court Challenge

On February 15, 2000, the Arizona chapter of the ACLU filed, in the federal district court in Arizona, a new and separate challenge to the tax credit statute (*Winn v. Killian*, case no. civ00-0287-phx-che). Because federal courts have ultimate authority and responsibility for interpreting the federal Constitution, this district court is in no way bound to follow the Arizona Supreme Court's decision as regards the First Amendment's Establishment Clause.

The state has moved to dismiss this federal lawsuit, putting forth claims asserting sovereign immunity and purported protection provided by a federal statute (the Federal Tax Injunction Act). To date, this dismissal motion is pending—and denial of the sovereign immunity argument may prompt an interlocutory (i.e., immediate) appeal. This federal action, therefore, may not be resolved for many years.

Conclusion

Imagine a law establishing the Gideons' religious organization as the "Official Church of the U.S.A." Such a law would strike at the heart of the constitutional prohibition against laws "respecting an establishment of religion." Upon challenge, it would be declared unconstitutional. Now imagine a law providing government grants to religious organizations that provide reading materials for hotel rooms. This law, too, would quickly be seen as violating the establishment clause, because its principal or primary effect advances religion. (See *Lemon v. Kurtzman*, 403 U.S. 602, 612-13 (1971), discussed in greater detail below.) Finally, imagine a law that provides a dollar-for-dollar tax credit to individuals who donate money to organizations that then grant the money to those who provide reading materials for hotel rooms. Although the Gideons would almost surely be the main beneficiary of this law, the reasoning of the majority in *Kotterman*, would seem to allow this last law to withstand a constitutional challenge.

The fact that Arizona's government might create a mechanism to encourage (actually, reimburse) targeted giving to religious organizations did not trouble the court. Yet, viewed in terms of effects, the practical distinction between the tax credits and a direct allocation (vouchers) is that the latter allocation is through representative democracy and the former is through direct democracy—*with the wealthy entitled to more votes*. Consequently, the tax credit mechanism results in the allocation of presumptive tax dollars to support those institutions (religious or otherwise) that are most popular with the state's wealthiest residents.

The hypothetical hotel-reading-material law is distinguishable from the Arizona tax credit law in at least two important ways. First, it does not necessarily serve a secular purpose. The *Kotterman* court found the Arizona law to serve the legitimate secular purpose of "bring[ing] private institutions into the mix of educational alternatives open to the people of this state," assuring the continued financial health of private schools, and producing "healthy competition" for public schools (972 P. 2d at 611). If a law does not have a secular purpose, then it violates the establishment clause whether or not its primary effect advances religion (recall that a law must have a secular purpose, as an independent "prong" of the *Lemon* test).

Second, the hypothetical hotel-reading-material is more difficult to characterize as an attempt to treat religious institutions in a neutral, accommodating way. Neutrality toward religion has long been a guiding principle of First Amendment jurisprudence. The evolution of Supreme Court decisions—and its recent modifications—can be understood as an evolution in how the Court's majority defines that neutrality. Three decades ago, the Arizona tax credit law would almost surely have been considered by the Supreme Court to provide an unconstitutional and extraordinary benefit to private, religious schools. Now, the Court may view that same law as a reasonable accommodation for the beliefs and needs of residents who feel ill-served by the public schools.

The Arizona Supreme Court grounds its *Kotterman* decision in such a neutrality argument: Basic education is compulsory for children in Arizona, but until now low-income parents may have been coerced into accepting public education. These citizens have had few choices and little control over the nature of their children's schooling because they could not afford a private education more compatible with their values and beliefs. Arizona's tax credit achieves a higher degree of parity by making private schools more accessible and providing alternatives to public education (972 P. 2d at 615). The court also notes that helping to pay for private school tuition helps to balance out the fact that the state already pays the cost of students' attendance at public schools. Such rationales (i.e., such definitions of "neutrality") if carried to their logical conclusion will carry the nation toward the privatization ideals of Milton Friedman (1963, 1990). As the *Kotterman* dissent points out, if the majority's interpretation of the First Amendment holds, then the government can use its taxing power (through tax credits) to direct unrestricted aid to support churches and other religious organizations. This could lead to a revolution in American schooling, and it is one that many fear will wipe out the educational and equity gains of the last century.

Notes

1. In fact, the wealthiest students appear to be receiving the vast majority of the law's benefits (Bland, 2000; Wilson, 2000). This is as true of the $200 donations to the public school fund as it is of $500 donations to the private school funds (Bland, 2000). Some funds, however, including the Catholic Tuition Organization, do means-test for their scholarships.
2. Presently, the Minnesota law allows a maximum deduction of $1,625 for elementary school expenses and $2,500 for secondary school expenses. This amendment was passed in 1997, along with an expansion in the types of expenses that the deduction covers, adding academic summer camps, summer school and up to $200 of the cost of a personal computer and education software. Further, persons who do not itemize deductions on their federal income tax form can now take the deduction. Perhaps most notably, the 1997 amendments created a refundable tax *credit* for families with incomes under $33,500 (now $37,500): up to $1,000 per student or $2,000 per family. (If a family owes no taxes or owes less than the amount of the credit, they receive the difference as a refund.) The credit is available for the same education expenses as the deduction (textbooks, transportation, academic summer camps, summer school and up to $200 of the cost of computer hardware and education software), except that it does *not* cover tuition. Expenses that exceed the credit amount may be used as a tax deduction.
3. Consider the governmental activities that have been upheld by the Court. Mitchell v. Helms (2000), discussed in the main text, upheld funding of hardware and software loans to public and parochial schools. Agostini v. Felton, 521 U.S. 203, 222 (1997), upheld government-funded remedial instruction in parochial schools. Other past cases have upheld government aid for a sign language interpreter for a deaf student attending a Catholic high school (*Zobrest v. Catalina Foothills Sch. Dist.*, 509 U.S. 1 (1993); government reimbursement to religious schools for the grading of tests that were prepared, mandated, and administered by the state (*Committee for Pub. Educ. & Religious Liberty v. Regan*, 444 U.S. 646 (1980); government reimbursement to the parents of parochial school students for the cost of public transportation to and from school (*Everson v. Board of Educ.*, 330 U.S. 1 (1947); and government aid in providing non-religious textbooks for students in parochial schools (*Meek v. Pittenger*, 421 U.S. 349 (1975); *Board of Educ. v. Allen*, 392 U.S. 236 (1968)).
4. The Milwaukee plan is the oldest surviving publicly funded voucher scheme. Several cities, however, including Washington D.C., New York City, Baltimore, and Dayton, Ohio, have privately-funded voucher plans. The most ambitious efforts are through the "Children's Scholarship Fund," which has already provided more than 40,000 "scholarships."
5. The U.S. Supreme Court similarly denied a *writ of certiorari* petition in the *Kotterman* case.
6. Arizona's laundering of state money through several intermediate steps certainly does serve to disentangle the government from those religious activities of parents and institutions that ultimately benefit from the government largess. Compare, on the one hand, the Milwaukee voucher program, which involves government monitoring to ensure that participating schools do not discriminate in admissions on the basis of religion and do not require vouchered students to participate in religious activities. The Arizona system, on the other hand, requires only that schools not "discriminate on the basis of race, color, sex, handicap, familial status, or national origin" (§43-1089(E)(1))—discrimination on the basis of religious adherence, preference, or observance is perfectly permissible.
7. The dissent notes that Arizona's tax credit statute actually has another loophole, allowing taxpayers a chance to make a profit: "After a taxpayer has contributed to the STO and received a dollar-for-dollar refund from the Arizona Department of Revenue, nothing in the Internal Revenue Code prevents him or her from reporting the contribution as a charitable deduction on the federal income tax return" (972 P.2d at 642, n. 17).
8. Some of the below discussion presents information provided on-line by the

Educational Commission of the States (ECS) at http://www.ecs.org/ecs/ecsweb.nsf.

9. The unpublished opinion is available in pdf format at http://www.ij.org/cases/index.html.

References

Bland, K (2000, April 9). School tax credits wide open to abuse: Millions are diverted from needy students. *The Arizona Republic*, A1.

Bowman, D. H. (2000, March 1). States giving choice bills a closer look. *Education Week*. 19(25), 1, 24.

Carter, S. (1993). *The Culture of Disbelief: How American Law and Politics Trivialize Religious Devotion*. New York: Basic Books.

Center for Market-Based Education and the Goldwater Institute (2000). *EXTRA-CREDIT* newsletter. [On- line.] Available at: **http://www.cmbe.org/extracredit.htm**.

Chubb, J. & Moe , T. (1990). *Politics, Markets, and America's Schools*. Washington DC: Brookings Institution Press.

Friedman, M. (1963). *Capitalism and Freedom*. Chicago: Univ. of Chicago Press.

Friedman, M. & Friedman, R. (1990). *Free to Choose: A Personal Statement*. San Diego: Harcourt Brace & Company.

Schnaiberg, L. (1999, April 14). As tax day nears, Arizona prepares to tally credits. *Education Week*, *18*(31), 17.

Wilson, G. Y. (2000). *Effects of Funding Equity on the Arizona Tax Credit Law*. Paper presented at the annual meeting of the American Educational Research Association, in New Orleans, LA, April 24-28, 2000.

Appendix
Other Voucher and Tax Credit Plans and Proposals

The voucher and tax credit plans discussed in the main text amount to just a small sampling of the plans underway nationally. Moreover, the pace of reform has recently intensified; this year's legislative sessions feature at least 21 states with bills to start voucher programs and 18 considering proposals that would offer tax breaks to help cover the costs of private schooling (Bowman, 2000). Some of these follow usual voucher formats, some follow the Arizona model, and some tie vouchers to school performance—patterned after Florida's plan. This Appendix provides some context for the Arizona tax credit scheme by offering an overview of these other voucher and tax credit plans and proposals. (Note 8)

Florida

The nation's only statewide voucher program was approved in Florida in the summer of 1999, but it was almost immediately held by a state court to violate the Florida constitution (*Holmes v. Bush*, No. 99-3370, Fla. Cir. Ct., filed June 22, 1999). (Note 9) Under the plan, each public school was to receive a grade, from A to F. Students at schools that earn a grade of "F" from the state two years out of four would be eligible for an "opportunity scholarship" worth at least $4,000 that could be used at a public, private, or religious school. In the first year, only two schools "qualified," both of them in Pensacola (Bowman, 2000). Private and parochial schools that might have accepted these students would have been prohibited from collecting additional

tuition and barred from requiring these students to participate in religious instruction, prayer or worship.

In its decision handed down on March 14, 2000, the Florida state court relied on the state's education clause (language passed by voters in 1998) in holding that vouchers supporting attendance at private schools would unconstitutionally undermine Florida's goal of providing a free public education. This education clause provides in part,

> Adequate provision shall be made by law for a uniform, efficient, safe, secure, and high quality system of free public schools that allows students to obtain a high quality education and for the establishment, maintenance, and operation of institutions of higher and other public education programs that the needs of the people may require. (Florida Constitution, Article IX, section 1.)

The court's reasoning is expressly grounded in a long- established constitutional principle in Florida that, when the constitution directs "how a thing shall be done, [this direction] is itself a prohibition against a different manner of doing it" (*Holmes v. Bush*, at p. 7). That is, because the education clause directs that the state's educational goals shall be obtained through free public schools, the use of vouchers (to private schools) to achieve this same aim is implicitly prohibited. The court therefore concluded,

> the statute provides that all students at designated schools who wish to do so may leave the public school system and instead receive their publicly funded education in private schools that offer the same services as do the public schools. This program supplants the system of free public schools mandated by the Constitution. (*Holmes v. Bush*, at p. 14.)

Given that most other states have education clauses similar to the above-quoted clause in the Florida constitution, this decision potentially has far-reaching ramifications.

Ohio

In 1995, Ohio created a scholarship and tutoring program in Cleveland. The program included the following provisions, which are similar to those of the MPCP: (a) the amount of the scholarship is the lesser of two numbers: the public, private or parochial school's tuition or a state-established amount not in excess of $2,500; (b) students whose family income is below 200% of the maximum level (established by the state superintendent of public instruction) for low-income families qualify for 90% of the scholarship amount; (c) students whose family income is at or above 200% of that level qualify for 75% of the scholarship amount; (d) students may use the vouchers at the public, private or parochial school of their choice; (e) participating schools must register with the state superintendent of public instruction; and (f) no more than 25% of the scholarships can be awarded to students enrolled in a private or parochial school at the time they apply for a scholarship, although the enabling legislation allows that proportion to eventually rise to 50%.

This original legislation was struck down in 1999 by the Ohio Supreme Court as unconstitutionally enacted (i.e., a technical flaw, not directly concerning the constitutionality of the legislation's contents) (*Simmons-Harris v. Goff*, 711 N.E.2d 203, 1999). The Ohio court, however, also stated that the program did not breach the separation of church and state in either Ohio or federal law. Accordingly, the legislation was (properly) re-enacted, then challenged in federal court—following the same pattern that we now see in Arizona. This new lawsuit was successful. Just seven months after similar legislation was stated to be constitutional by the Ohio Supreme Court, the federal district court disagreed, ruling that it violates the federal establishment clause (Simmons-Harris v. Zelman, 72 F. Supp. 2d 834, N.D. Ohio, 1999). That decision is presently on appeal.

Illinois

In 1999, Illinois enacted legislation granting tax credits to parents of

children in public, private or parochial schools. Under the law, parents may reduce their state income tax bill by 25 percent of whatever they spend for their children's tuition, book fees, and lab fees. In order to be eligible for the tax credit, parents must spend at least $250, and the tax credit may not exceed $500 per family. Illinois' tax credit program is presently being challenged in court (*Griffith v. Bower*, No. 99-CH-0049, Ill. Cir. Ct., filed July 12, 1999).

Iowa

In 1987, Iowa enacted a law that allowed parents with a net income of less than $45,000 to claim a tax deduction of up to $1,000 for each dependent's acceptable education expenses. These acceptable expenses include tuition and textbooks but exclude the costs of religious materials. The state has since shifted from a deduction to a tax credit, and the income ceiling has since been eliminated. All parents may now claim a tax credit of up to 25% of the first $1,000 for each dependent's acceptable education expenses.

Puerto Rico

Pursuant to a 1993 Puerto Rico law, parents with annual incomes of less than $18,000 may receive vouchers for up to $1,500 toward tuition at the public, private or parochial school of their choice. However, the Puerto Rico Supreme Court ruled in 1994 that this voucher program violated Puerto Rico's constitution. In 1995, however, Puerto Rico established the "Educational Foundation for the Free Selection of Schools, Inc," a nonprofit corporation which provides financial aid for elementary and high school students in public, private or parochial schools.

Donors to the Educational Foundation are eligible for a tax credit up to $250 for individual taxpayers or $500 for corporations and partnerships. The amount of donations in excess of the credit can be used as a tax deduction. The program includes the following provisions: (a) the annual income of a student's family cannot exceed $18,000; (b) the amount of education financial aid cannot exceed $1,500 per student; and (c) participating schools must be licensed by the General Council of Education and have an admission policy free of discrimination.

Vermont and Maine

Given their large areas containing small populations, Vermont and Maine have both enacted legislation allowing students with no nearby public school to attend private, non-parochial schools at state expense. Both programs have survived legal challenges to the exclusion of parochials from their programs.

In Maine, both the Supreme Judicial Court of Maine (*Bagley v. Raymond School Department*, 728 A.2d 127, 1999) and the U.S. Court of Appeals for the 1st Circuit (*Strout v. Albanese*, 178 F.3d 57, 1999), in two separate cases, have ruled that the exclusion does not violate parents' right of free exercise of religion and that the inclusion of religious schools in the program would violate the federal constitution's establishment clause. The Vermont case arose out of the 1996 decision by the town of Chittenden to pay the parochial school tuition for about a dozen families. In 1999, the Vermont Supreme Court ruled that Chittenden's program violated the clause of the Vermont constitution prohibiting "compelled support" of places of religious worship (*Chittenden Town School District v. Vermont Department of Education*, 738 A.2d 539 (1999)).

Pennsylvania

In 1998, the Southeast Delco School District, located near Philadelphia, Pennsylvania, adopted a voucher plan reimbursing—up to $1,000 annual tuition per child—parents who send their children to private and religious schools. On December 23, 1999, the Commonwealth Court of Pennsylvania unanimously upheld a lower court's ruling that, under Pennsylvania law, a local school board has no authority to initiate such a plan (*Giacomucci v. Southeast Delco Sch. Dist.*, 742 A.2d 1165 (1999)).

Ballot Measures

Ballot initiatives designed to create statewide voucher systems have failed in Michigan (1978), Oregon (1990), Colorado (1992) and California (1993). However, similar efforts continue in all these states. In fact, a Michigan group called "Kids First! Yes!," announced in February that it had collected the signatures necessary for a statewide vote in November on its initiative to allow vouchers for private and religious schools (Bowman, 2000).

Proposed Legislation

In Connecticut, a tuition tax credit bill has been referred to committee. In Virginia, the legislature recently tabled—until next year—a proposal that would allow parents of private school students to receive state income- tax credits, starting at $500 in 2001 and increasing to $2,500 over five years. Legislators in at least seven states—California, Colorado, Georgia, New Mexico, Pennsylvania, Vermont, and Washington—have proposed legislation similar to Florida's voucher law, although these efforts likely lost some steam after the Florida court's unfavorable decision. In New York City, Mayor Giuliani included $6 million in this year's budget plan for an experimental voucher program.

In Congress, Republican leadership in both houses have, in every recent session, been pushing for vouchers. For instance, in the 106th Congress, Senator Jon Kyl (R-AZ) introduced a tuition tax credit bill (S.138) in the U.S. Senate for K-12 expenses. It would have given a tax credit to parents for their children's educational expenses and to other individuals who contribute to a nonprofit scholarship program to fund education for low-income students. The bill would phase in a credit up to $250 per individual (or $500 per joint return) by 2002. In the U.S. House of Representatives, Congressman Jim Rogan (R-CA) introduced a similar bill (H.R. 600) which allowed a much larger credit of $1000 per individual. When last I checked (in March, 2000), the Senate bill had been referred to the finance committee; the House bill had been referred to the ways and means committee.

About the Author

Kevin G. Welner
School of Education
University of Colorado
Boulder, CO 80309-0249
(303) 492-8370

E-mail: kevin.welner@colorado.edu
B.A. University of California, Santa Barbara, 1985
J.D. University of California, Los Angeles, 1988
Ph.D. University of California, Los Angeles, 1997

Kevin Welner is an Assistant Professor of policy, law, and evaluation at the University of Colorado at Boulder's School of Education. His forthcoming book is entitled, *Legal rights, local wrongs: When community control collides with educational equity*, published by SUNY Press. His homepage is at http://www.colorado.edu/education/contacts/faculty/welner.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary

Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina — Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doc.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

BEST COPY AVAILABLE

This article has been retrieved 118 times since August 1, 2000

Education Policy Analysis Archives

Volume 8 Number 37 August 1, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

The Arizona Education Tax Credit and Hidden Considerations of Justice: Why We Ought to Fight Poverty, Not Taxes

Michele S. Moses
Arizona State University

Abstract
The current debate over market-based ideas for educational reform is examined, focusing specifically on the recent movement toward education tax credits. Viewing the Arizona education tax credit law as a voucher plan in sheep's clothing, I argue that the concept of justice underlying the law is a crucial issue largely missing from the school choice debate. I question the libertarian conception of justice assumed by voucher and tax credit advocates, and argue instead that a contemporary liberal democratic conception of justice ought to undergird attempts at school reform. A call for educators and policymakers to concentrate energies on efforts to help needy students rather than on efforts to channel tax dollars toward self-interested ends concludes the article.

This article is one of four on the Arizona Tax Credit Law:

- Welner: Taxing the Establishment Clause
- Wilson: Effects on Funding Equity
- Rud: Moral Considerations

Why will conservative politicians and policymakers put money and effort behind voucher plans – a reform idea wholly lacking evidentiary support – yet they will make symbolic efforts at best to break the cycle of poverty, which is perhaps the most serious problem facing the United States in general and public schools in particular? Worse yet, these same persons claim to support vouchers primarily because they will benefit the neediest students. Presidential candidate George W. Bush, for one, touts vouchers (or "scholarships" as he prefers to call them) as the great hope for children and schools. In the state of Arizona, voucher initiatives have repeatedly failed to gain legislative support (Garn, 1999), so the Republican-dominated legislature passed a bill to establish a state- wide education tax credit for private school tuition, in essence a voucher in sheep's clothing. I focus here on an analysis of Arizona's education tax credit as a recent and popular example of market-based reform. School tax credits are seen as different from vouchers, and thus are becoming state law and policy with little public notice or debate.

As the political climate surrounding education policy becomes increasingly tolerant of market-based educational changes like voucher and education tax credit plans, a fundamental question is largely missing from the debate: What conception of justice undergirds these plans? And, intricately connected to this, what implications do these market notions have for social justice for poor students and students of color? I will attempt to answer these crucial questions. I shall argue that even though proponents of education tax credits and other such school choice plans claim to be most concerned with improving education for disadvantaged students, in fact, poor students and students of color will ultimately be further disadvantaged by such schemes. In so doing, I will examine Arizona's education tax credit law by placing it within the larger debate over voucher plans, relying often on the Milwaukee, Wisconsin, voucher program as an example. I shall then use this examination of the issues surrounding education tax credits as a backdrop in assessing what I take to be opposing concepts of justice assumed by the proponents and opponents of education tax credits.

The Arizona Education Tax Credit

For four years running, the Arizona state legislature voted against a school voucher plan. After this string of defeats, in 1997, Republican legislators formulated an alternative school choice bill, this time in the guise of education tax credits (Laitsch, 1998). The bill was passed into law, (Note 1) allowing Arizona state income tax payers to claim two different types of dollar-for-dollar tax credit. First is a private school tuition tax credit of up to $500 that can be donated to a School Tuition Organization (STO), which then awards tuition scholarships to students who wish to attend private or religiously affiliated schools. Such tax credits closely resemble privately funded voucher programs, within which private organizations establish so-called scholarship funds for students in private schools (Witte, 2000). These organizations are usually religiously affiliated and have varying selection criteria. For example, in Indianapolis, low-income students could receive up to $800 toward private school tuition, which would cover roughly half of their tuition. Of course, families would have to cover the rest (Witte, 2000). The second Arizona tax credit is a $200 credit that can be donated to public schools for use only on extra-curricular activity fees (e.g., band uniforms or sports equipment). (Note 2)

Soon after the tax credit law was enacted, the Arizona Education Association and others brought suit challenging its constitutionality. In *Kotterman v. Killian*, the Arizona Supreme Court upheld the constitutionality of the education tax credit law on January 26, 1999. The United States Supreme Court refused to hear the appeal. Thus the high court sent the message that the tax credit law could stand as constitutional in the state of Arizona. However, by not hearing the case, they declined to issue a ruling with direct national consequences on the question of education tax credits. By extension, the high court declined as well to rule on the question of vouchers, as an education tax credit is a form of voucher, though not widely recognized as such. In fact, the U.S. Supreme Court also declined to hear a case specifically regarding Milwaukee's voucher program (Pardini, 1999). The Justices seem to be reserving judgment for the time being.

Conflicting Ideas of Vouchers

Although the debate over education tax credits is just heating up, the debate over vouchers has been boiling over since the voucher idea was first introduced by Milton Friedman in the 1950s (Friedman, 1955). There are two major strands in the debate over vouchers, one concerning freedom and the other concerning equity. Of course, the issues within each are complex and overlap significantly. The first strand centers on the broad issue of freedom of choice. Supporters of vouchers argue that parents need more freedom of choice when it comes to their children's schooling, and they believe that vouchers will provide the avenue for those choices. Parents will not be forced to support failing public schools with their tax dollars, especially if they would prefer to send their children to private schools. Once parents threaten flight to better- achieving private schools, the argument goes, the public schools will be forced to improve in order to be able to compete. Voucher foes disagree, arguing instead real freedom of choice will exist only for higher-income parents. They believe that tax money is better invested in public schools that are available to all children. The issue is less about free choice than it is about self-interest. After all, parents are at least nominally free to enroll their children in private schools whether or not vouchers or tax credits exist. Some wealthy families that choose private schools simply do not want their tax dollars supporting public schools if their children are not attending them. Some voucher proponents calling for greater school choice contend that private school vouchers will allow families greater control over their children's schooling, particularly where it concerns religion and morality. With education tax credits and many voucher plans, students can attend religiously affiliated schools, thus enabling them to learn according to their parents' values (e.g., creationism instead of evolution). Opponents declare this an unconstitutional breach of the separation of church and state. Public monies, they maintain, must not go to support private, religious education.

The second strand of argument over vouchers and education tax credits focuses on the issue of who really benefits from these initiatives. Defenders of vouchers often try to take the high ground by arguing that voucher and tax credit plans primarily benefit the least advantaged students and families. With a voucher or tax credit, poor families, they say, will no longer be held captive in bad public schools. They will be able to send their children to private schools for a better education. According to voucher opponents, the least advantaged students are not only not the primary beneficiaries of voucher plans, they are the ones who are most likely to be harmed. Most voucher plans, and the Arizona education tax credit, do not restrict private school tuition aid to needy students. Therefore, there exists a risk that higher-income families will take most advantage of the voucher opportunities, leaving the neediest students in underfunded public schools. The vouchers and tax credits thus function as subsidies for middle- and upper-income families.

In what follows, I will delve more deeply into these two strands of argument, paying close attention to Arizona's education tax credit and Milwaukee's voucher plan.

Strand One: Vouchers and Tax Credits as Freedom of Choice

According to John Chubb and Terry Moe (1990), a free market in schooling will respond to and rectify what they see as public school failure (at least as defined by academic achievement as measured by standardized test scores), stemming from a system of direct democratic control. Democratic control by means of elected school boards, they contend, is responsible for the creation of an unwieldy bureaucracy of school governance that erodes student achievement, parent satisfaction, and educational innovation. Under market-based reform, education would instead be treated as a consumer good. Individual schools would perform well or risk students and parents taking their "business" elsewhere. Families would have greater freedom to have their children attend better private schools as well as parochial schools more in line with their moral values. In addition, these new public schools, unfettered from democratic structures, would rise to the higher levels of academic achievement claimed for private schools (Chubb & Moe, 1990).

Chubb and Moe's (1990) arguments exemplify those of school choice proponents in general. The concern for greater freedom of choice can be separated

into three main issues. These are: 1) unfair taxation, i.e., it is unfair to force parents to give their tax money to public schools when they would rather send their children to private or parochial schools; 2) private school performance, i.e., private schools promise higher academic achievement; and 3) moral values, i.e., religious schools better support conservative moral values.

Unfair Taxation

In *Kotterman v. Killian*, the Arizona Supreme Court looked favorably on education tax credits for private and religious schools because such schools were said to serve the public interest by relieving parents' tax burdens. Critics of vouchers and education tax credits fear the loss of tax dollars for public schools. This fear appears to be warranted. A study of the Milwaukee voucher program has shown that the Milwaukee Public Schools lost over $22 million dollars that they would have received from the state were there no charter schools or vouchers redeemed at private schools (Miner, 1998/1999). Tax credit advocates, however, do not see tax credit money as belonging to the state, since, because the money is donated directly to STOs or schools by the taxpayers, it is never actually in the state's possession. This was the reasoning used by the majority of the *Kotterman v. Killian* justices, who overlooked two things. First, if state income tax money is withheld from a person's paycheck and that person makes a $500 donation to an STO, then when she or he receives the tax credit, the state is in effect returning the $500 – money in the state's possession – to the taxpayer. Second, tax credit donations to STOs or public schools are not simple philanthropic donations of individuals' own money; by taking advantage of the tax credit, individuals are choosing where to place their tax dollars. Without such a choice, that money would go to the state.

In the same vein, if voucher plans do not reduce public school funds, then why would Governors Jeb Bush and George W. Bush both use the threat of federal vouchers to punish schools whose test scores do not improve? George W. Bush has proposed taking money from federal Title I programs targeted on the neediest students and schools to finance vouchers for students in low-performing schools. His plan stipulates that after three years without test score improvement, Title I funding would be taken away from the school and given to the state to set up voucher programs for students (Herman, 1999). It is doubtful, however, that such a plan would withstand the inevitable court challenge.

The nature of taxation is connected with citizens contributing to the public good. If public tax dollars are used to subsidize some students' private school attendance, then tax monies are contributing to some citizens' private good. Foes of education tax credits have legitimate worries about increased inequality and segregation brought on by such school choice programs (see Cobb & Glass, 1999). In Dan Goldhaber's (1999) optimistic view, school choice efforts could break the ties between low-income neighborhoods and poor schools, which could result in fewer white and higher-income families fleeing to suburban school districts. Of course, changing school funding schemes could have this effect as well and would focus on improving public schools rather than escaping them.

Consider the state of Florida's recent school voucher law, which holds that public school students in schools that received failing grades on their state school report cards for two consecutive years could receive vouchers that use tax dollars to fund private school tuition. In March, 2000, it was declared unconstitutional by a state judge. The judge ruled that the voucher law violated the state constitutional mandate to provide students with a free education in public schools (Hallifax, 2000). (Note 3)

Still, the voucher proposals keep coming. New Mexico's Republican Governor Gary Johnson has proposed the most comprehensive U.S. voucher program yet (Janofsky, 2000). Underlying Johnson's proposal is the assumption that private schools are better than public ones; many politicians play into this assumption by proposing voucher and tax credit plans to help families escape public schools. Perhaps it helps avoid the more important discussion about what ought to be done to improve those public schools that are not serving children well.

Private School Performance

Those in favor of tax credits and vouchers contend that private school students by and large have better academic achievement than public school students. As such, more students should be provided with the opportunity to attend private schools, and school choice plans should help them afford the costs. This argument is compelling, because as it stands, only middle- income and high-income families can afford private school education for their children. Tuition tax credits and vouchers, then, can make the difference for some lower- income families. The problem is that this logic assumes that private schools do indeed provide students with a better education than do public schools. Goldhaber (1999) points out that on average, private schools tend to produce higher standardized test scores, and high school graduation and college attendance rates. However, and this is the key point, Goldhaber also mentions that this finding does not take account of the differences between private school selection criteria and public school open admissions (Goldhaber, 1999). In fact, the findings on student achievement do not support the contention that private schools do a better job educating students (Goldhaber, 1999; Witte, 2000). Consider that when John Witte (2000) compared the reading and math achievement of Milwaukee Choice students and Milwaukee public school students, he found no statistically significant differences. This finding led him to conclude that "the battle and politics over vouchers may have more to do with money and with the allocation of power than with education" (Witte, 2000, p. 157).

In addition to perceptions of higher achievement, private schools are also perceived as having more involved and active parents. Similarly, it is often the most involved parents who take advantage of school choice opportunities (Witte, 2000). As such, these involved parents leave the public schools rather than using their energies to help improve them. According to Witte

> One could...reasonably argue...that if these students and families remained in their prior [public] schools, they could exercise considerable influence in attempting to improve those schools. [Choice] Parents were educated, angry, involved, and had high expectations for their children. If engaged and given the opportunity, they could push the public system rather than leave it (Witte, 2000, p. 73).

It defies common sense that in order to foster improvement in the public schools, we should shift more good students, active parents, and financial resources to private schools. It seems that this would render public schools less likely to improve. In a review of the empirical literature regarding school choice, Goldhaber found that competition does seem likely to spur public schools to change (Goldhaber, 1999). But why promote public school reform in ways that risk making them worse? Why act as if schools and students in poor neighborhoods deserved to be punished for having to deal with myriad issues that high-income schools do not face? Why not instead foster action by involved parents within the public system?

Moral Values

Overall, approximately 85% of all private school students attend religious schools (Witte, 2000). Although it is reasonable for families to choose to send their children to schools that uphold their religious tradition and moral values, it is not right for them to do so using public tax dollars. To do so threatens the separation of church and state. In upholding the constitutionality of Arizona's tuition tax credit, the majority opinion in *Kotterman v. Killian* stated that before the tax credit initiative, low-income parents

> may have been coerced into accepting public education. These citizens have had few choices and little control over the nature and quality of their children's schooling because they have been unable to afford a private education that may be more compatible with their own values and beliefs. Arizona's tax credit achieves a highe degree of parity by making private schools more accessible and providing alternatives to public education. (Note 4)

Three points can be raised regarding this portion of the *Kotterman* court's opinion. First, the Justices appear to place considerable value on an education that is in harmony with families' moral values and beliefs. They are presumably referring to the generally conservative moral values espoused in religious schools. If tax credits are to be valued because they help achieve greater economic parity, then why should moral values play a part in their defense? Second, the Justices assume that the inability to afford private school has *caused* low-income parents to have "few choices and little control" over their children's education. Poverty and isolation are more likely to have caused these families' educational difficulties, and simply placing children in private schools is not likely to solve the problems associated with poverty. Third, the Justices ignore the fact that the tax credit also makes private school more affordable for students already enrolled as well as other middle- and high-income students. Subsidizing the attendance of economically privileged students at private schools will not help Arizona achieve the higher degree of parity the Justices seek.

Thus far, it has been the Catholic Tuition Organization of Phoenix and the Arizona Christian School Tuition Organization, both of which support schools affiliated with their respective religions, that have benefited disproportionately from the Arizona tuition tax credit (Schnaiberg, 1999; Wilson, 2000). As of January, 2000, these two STOs had received over $1,375,000 of the total $1,800,000 donated to 15 of Arizona's STOs (Center for Market-Based Education and the Goldwater Institute, 2000). Apart from reporting to the Arizona Department of Revenue on the amount of their scholarship money that was allocated, there is no accountability for how the STO scholarships are disbursed. While donors cannot designate a donation to benefit their own children directly, there is nothing stopping them from, say, earmarking their donation for a friend's child. The STO has complete freedom to determine how money is allocated among applicants and the amount of aid each will receive (Center for Market-Based Education and the Goldwater Institute, 2000). In addition, as Justice Feldman points out in the dissenting opinion in *Kotterman v. Killian*, "contrary to the majority's assertion, the [tax credit] statute promotes support of religious schools. It does this without prohibiting use for sectarian instruction, thereby allowing direct state subsidy of religious instruction and observance." (Note 5) In essence, Arizona tax dollars are funding the teaching of specific religious and moral values.

We arrive at what Kenneth Howe (1997) identified as a case of the slippery slope. Once voucher and tax credit plans are introduced, even the most restrictive ones, they tend to slip toward subsidy of religious schools and toward benefiting primarily higher-income students as well. The Milwaukee voucher experiment provides an example of how the slippery slope can turn a successful limited voucher program into a program to subsidize religious school attendance. Beginning in 1991, only low-income students from Milwaukee public schools were eligible to receive approximately $5,100 to attend nonsectarian private schools. By 1995, the program had expanded to include religious schools, and by 1999, 69% of the participating choice schools were affiliated with a religion (Witte, 2000). Once the Milwaukee Choice program expanded, it changed in significant ways. By the 1998-1999 school year, only 23% of "choice students" came from the Milwaukee Public Schools; the rest had already been enrolled in private schools; 55% of these were already private school students or new private school students, and 22% were continuing choice students. Perhaps the best evidence of the slippery slope phenomenon is that the mayor of Milwaukee has proposed removing the program's income restriction, which will likely result in private school subsidies for higher-income families (Witte, 2000).

Or consider the Cleveland, Ohio program. It began with many fewer restrictions than Milwaukee's, and as a result, saw discouraging results. The voucher schools were mostly parochial, the vouchers were not specifically targeted toward needy students, and it subsequently became a program primarily benefiting students who were already attending private schools (Witte, 2000). Despite the evidence that voucher and tax credit plans will inevitably encounter the slippery slope phenomenon where they end up benefiting mostly high-income students and private and religious schools, school choice supporters continue to insist that these programs

are designed to help the neediest students and schools.

It would be difficult to dispute the notion that many public schools are in need of improvement; studies by Jonathan Kozol (1992) and Jean Anyon (1997) provide compelling stories of the plight of urban public schools. In that regard, voucher proponents join with liberals in decrying the sorry state of some public schools. However, school choice advocates ignore evidence that the democratically controlled U.S. public school system is doing a remarkable job. More people than ever are guaranteed a free education, and the U.S. high school graduation rate is at its highest point in history (Berliner & Biddle, 1995). Market-driven reform supporters stretch facts when they paint the public school system as an unredeemable villain. Yes, it is crucial for our neediest students that something be done to improve the school infrastructure, teacher quality, and achievement in the poorest schools. But market-based ideas such as vouchers and tax credits hold little, if any hope for beginning that long and complicated process. Vouchers and tax credits are far more likely too do further harm to poor students and students of color.

Strand Two: Vouchers and Tax Credits as Benefit for the Least Advantaged

Proponents of school vouchers and education tax credits argue that these programs serve primarily to combat the inequality faced by low-income students. Opponents claim the opposite, viz., that vouchers and tax credits will serve to perpetuate the already unjust funding disparities between public schools in high-income and low-income areas and consequently, hinder progress toward social justice goals.

Andrew Coulson (1996) looked historically at market-inspired reforms and maintains that such fears of injustice are unfounded. He complains that opponents assume that too many families would not actively find out about the options for choice. In making the case for market-driven change, he writes that "Members of the minority groups assumed to be incompetent or uninterested in their children's education are foremost in defending their ability and prerogative to choose" (Coulson, 1996, p. 3). This is a misleading characterization of an important objection to market-based choice schemes. In implying that choice opponents discount the intelligence and power of low-income parents and parents of color, Coulson dismisses the very real danger that a large portion of low-income families are harmed by voucher and tax-credit plans. In initial data analyses of donations in the first year of the Arizona education tax credit, Glen Wilson (2000) found that the education tax credits exacerbate the already disastrous inequities in Arizona public school funding.

With the Arizona education tax credit, only taxpayers can benefit from this so-called expanded choice. Families that do not earn enough to pay taxes – those whose children are arguably the poorest and most in need of expanded options – will not be able to contribute either to their children's public school or to an STO. Thus, the likelihood of their pursuing the benefits of these tax credits seems remote. Why then, as Coulson (1996) notes, do some low-income families seem to support vouchers plans such as these? Voucher advocates often point to the fact that the Milwaukee voucher program was initially strongly supported by urban African American parents (Witte, 2000). The abstract promise of vouchers and tax credit monies to pay for a better school for one's children is indeed hard for many families to resist. The problem comes when the voucher plans are put into action. Wisconsin state Representative Polly Williams, an outspoken early architect of Milwaukee's voucher program, learned this first-hand. Within five years of the outset of the Milwaukee program, Representative Williams and most other African American leaders in Wisconsin had rescinded their support of the program. (Note 6) In Cleveland, no African American leaders support the voucher plan (Witte, 2000). They were disillusioned once the promise that the voucher plan held for urban students in general and students of color in particular failed to materialize. Worse, the voucher program was expanding and was benefiting principally higher- income families (Witte, 2000). Representative Williams pointed out that "'This is what you call hijacking the program. There are people in the coalition who never intended to help low-income children.'" (Quoted in Witte, 2000, p. 170). According to Timothy McDonald, Chair of the national African American Ministers Leadership Council,

"'Inner city parents whose schools are not performing well are desperate for solutions and the Religious Right is exploiting that frustration. This is really an attempt to divide the African American community against itself'" (Rethinking Schools, 1999, p. 2). Voucher and tax credit plans often use the term "scholarship" rather than voucher to skew perceptions of the program. For example, the Cleveland program is called the Ohio Pilot Project Scholarship Program, and Florida's vouchers are termed opportunity scholarships by Governor Jeb Bush. Why not just use the language of vouchers? If there is nothing to be ashamed of or opposed to regarding voucher plans, then why are such circumlocutions employed?

Vouchers and education tax credits serve to perpetuate *status quo* class arrangements within schools. As Howe (1997) observed,

> we've now taken our first step on a slippery slope and will inevitably slide to the bottom, where privatized schooling will be there to greet us. The only remedy for this problem is to demand significant restrictions, to keep equality of educational opportunity at the forefront, and to insist on continued efforts to improve existing public schools (Howe, 1997, p. 123).

Perhaps an even better remedy would be not to allow voucher and tax credit plans in the first place. Perhaps then we could go about the work of improving schooling for poor children without the distraction and detraction of market- based choice schemes.

One is drawn to Witte's conclusion: "the most plausible explanation for the continuity and expansion of vouchers has little to do with aiding poor, minority students, and much more to do with distributing subsidies to those who now attend private schools, or would do so in the future" (Witte, 2000, p. 158). Data on the first year of the Arizona tuition tax credit show that poor students and students of color indeed are not receiving the majority of the financial benefit (Wilson, 2000). In Presidential Candidate George W. Bush's plan for education, parents could place up to $5,000 per year in tax-free education savings accounts to be used for K-12 private school expenses. The current limit on such accounts is $500 per year for college fees (Johnson, 2000). In the spirit of education tax credits, this type of tax-free saving serves to benefit wealthier families. If a family puts $5,000 into a tax-free savings account to be used for private school tuition, and they are in, say, a 25% tax bracket, then they are gaining $1,250 tax dollars. The corollary effect of this, of course, is that the tax dollars available to fund public schools are significantly reduced. In addition, poor families would have a much harder time contributing to such a savings account, and so could not take advantage of either the tax break or the additional funds to put toward private school tuition, if they so desired. It is difficult to believe that school choice proponents such as Bush still attempt to have the public believe that they are really just trying to help our neediest students. Perhaps we should not be surprised, given the presence of similar attempts in our nation's history. It was also argued that the doctrine of "separate, but equal" would be in the best interest of people of color. Even the U.S. Supreme Court supported that position in *Plessy v. Ferguson* (1896).

In addition to the harm that vouchers and tax credits do to low-income students is the harm that they may do to students with disabilities. By virtue of the fact that private schools need not make any adjustments for students with disabilities, private schools are much less likely to serve this group of students than are public schools, who are required by law to do so. Consider that under Milwaukee's Choice plan, private schools do not have to accept students with disabilities; these students can, in essence, be officially excluded (Witte, 2000). The issue of students with disabilities provides further evidence that voucher and tax credit plans are not truly concerned with aiding the least advantaged students.

Both Witte (2000) and Goldhaber (1999) found that families with more formal education are more likely to take advantage of opportunities for school choice. Because he looked at all types of choice program and not just a restricted one like Milwaukee's, Goldhaber also found that higher-income parents are more likely to use charter schools, vouchers, and the like when they were available. These outcomes could be predicted, especially when we look at data from the Milwaukee

voucher program showing that information about the program reached families most often through personal contacts and word-of-mouth (Witte, 2000). If the same is found to be true in Arizona, then it is much less likely that poor students and schools in poor neighborhoods will learn about and benefit from the tax credit in large numbers. It is no wonder then that Goldhaber concludes that "unfettered school choice would likely lead to increased racial and economic segregation (Goldhaber, 1999, p. 23).

For a public school donation to fall under Arizona's tax credit rules, it has to be earmarked for activities for which there are *extracurricular* fees, say, for example, a Spanish class trip to Spain. The public school tax credit thus raises two major problems. First, the extra-curricular activities for which these donations can be used are rather unlikely to significantly improve students' academic experience. While there are certainly benefits to be had from having better band uniforms, or traveling to Spain, perhaps money could be better spent on teacher salaries, for instance, so that lower-income schools can at least try to compete with higher-income schools for the best prepared teachers. Second, there is no guarantee that anyone will volunteer to donate anything to the neediest public schools. However, we can imagine that school administrators would try to take advantage of the tax credits and mobilize parents and members of their community to make their donations. But, again, which schools would be more likely to have the time and energy to devote to such a campaign? Higher-income schools, to be sure. I am *not* saying that educators and families in low-income schools do not care or do not value education, as has been suggested in other research (Coulson, 1996; Goldhaber, 1999). (Note 7) What I *am* saying is that if parents are struggling to provide food and shelter for their children, and if educators are struggling to keep kids safe and in school, it seems unlikely that these parents and educators will manage also to wage a community campaign for tax credit donations. Who cares about new band uniforms if the school has no band? In addition, the tax credit money that wealthy families contribute to public school extra- curricular activities does not go into the state's general tax fund, leaving the state less money to fund social programs that could help poor students and their families.

Opposing Conceptions of Justice

How can it be that voucher and education tax credit programs are held up simultaneously as 1) needed help for both the disadvantaged and the unsatisfied, and 2) yet another way that the wealthy and powerful can help themselves to the detriment of our neediest students? The problem is that the opposing camps are, in many ways, entering into the debate holding vastly different assumptions about freedom and, most importantly, about justice. Whether one supports vouchers and education tax credits depends largely on which set of philosophical assumptions one holds. One set of assumptions —those that voucher proponents subscribe to —stems from a libertarian conception of justice. The other set—those that voucher opponents subscribe to—stems from a contemporary liberal democratic conception of justice. In the next section, I explore these opposing ideas, settling on the liberal democratic conception as most genuinely concerned with important considerations of justice for the least advantaged students.

Libertarians Versus Contemporary Liberal Democrats

Libertarian theories of justice are primarily concerned with issues of freedom and individual choice, specifically as exemplified in the free market. As long as the distribution of goods stems from free exchanges, any inequalities in the distribution that result are just (Howe, 1997). Therefore, justice is served when society functions as freely as possible, there is little state involvement in the affairs of individuals, and persons are free to choose the good. Hence, school choice plans like education tax credits fit perfectly with these notions of justice. By allowing citizens to determine for themselves where (at least part) of their tax money will go, and allowing them to have greater choice as to where to educate their children, according to their own belief system, tuition tax credits serve libertarian notions of justice well. It simply does not matter to libertarians that this is unfair for families and schools in

low-income neighborhoods. The education tax credit is therefore seen by its proponents as one more device for empowering parental choice. Defenders of school choice view it as being fair and just, arguing that the system of financing public education is coercive and discriminatory. According to libertarian theory, the tax credit is available equally to all; low-income parents have the same chance as anyone else to use it. If they do not take advantage of the opportunity, it is their own choice. The entire notion of market-based educational change thus assumes a libertarian conception of justice.

While there is some merit to the virtues of freedom and of choice, libertarians run into significant problems when contemporary liberals (myself included) question them on the aims of education in a democracy and on the actual outcomes of education tax credits for the neediest students. First, tax credits, which add up to public financial support for private and parochial schools, are detrimental to the major aim of education to prepare all students for democratic participation. Second, while Arizona tax credit supporters claim it will indeed result in expanded choices for low- income students, the actual outcomes of the first year of the tax credit program show that this is not at all the case (Wilson, 2000). Advocates of education tax credits are operating under an impoverished notion of social justice for students, especially low-income students. They are asking the wrong questions, and trying to fix the wrong problems. Instead of combating the woes of low-income public schools head-on, they are attempting to shift the emphasis from reform and social change to an emphasis on individual freedom (Cookson, 1992). But this notion of freedom is empty, as it is disconnected from both the sociopolitical context and a concern for others, resulting in a type of politics of disconnected freedom. Within these politics, parents and students are merely self-interested consumers who would use the libertarian free market rationale to justify fleeing public schools, or, at least, finance summer cheerleading camp. Yet public schools are the very institutions that serve to sustain a notion of education that aims to prepare all students for meaningful and critical participation in our democracy. Contemporary liberal theories would advocate instead a sense of reciprocity, within which persons do not act only in their own self-interest, but instead aim to act fairly and justly by trying to understand intimately the perspectives and standpoints of others (Gutmann & Thompson, 1996; Rawls, 1971). According to John Rawls, the leading liberal democratic theorist of justice, the only way that any inequalities, such as the ones exacerbated by education tax credits, can be tolerated is if they serve to make all students somehow better off (Rawls, 1971). Education tax credits simply do not fit into this stipulation.

Nonetheless, it seems that, *prima facie*, school choice proposals like education tax credits are in keeping with the tradition of democratic education (Gutmann, 1987). Parents have more freedom of choice, public schools feel compelled to improve in order to compete with private and parochial schools, and poor families have an increased opportunity to place their children in the best schools. However, as Howe (1997) points out, school choice schemes are actually incompatible with equal educational opportunity and democracy. Libertarian advocates for education tax credits tout this reform without taking into account the social and political context underlying the inequalities they contend tax credits will help minimize. In actuality, market-based educational reforms allow us to avoid rather than deal with debates over the nature of schooling in a democracy (Gutmann, 1987). David Berliner and Bruce Biddle (1995) remind us that while the free market generally guarantees efficiency, it does not guarantee equality (Berliner & Biddle, 1995). This is certainly true, and libertarians are aware of this point. Nevertheless, the guiding philosophy behind market-based schemes does not hold the concepts of equality and justice paramount. As long as everyone has nominal freedom of choice, justice will be done.

Libertarian theories, then, allow market-driven reform proponents to idealize the notion of freedom, conveniently forgetting that, as contemporary liberals argue, the state cannot be neutral about civil rights (Moses, 2000). That is to say, left to their own devices, there is little guarantee that private and parochial schools will pay any attention whatsoever to issues of equality and social justice. What is more, history shows that private and parochial schools will not pay any such attention, and may in some instances be designed so that they may exclude certain students. Chubb and Moe (1990) claim that it is good that private schools are separated from

democratic political processes. Contemporary liberals disagree, responding that that is precisely the problem if we truly care about social justice. Perhaps libertarians do not. They should.

Concluding Thoughts

Consider the following two student cases.

1. Jonathan is a young boy who is having behavioral difficulties in his public elementary school in the inner city of Newark, New Jersey. In recent weeks, Jonathan has switched from living with his father to living with his mother. When he passes his father on the street, his father refuses to even say hello to him. Since Jonathan began living with his mother, his father ignores him. (Note 8)
2. Harold is an elementary schooler in Los Angeles whose mother works an alternating shift in the local hospital's laundry. She is the family's sole source of financial support. His father left when he was five and is now in jail. Harold has for years been labeled a distracted problem child. His mother usually does not have time to attend parent-teacher conferences to discuss Harold's situation. (Note 9)

In the first case, Jonathan is just one example of the children in his elementary school who Jean Anyon (1997) describes as having "hard lives" (Anyon, 1997, p. xiii). A teacher in the school describes the myriad troubles faced by the children she teaches: "Derrick's father died of AIDS last week;...One girl's father stole her money for drugs...One boy had a puffy eye because his mother got drunk after she got laid off and beat up the kids while they were sleeping" (Anyon, 1997, p. xiii). In the second case, Harold's mother is having great difficulty being her family's sole provider. A medical problem is plaguing her, she has little time to spend with Harold, and she seems resigned to a life of poverty.

Think again of the debates over education tax credits. Is it likely that Harold's mother or many of the parents of students in Jonathan's school will be able to donate their $200 tax credit to the school, or $500 to a School Tuition Organization? Will they then help their child apply first for a private school tuition scholarship, and then for admission to a private school, and last obtain transportation to and from a private school that is not in their neighborhood? Would a private school even choose to admit students like Jonathan and Harold? If the parents were able to donate their $200 tax credit to the public school, for what would the school use the money? There is no school band or science laboratory or athletic team in need of new equipment. The things the students need most are not fee-driven extra-curricular activities; they need functional bathrooms, bilingual teacher aides, and up-to- date classroom computers (Kozol, 1991). Of course, in no way do I mean to imply that low-income students do not deserve to participate in extra-curricular activities, only that the things that Arizona's $200 tax credit will pay for are not the highest priority for school improvement. Perhaps in some schools in higher-income neighborhoods that is not the case. However, tax credit proponents claim that they will help revitalize poor public schools that serve needy students. I do not see how, for it does not even seem that the tax credits were formulated with such students in mind.

It would be best, and most just, to focus school reform efforts on improving public schooling, especially for the neediest students, and celebrating what is good about public schools, rather than demonizing public education in an effort to serve special interests, as proponents of school choice tend to do. In the interests of justice, it is liberal democratic rather than libertarian market principles that should guide public schooling. It is public, not private, education that is a primary good in the United States, for at its best, it serves the critical social purpose of educating all students for meaningful democratic participation (Gutmann, 1987). In the end, education tax credits and other school choice schemes will not help to reform public schools that do not successfully serve that social purpose or diminish social inequalities. Rather, they will result in injustice by exacerbating the very inequalities that they claim to erase.

Education tax credit schemes cannot change the fact that school funding is

largely based on property taxes. Family incomes and therefore neighborhood and housing situations determine children's school possibilities and too often a school's quality as well. This is a serious fundamental problem. Schools in poor areas ought to receive more state dollars than their wealthy counterparts. Perhaps restricted school choice plans (like the original Milwaukee voucher program) may help some students. That is perhaps acceptable. But realistically, the case cannot be made that market-based reforms will save children from poverty or even save them from unfair school funding arrangements (Traub, 2000).

As Berliner and Biddle (1995) document, overall public schools in the U.S. are doing a good job (Berliner & Biddle, 1995). Similarly, in his journey across the U.S. to visit public schools in places such as New York City; Chicago, Illinois; Hattiesburg, Mississippi; and Tucson, Arizona, Mike Rose learned that public school "classrooms offer a collective public space in which America is being created" (Rose, 1996, p. 10). This, despite rampant poverty and racism. Imagine the possibilities if we were to focus our collective societal energy and power on eradicating poverty rather than on things like self- interested tax credit schemes.

Notes

1. A.R.S. § 43-1089 (1997).
2. Apparently, these donations can also be deducted from one's federal taxes as well - in essence providing a double benefit. See paragraph 148 of the dissent of Kotterman v. Killian, 972 P.2d 606 (1999).
3. As of this writing, the state was in the process of appealing the decision to the 1st District Court of Appeals (Hallifax, 2000).
4. See paragraph 22 of the majority opinion of Kotterman v. Killian.
5. See paragraph 93 of the dissent of Kotterman v. Killian.
6. In fact, only one African American leader, Dr. Howard Fuller, continues to support Milwaukee's voucher plan. (For the complete story, see Chapters 7 and 8 of Witte, 2000).
7. Goldhaber (1999) suggests that better-targeted publicity can help low-income families find out about school choice opportunities because they "may have less knowledge about the workings of the educational system and the value of education" (Goldhaber, 1999, p. 23). Better-targeted publicity about options cannot hurt, but there is no need to perpetuate the idea that low-income people do not know the value of education.
8. This story comes from a case described in Jean Anyon's book *Ghetto Schooling* (Anyon, 1997).
9. Harold's story is recounted in Mike Rose's book *Lives on the Boundary* (Rose, 1990).

References

Anyon, J. (1997). *Ghetto schooling: A political economy of urban educational reform.* New York: Teachers College Press.

Berliner, D. C., & Biddle, B. J. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools.* White Plains, NY: Longman.

Center for Market-Based Education and the Goldwater Institute. (2000, January 26). EXTRA-CREDIT newsletter. [Online.] Available at: http://www.cmbe.org/extra_info.htm.

Chubb, J. E., & Moe, T. M. (1990). *Politics, markets, and America's schools.* Washington, D.C.: The Brookings Institution.

Cobb, C. D. & Glass, G. V. (1999). Ethnic segregation in Arizona's charter schools. *Education Policy Analysis Archives, 7*(1). [Online] Available: http://epaa.asu.edu/epaa/v7n1/ [August 1, 2000]

Cookson, P. W. (Ed.). (1992). The choice controversy. Newbury Park, CA: Corwin

Press.

Coulson, A. (1996). Markets versus monopolies in education: The historical evidence. *Education Policy Analysis Archives, 4*(9). [Online] Available: http://epaa.asu.edu/epaa/v4n9.html [August 1, 2000]

Friedman, M. (1955). The role of government in education. In R. A. Solow (Ed.), *Economics and the public interest* (pp. 123-144). New Brunswick, NJ: Rutgers University Press.

Garn, G. A. (1999). Solving the policy implementation problem: The case of Arizona charter schools. *Education Policy Analysis Archives, 7*(26). [Online] Available: http://epaa.asu.edu/epaa/v7n26.html [August 1, 2000]

Goldhaber, D. D. (1999). School choice: An examination of the empirical evidence on achievement, parental decision making, and equity. *Educational Researcher. 28*(9), 16-25.

Gutmann, A. (1987). *Democratic education.* Princeton, NJ: Princeton University Press.

Gutmann, A., & Thompson, D. (1996). *Democracy and disagreement: Why moral conflict cannot be avoided in politics, and what should be done about it.* Cambridge, MA: The Belknap Press.

Hallifax, J. (2000, March 23). Voucher law thrown out, [Online article]. *The Associated Press/ABCNews.com.* Available: http://more.abcnews.go.com/sections/politics/dailynews/florida000314.html [2000, March 23].

Herman, K. (1999, September 3). Bush pushes plan for failing schools. *Denver Rocky Mountain News.* p. 45A.

Howe, K. R. (1997). *Understanding equal educational opportunity: Social justice, democracy, and schooling.* New York: Teachers College Press.

Janofsky, M. (2000, February 6). N.M. considers most ambitious voucher program. *The Arizona Republic,* p. A23.

Johnson, G. (2000, March 24, 2000). Gore, Bush make visits to schools, [Online]. *Associated Press.* Available: http://news.excite.com/news/ap/000324/11/education-candidates [2000, March 24].

Kozol, J. (1991). *Savage inequalities: Children in America's schools.* New York: Harper.

Laitsch, D. (1998, October 1998). School choice and privatization efforts in the states: A legislative survey, [Online]. AACTE. Available: http://www.edpolicy.org/nat.iss/eval/priv.htm [2000, February 16].

Miner, B. (1998/1999). $22 million siphoned from MPS to pay for private schools. *Rethinking Schools, 13*(2), 3.

Moses, M. S. (2000). The relationship between self-determination, the social context of choice, and authenticity. Paper presented at the Philosophy of Education Society Annual Meeting, Toronto, Canada.

Pardini, P. (1999). Chruch/state complexities. *Rethinking Schools, 14*(2).

Rawls, J. (1971). *A theory of justice.* Cambridge, MA: Harvard University Press.

Rethinking Schools. (1999, Spring). False choices: Vouchers, public schools, and our children's future, pp. 1-4.

Rose, M. (1990). *Lives on the boundary: A moving account of the struggles and achievements of America's educational underclass.* New York: Penguin Books.

Rose, M. (1996). *Possible lives: The promise of public education in America.* New York: Penguin Books.

Schnaiberg, L. (1999, April 14). As tax day nears, Arizona prepares to tally credits. *Education Week,* p. 17.

Traub, J. (2000, January 16). Schools are not the answer. *The New York Times Magazine,* 52-57,68, 81, 90-91.

Wilson, G. Y. (2000). Effects on Funding Equity of the Arizona Tax Credit Law. *Education Policy Analysis Archives, 8*(37). [Online]. Available: http://epaa.asu.edu/epaa/v8n38.html [August 1, 2000]

Witte, J. F. (2000). *The market approach to education: An analysis of America's first voucher program.* Princeton, NJ: Princeton University Press.

About the Author

Michele S. Moses
College of Education
Arizona State University

(480) 965-4673
Email: michele.moses@asu.edu

Michele S. Moses is Assistant Professor in the Division of Educational Leadership & Policy Studies at Arizona State University. She received her Ph.D. from the University of Colorado at Boulder, specializing in the philosophy of education and education policy analysis. Her research interests include issues of educational equity and social justice and the politics of education policies concerned with multiculturalism.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografía e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 92 times since August 1, 2000

Education Policy Analysis Archives

Volume 8 Number 38 August 1, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Effects on Funding Equity of the Arizona Tax Credit Law

Glen Y. Wilson
Arizona State University

Abstract
This article examines the results from the first year (1998) of the Arizona Education Tax Credit program. The tax credit law allows individuals a dollar- for-dollar tax credit of $500 for donations to private schools and a dollar-for-dollar tax credit of $200 for donations to public schools. Although one justification for this statute was that it would help lower income students, the primary beneficiaries of this program tend to be the relatively well off. The author concludes that Arizona's tax credit law increases educational funding inequity in Arizona. Data for 1999, only recently made available, show a 159.1 percent increase in total contributions and an exacerbation of the trends noted here.

This article is one of four on the Arizona Tax Credit Law:

- Welner: Taxing the Establishment Clause
- Moses: Hidden Considerations of Justice
- Rud: Moral Considerations

Introduction

Education tax credits are a relatively new mechanism intended to promote and fund school choice by means of the tax system. In Arizona's first regular legislative session in 1997, House Bill 2074 was passed and on April 7, 1997 was signed into law by Arizona Governor Fife Symington as A.R.S. § 43-1089. Beginning with the 1998 tax year, A.R.S. § 43-1089 created a private school tuition organization individual income tax credit and a public school extracurricular activity fee

individual income tax credit.

With the private school tax credit, Arizona taxpayers were granted a full and direct credit against state income taxes for contributions up to $500 to school tuition organizations (STOs). STOs then provide grants to students to attend private schools. A.R.S. § 43-1089 contains very few restrictions as to how the proceeds from this tax credit are to be used. The major restrictions are: that taxpayers claiming this credit may not earmark their donation to their own dependents, that STOs allocate at least 90 percent of their annual revenue for "educational scholarships" or "tuition grants," and that STOs provide scholarships or grants without limiting availability to only students of one school (A.R.S. § 43-1089).

A similar $200 tax credit is also available for contributions to public schools; however, these contributions may only be used for extracurricular activities that require a student fee. Examples provided in the statute include: band uniforms, equipment or uniforms for varsity athletic activities and scientific laboratory materials (A.R.S. § 43-1089.01). Originally, contributions to public schools did not qualify for this credit because the legislative bill restricted the tax credit to "a nongovernmental primary or secondary school" of the "parents' choice" [A.R.S. § 43-1089 (E) (1), (2)]. As a compromise with opponents of the legislative bill, the law as finally enacted included a $200 tax credit for contributions to K-12 public schools.

To tax professionals, provisions such as tax credits and tax deductions are known as tax expenditures. Tax expenditures are special preferences embedded in the tax code that are intended to benefit particular activities or groups. Tax expenditures cause a loss of tax revenue and thus, are functionally equivalent to government spending programs. Surrey and McDaniel (1985) stated the following about tax expenditures:

> Whatever their form, these departures from the normative tax structure represent government spending for favored activities or groups, effected through the tax system rather than through direct grants, loans, or other forms of governmental assistance....These tax expenditures in effect represent monetary assistance provided by the government (p. 3).

It should be noted that unlike tax deductions allowed for general charitable giving, Arizona's education tax credit provides a full reimbursement to those who contribute. Thus, the tax credit plan does not function as a stimulus to charitable giving, but instead functions to allow self-selected taxpayers to redirect funds, that would otherwise flow into state accounts, to private entities of their own choosing.

A major justification for school choice programs has been to offer additional educational alternatives to low-income families. The Arizona tax credit law was promoted with a similar justification. *The Arizona Republic*, in a recent story on the tax credit program reported that "Supporters of the credit for private school scholarships, including Rep. Mark Anderson, R-Mesa, who sponsored the legislation, touted it as a way to send kids to private school who otherwise couldn't afford to go" (Bland, 2000). Arizona Supreme Court Chief Justice Thomas B. Zlaket offered similar reasoning in the opinion upholding the school tax credit law. Zlaket wrote: "Until now, low-income parents may have been coerced into accepting public education...Arizona's tax credit achieves a higher degree of parity by making private schools more accessible and providing alternatives to public education" [Kotterman v. Killian, No. CV-97-0412-SA (1999)]. If such published accounts were accurate, it would appear that the primary intended beneficiaries of the law could be construed as low-income students and their families with a primary intended effect of increased educational choice (increased access to private schooling). For public schools, the justification appears to be to assist parents in paying for public school extracurricular activities. To extend the justification for the private school tax credit to the extracurricular public school tax credit would logically mean that the primary beneficiaries of the public school tax credit should be students and families that face hardship in paying extracurricular fees.

However, to opponents, education tax credits are poor public policy and a dangerous road on which to travel. In addition to fundamental constitutional questions of separation of church and state, many critics believe that tax

expenditures, such as tax credits, tend to be highly inequitable. Wealthy individuals may be much more likely to take advantage of them than lower-income individuals, who may not even earn enough income to participate in the program. For example, Weinberg (1987) calculated that for FY 1985, at least 50 percent of the total benefits provided by tax expenditures through the U.S. individual income taxation system went to the top 20 percent of families (in terms of income). The poorest 40 percent of families (by income) received less than 20 percent of the total benefits offered through tax expenditures. Under Arizona's plan, those participating receive a full reimbursement of their contribution and thus, do not actually incur any costs at all. Therefore, Arizona's plan appears to allow higher-income individuals to direct a portion of state tax revenue to public or private schools while possibly denying lower-income individuals an equal real opportunity to do the same. Another objection to the use of tax credits relates to the distributional pattern that critics believe will occur. Critics have charged that under this plan, resources will not flow to where needs are the greatest—that in the end, this plan will be just another subsidy for the middle-class.

Research Design

The purpose of the quantitative analysis reported here is to describe the distribution of tax credit contributions in terms of student poverty/wealth, contributor poverty/wealth, enrollment and student achievement. Since the data in hand constitute a full census of the education tax credit records for the 1998 tax year, no questions of statistical inference arise. Rather, the purpose of the data analysis will be to show the different levels of contributions in terms of different factors.

Data Collection and Preparation

Complete records of all Calendar Year 1998 contributions (as of March 26, 1999) under the education tax credit law were obtained from the Arizona Department of Revenue (ADOR). Approximately 60,000 contributions were documented, accounting for about $7.7 million dollars. The number of contributions and the total amount contributed to the recipient school were provided; no taxpayer identification (neither personal identity, location nor income level) was included. The data contained listings for 1,144 K-12 public schools. Data on public schools participating in the federal free/reduced meal program (F/R meal) were obtained from the Arizona Department of Education (ADE). The number of students eligible for the F/R meal program as well as the total school enrollment were contained in the data from ADE. After combining the two data records, there were 929 public schools (81.2% of the total) for which there was data on both measures (tax credit contributions and F/R meal program). Schools for which there was no tax credit contribution listing and/or no free/reduced meal program data were not included in the analysis. For the public schools with data on the two elements of interest, information as to the school's 1997-98 student performance on the state-mandated Stanford-9 Achievement Test was added for each school. For elementary schools, the 4th grade reading and math individual percentile ranks were used; for middle/junior high schools, 7th grade reading and math individual percentile ranks were employed; and for high schools, 9th grade reading and math individual percentile ranks were used. If the particular score for a school was missing, the closest available score was used. For example, if the 4th grade reading or math score was missing for an elementary school, then the closest available score such as the 3rd grade score for that particular school was used. The reading and math individual percentile ranks were summed and divided by 2 to provide a combined score for each school. The 929 public schools in the dataset were placed into quarters based on the percentage of a school's students eligible for F/R meal program. In this dataset, these percentages ranged from 1 to 100 percent of schools' enrollment.

The data on tax credit claimants (Tables 4 – 7) are based on ADOR's review of individual tax returns. As of September 23, 1999, approximately 25,000 individual

tax returns have been reviewed. ADOR estimates that nearly 17,000 tax returns filed prior to September 1, 1999 have yet to be reviewed. Any tax returns filed after September 1, 1999 and before the end of calendar year 1999 will also require review in order to have complete first year results. The data concerning private schools and School Tuition Organizations (table 8) were obtained from ADOR, the Center for Market-based Education, and telephone calls to individual STOs.

Findin[illegible] · Public Schools

After the ADOR tax credit and ADE F/R Meal Program data records were combined, there were 929 public schools enrolling 672,211 students, for which there was data on both measures of interest (contributions under the tax credit program and F/R meal program). Stanford Achievement Test data were then added to the dataset and schools were arranged into quarters on the basis of relative poverty/wealth. Summary tables were developed for several items of interest (school characteristics, school basis contribution data and student basis contribution data). Characteristics of the schools in the dataset are shown in Table 1.

Table 1
Public School Characteristics

	All Schools	*Poorest Quarter*	*Second Poorest Quarter*	*Second Wealthiest Quarter*	*Wealthiest Quarter*
Number of Schools	929	232	232	233	232
School Enrollment	672,211	142,760	164,087	168,025	197,339
Percent of Total School Enrollment	100.0%	21.2%	24.4%	25.0%	29.4%
Mean School Enrollment	723.6	615.3	707.3	721.1	850.6
Mean Percentage of Students Eligible for F/R Meal Program	51.2%	87.1%	63.3%	40.5%	14.0%
Mean Combined Reading/Math SAT-9 Percentile Rank Score	48.7	30.4	43.3	53.5	66.6

Sources: Arizona Department of Education and Arizona Department of Revenue

Table 1 shows the extent of the differences in the poverty/wealth measure and achievement measure between the quarters formed around relative poverty/wealth of the schools. The mean percentage of students eligible for the F/R meal program represents relative differences in poverty for a school's student body. The overall mean percentage of students eligible for the F/R meal program was 51.2 percent with a standard deviation of 28.01. When viewed by quarters based on poverty/wealth, the mean percentage of students eligible for the F/R meal program ranged from 87.1 percent (SD = 6.94) in the poorest quarter to 14.0 percent (SD = 7.36) in the wealthiest quarter. As for achievement differences represented by Stanford-9 results, the mean combined reading/math individual percentile rank score for all schools was slightly below midpoint at 48.7 (SD = 18.75); for schools in the poorest quarter the score was 30.4 (SD = 11.85) and for the wealthiest 25 percent of schools it was 66.5 (SD = 9.61).

Table 2 accounts for a total of $5,925,436 contributed to 929 public K-12 schools from 53,294 separate donations. 163 schools (17.5%) did not receive any money under this program. A comparison of the distribution of tax credit contributions between the poorest and wealthiest quarters reveals that wealthy

schools received a disproportionately large number of donations as well as a disproportionately large amount of the total resources that were distributed under this program. In terms of the number of contributions, the wealthiest quarter of schools received 29,756 separate donations, a mean of 128.3 (SD = 204.94) donations per school. The poorest quarter received 4,097 separate donations, a mean of 17.7 (SD = 39.62) donations per school. Thus, the wealthiest quarter received 55.8 percent of all contributions while the poorest quarter accounted for 7.7 percent. This resulted in schools in the wealthiest quarter receiving a mean amount of $13,448 (SD = $14,858) and the schools in the poorest quarter receiving a mean amount of $2,859 (SD = $6,763). In the wealthiest group, 5 schools (2.2%) did not receive any money, while in the poorest quarter, 79 schools (34.1%) did not receive any funds. Fully 52.7 percent of the amount contributed to public schools went to the wealthiest 25 percent of schools while the poorest 25 percent of schools received 11.2 percent.

Table 2
School Basis Contribution Data

	All Schools	*Poorest Quarter*	*Second Poorest Quarter*	*Second Wealthiest Quarter*	*Wealthiest Quarter*
Amount Donated	$5,925,436	$663,272	$782,417	$1,359,790	$3,119,958
Percent of Total Amount Donated	100.0%	11.2%	13.2%	22.9%	52.7%
Number of Donations	53,294	4,097	6,218	13,223	29,756
Percent of Total Donations	100.0%	7.7%	11.7%	24.8%	55.8%
Per School Donation	$6,378.29	$2,858.93	$3,372.49	$5,836.01	$13,448.09

Sources: Arizona Department of Education and Arizona Department of Revenue

A regression analysis was conducted to evaluate the relationship between the dependent variable of donation amount to public schools and the independent variable of percentage of a public school's students eligible for F/R meal program. A first-order quadratic regression model provided the best fit between the independent and dependent variables, $R = .409$, $R^2 = .167$, Adjusted $R^2 = .165$, $F(2, 926) = 92.75$, $p < .001$. The beta weight for the independent variable was negative, indicating that schools with higher percentages of students eligible for the F/R meal program (higher poverty) tended to receive lower donation amounts through the tax credit program.

Table 3 presents tax credit donation data on a per student basis. A comparison of the wealthiest quarter and the poorest quarter shows that the wealthiest quarter received an average of $15.81 per enrolled student while the poorest quarter received an average of $4.65, a difference of 70.6 percent. In the wealthiest quarter, there was 1 donation received for every 6.6 enrolled students, compared with 1 donation received for every 34.8 enrolled students in the poorest quarter.

Table 3
Student Basis Contribution Data

	All Schools	*Poorest Quarter*	*Second Poorest Quarter*	*Second Wealthiest Quarter*	*Wealthiest Quarter*
Per Student Donation	$8.81	$4.65	$4.77	$8.09	$15.81
Number of Students Per Each Donation	12.6	34.8	26.4	12.7	6.6

Sources: Arizona Department of Education and Arizona Department of Revenue.

Table 4 presents available data on the distribution of public school tax credits by the claimant's federal adjusted gross income (FAGI). Placing the tax credit claimants into groups based on their FAGI shows that the largest group of claimants (49. 2%), fall into the $50,000 to $100,000 group. This group accounted for 49.1 percent of the total credits for public schools.

Table 4
Public School Tax Credit by Claimants' Federal Adjusted Gross Income

	Total	*$20,000 or less FAGI*	*$20,000 to $50,000 FAGI*	*$50,000 to $100,000 FAGI*	*$100,000 to $500,000 FAGI*	*Over $500,000 FAGI*
Number of donations	16,930	389	3,999	8,322	4,100	120
Percentage of Total Donations	100.0%	2.3%	23.6%	49.2%	24.2%	0.7%
Total Credits	$3,043,456	$65,887	$693,208	$1,493,354	$768,253	$22,754
Percentage of Total Credits	100.0%	2.2%	22.8%	49.1%	25.2%	0.7%
Average Size of Donation	$179.77	$169.38	$173.35	$179.45	$187.38	$189.62

Source: Arizona Department of Revenue (Data as of August 1999)

Findings: Private Schools

According to ADOR tax credit records, there were 15 STOs actively soliciting donations in calendar year 1998. Of these 15 STOs, 10 were religiously affiliated, three were nonreligious, one is of unknown status, and one is no longer active. The 15 STOs reported receiving $1,815,799 from 4,246 separate donations. Table 5 shows the distribution of donations by type of STO. Fully 95.3 percent of the funds donated went to religiously oriented STOs.

Table 5
Donation Data Reported by School Tuition Organizations (STOs)

	Total STOs	*Religious affiliated STOs*	*Nonreligious STOs*	*STOs No longer active or of Unknown status*
Number of STOs	15	10	3	2
Number of Donations	4,246	4,045	132	69
Amount Donated	$1,815,799	$1,731,019	$54,805	$29,975
Percentage of Total Amount Donated	100.0%	95.3%	3.0%	1.7%

Source: Arizona Department of Revenue (Data as of August 1999)

The U.S. Department of Education in the *Digest of Education Statistics, 1999*, estimates that in the fall of 1997 there were 44,991 students enrolled in private elementary and secondary schools in Arizona. From the Fall of 1993 to the Fall of 1997, there was an increase of 1,226 private school students for an average annual increase of 307 students. Applying this rate of increase to the Fall 1997 figures produces a Fall 1998 private school enrollment estimate of 45,298. Therefore, the average per student donation for private schools is estimated to be approximately $40.09 (Table 6).

Table 6
Estimated Per Student Basis Donation Data for Public and Private Schools

	Public Schools	*Private Schools*
Per Student Donation	$8.81	$40.09
Number of Students Per Each Donation	12.6	10.7

Sources: Digest of Education Statistics, 1999 and Arizona Department of Revenue

For the first year of the tax credit, many STOs were reportedly reluctant to distribute revenues for scholarships until the court challenges were decided (Meyer and Smith, 1999). Seven STOs reported information about the amount and numbers of scholarships given (one STO did not provide the number of scholarships given). These data are summarized in Table 7.

Table 7
Scholarship Data Reported by School Tuition Organizations (STOs)

STO	*Number of Donations*	*Total Amount*	*Average Scholarship Amount*
Arizona Scholarships and Tuition	42	$26,360	$627.62
Christian Scholarship Fund of Arizona	163	$68,235	$418.62
Higher Education for Lutherans Program	116	$31,380	$270.52
Northern Arizona Christian School Scholarship	30	$35,000	$1,167.67
St. Gregory/Green Fields Scholarship	82	$32,480	$396.10
Southern Arizona Foundation for Education	56	$22,250	$397.32
Total	489	$215,705	$411.11

Source: Arizona Department of Revenue (Data as of August 1999)

STO reports to ADOR indicated that 417 scholarships (85.3%) averaged below 500 dollars, with 42 (8.6%) between $500 and $1,000 and 30 (6.1%) above $1,000. The low scholarship award amounts suggests that the tax credit is functioning more as a middle class subsidy rather than offering increased access for low income students. Low-income families would likely continue to find it financially difficult to enroll their children in private schools with such low scholarship assistance.

Similar to Table 4 for public schools, Table 8 presents available data on the distribution of tax credits by the claimant's federal adjusted gross income (FAGI), but this time for private schools. Placing the tax credit claimants into groups based on their FAGI shows that the largest group of claimants (40.9%), fall into the $50,000 to $100,000 group. The median FAGI for the $50,000 to $100,000 group was slightly over $70,000. This group claimed 41.7 percent of the total credits for public schools claimed.

Table 8
Private School Tax Credit by Claimants' Federal Adjusted Gross Income (FAGI)

	Total	*$20,000 or less FAGI*	*$20,000 to $50,000 FAGI*	*$50,000 to $100,000 FAGI*	*$100,000 to $500,000 FAGI*	*Over $500,000 FAGI*
Number of donations	2,579	52	492	1,055	906	74
Percentage of Total Donations	100.0%	2.0%	19.1%	40.9%	35.1%	2.9%
Total Credits	$1,133,636	$14,311	$187,130	$472,345	$424,500	$35,350
Percentage of Total Credits	100.0%	1.3%	16.5%	41.7%	37.4%	3.1%
Average Size of Donation	$439.56	$275.21	$380.35	$447.72	$468.54	$477.70

Source: Arizona Department of Revenue (Data as of August 1999)

Compared with public schools, the results for private schools were somewhat more skewed toward the wealthy, with those in the $100,000 to $500,000 FAGI group accounting for 37.4 percent of the STO credits versus 25.2 percent of the public school credits.

Conclusion

Arizona's education tax credit law results in serious inequities in who has access to this credit, and who receives the proceeds. The strongest argument and major justification for this tax credit program was that it would benefit lower income students and offer them increased access to private schooling. Overall, the evidence strongly suggests that lower income students are not benefiting from this program. In public schools, the schools with wealthier families and higher standardized test scores are receiving most of the proceeds from this program while schools with students from poorer families and lower test scores are receiving much less. According to the analysis, 52.7 percent of the total amount contributed went to the wealthiest 25 percent of schools while the poorest 25 percent of schools received 11.2 percent. The average STO scholarship award amount was $411.11, which tends to cast doubt that such scholarships are enabling many low- income students to begin attending private schools.

The evidence also suggests inequity in who has access to this tax credit. The data showed that 75.1 percent of the public school portion of tax credits provided through the education tax credit program went to donators with federal adjusted gross income of $50,000 or more. For private school donations, the results were even more highly skewed toward the wealthy. For private school donators, 82.2 percent of the tax credits claimed went to those with federal adjusted gross income of $50,000 or more.

The tax credit for school tuition organizations that provide scholarships for students attending private or religious schools is almost solely benefiting religiously oriented schools. The data shows that 95.3 percent of all private tax credit donations went to religiously oriented school tuition organizations.

Data for the second year of the Arizona's Education Tax Credit program, only recently made available, show a 60.4 percent increase in public school donations and a 633.3 percent increase in private school donations over the prior year's results. Preliminary indications are that the second year data shows an exacerbation of the trends noted in the first year data (Bland, 2000).

Overall, the evidence from this analysis indicates that students from wealthier families and wealthier donators are the primary beneficiaries of this tax credit statute, rather than low-income students and families. This tax credit has functioned to increase the funding inequity which was already a problem and source of contention in Arizona's school system.

References

Bland, K. (2000, April 9). School tax credits wide open to abuse. *The Arizona Republic*, pp. A1, A22.

Center for Market-based Education. (2000, March 15). Extra Credit: Arizona Tuition Tax Credits. [Online]. Available: http://www.cmbe.org/extracredit.htm. [August 1, 2000]

Kahne, J. (1996). *Reframing educational policy: Democracy, community and the individual*. New York: Teachers College Press.

MacRae, D., & Whittington, D. (1997). *Expert advice for policy choice: Analysis & discourse*. Washington DC. Georgetown University Press.

Meyer, G. & Smith, E. (1999, September 23). Arizona's individual income tax credits for schools. Phoenix: Arizona Department of Revenue.

Peterson, P. E. (1999). Top ten questions asked about school choice. In D. Ravitch (ed.), *Brookings papers on education policy, 1999* . (pp. 371-418). Washington DC: Brookings Institution Press.

Peterson, P. E. (1998). School choice: A report card. In P. E. Peterson & B. C. Hassel (eds.), *Learning from school choice.* (pp. 3-32). Washington DC: Brookings Institution Press.

Surrey, S. S., & McDaniel, P. R.. (1985). *Tax expenditures.* Cambridge MA: Harvard University Press.

Walsh, M. (1999, February 3). Tax credits pass muster in Arizona. *Edweek* [Online]. Available: http://www.edweek.org/ew/vol-18/21ariz.h18 [August 1, 2000]

Weinberg, D. H. (1987). The distributional implications of tax expenditures and comprehensive income taxation. *National Tax Journal, 40*, 237-253.

About the Author

Glen Y. Wilson
College of Education
Arizona State University
Tempe, AZ 85287-0211

Email: glen.wilson@asu.edu

B.S. Arizona State University 1985
M.B.A. Arizona State University 1986
Ed.M Harvard University 1998

Glen Wilson is a Ph.D. student in the Division of Educational Leadership and Policy Studies in the College of Education at Arizona State University. His interests include K-12 education policy with specific interests in the school choice movement as a reform strategy and in issues of fairness and equity related to high stakes testing and school finance.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doc.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografía e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 73 times since August 1, 2000

Education Policy Analysis Archives

Volume 8 Number 39 August 1, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES.**
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Moral Considerations Regarding the Arizona Tax Credit Law: Some Comments

Anthony G. Rud Jr.
Purdue University

Abstract
I begin by commenting on the language used, both by the Arizona tax credit law, and by our commentators, and then turn to a discussion of a factor I believe fuels the impetus for sectarian education. I end with a consideration of questions related to the social, cognitive, and moral costs of such privatization, in contrast to a democratic commitment to education.

This article is one of four on the Arizona Tax Credit Law:

- Welner: Taxing the Establishment Clause
- Moses: Hidden Considerations of Justice
- Wilson: Effects on Funding Equity

Language

Language can mask, or be used to deconstruct, purpose and motive. George Orwell's speaks about the importance of clear expression in *Politics and the English Language* (1946/1981):

> Now it is clear that the decline of a language must ultimately have political and economic causes: it is not due simply to the bad influence of this or that individual writer. But an effect can become a cause, reinforcing the original cause and producing the same effect in an intensified form, and so on indefinitely. A man may take to drink because he feels himself to be a failure, and then fail all the more completely because he drinks. It is rather the same thing that is happening to the English language. It becomes ugly and inaccurate because our thoughts are foolish, but the slovenliness of our language makes it easier for us to have foolish thoughts. The point is that the process is reversible. (pp. 156-57)

Orwell was writing in a different time, but his words apply in many instances today. I hear Orwell when I read about the Arizona tax credit law discussed by Welner, Moses and Wilson. Our authors claim that deception through the use of language has occurred in this issue. The very title of the session at which the papers were originally delivered suggests such linguistic deception. To don a costume, we all know, is to dress up better or differently than we really are.

What kind of costume do our authors tell us that vouchers wear? The term proposed is a "scholarship," implying that academic merit is rewarded and inequity redressed. However, as Welner points out, this is not the case. Our authors claim that more likely terms for the Arizona tax credit law are vouchers, tax credits, and so forth. Indeed the language of "scholarship" is used to manipulate sentiments toward more lofty goals than mere personal gain. Wilson concludes that these scholarships are tax credits, while Moses more bluntly calls this usage a deception.

The Move to Sectarian Education

Such use of language masks an important issue that give impetus for this kind of law. The papers all talk about how religious schools are disproportionately represented in the funding. There is a deeper motivation for such that is not sufficiently discussed in the public debate in Arizona. Why are religious schools chosen overwhelmingly by these parents? What do some parents believe they are not getting from public education that makes them want to opt for this kind of instructional environment for their children?

Warren Nord (1990, 1995) has written on the absence of the study of religion in public schools. He has criticized this lack on curricular grounds, in that religion can explain a great deal about history and other aspects of culture. When a religious explanation for certain events or theories is absent, Nord argues, that event or theory is meaningless.

Unfortunately, a discussion of religion in the public schools brings up many knee jerk responses, and a worry about indoctrination rather than education. This kind of reaction is understandable, however, it confuses the *study* of religion with its *practice*. Certainly this is a fine line, but a line that must be treaded in our public schools, and I believe that it is compatible with a democratic view of education. Leaving out a religious explanation for many phenomena, such as the birth of mathematics, the Crusades, the motivation of a Thomas More, the theories of Copernicus, and so forth, can be criticized on curricular grounds. If religion is left out as a *curricular element*, the student gets an impoverished and incomplete view of how certain events in history came about, as well as the genesis and rationale of certain scientific theories that ground much of the curriculum. I would argue that missing the element of the study of religion in our curricula might contribute to the choice of private, sectarian education by some parents.

However, my advocacy of an element of the study of religion in the curriculum may not satisfy all. Many families choose sectarian education because of a lack of

perceived order and authority in public schools (usually such parents, in my experience, especially complain about profanity). In doing so, they move more toward what has been called a "lifestyle enclave" (Bellah et. al, 1985/1986, p. 335) where an aspect of private life is shared, and consequently, the benefits of a democratic and diverse way of life diminished.

Retreat from Democracy

Let us look at some other items that can be seen through the lens of the retreat from the public and the publicly supported that the Arizona tax credit law permits. Perhaps most distressing to an educator is the learning theory that supports this movement. There is a retreat from a Deweyan learning from others who are different, to a kind of learning within what I termed above a lifestyle enclave. There are benefits from open dialogue. As Dewey pointed out, "A democracy is more than a form of government; it is primarily a mode of associated living, of conjoint communicated experience" (Dewey, 1916/1989, p. 93). One learns from the other, and with learning comes growth.

The notion of freedom that underlies the movement toward sectarian and privatized education is also distressing. As Moses points out, the move to privatization contrasts the clash of individual, atomized, freedom, (her apt phrase is "the politics of disconnected freedom"), to the more fragile notion of contextual, participatory freedom. Our authors point out that similarly, justice takes a back seat in these arrangements too. Democracy is cumbersome and in a sense bothersome, but the alternative leaves out, and leaves behind, too many students and families, as well as offering the chosen families and students a narrow education.

Markets and Education

Sergiovanni (2000) reminds us of the difference between markets and education:

> In markets, individuals, motivated by self-interest, act alone in making preferred choices. Democratic choice, by contrast, is collective, complex, cumbersome, time- consuming, and sometimes combative. Further, and unlike market choices where the will of the majority is not supposed to be imposed on everyone, once a democratic decision is made it applies to everyone. (p. 163)

Efficiency does not equal or even lead to equality. Moses makes a convincing argument in contrasting the libertarian market determined, efficient conception with the liberal democratic, participatory conception. Is the improvement of education best served by the market, or by other forces? Is it a question of money and power, or schooling and justice?

Concluding thoughts

In sum, I am of at least two minds about these issues surrounding the Arizona tax credit law. I look toward democratic participation as essential in schooling. Yet, I want to keep in mind the existential needs seemingly expressed by these parents regarding the need for sectarian education. I believe many of their concerns could be addressed with a robust and critical curriculum that takes into account the role of religion in culture. Since our authors are discussing an issue that is very much alive in Arizona, and in other parts of the country as well, I think it is urgent that we all ask what kinds of action are best suited to bring about and enhance a participatory and democratic ideal. I join many others in being prepared to defend this ideal on moral, and cognitive, grounds.

References

Bellah, R. N., Madsen, R., Sullivan, W. M., Swidler, A., & Tipton, S. M. (1986) *Habits of the heart: Individualism and commitment in American life*. New York: Harper & Row. (Original work published 1985).

Dewey, J. (1989). *Democracy and education*. In J. Boydston (Ed.), The collected works of John Dewey (Vol. 9). Carbondale: Southern Illinois University Press. (Original work published 1916).

Orwell, G. (1981). Politics and the English language. In Orwell, G., *A collection of essays* (pp. 156-71) San Diego: Harcourt Brace and Company. (Original work published 1946).

Nord, W. A. (1990, September). Taking religion seriously. *Social Education*, 287-290.

Nord, W. A. (1995). *Religion and American education: Rethinking a national dilemma*. Chapel Hill: University of North Carolina Press.

Sergiovanni, T. J. (2000). *The lifeworld of leadership: Creating culture, community, and personal meaning in our schools*. San Francisco: Jossey-Bass Publishers.

About the Author

Anthony G. Rud Jr.
School of Education
Purdue University

Email: rud@purdue.edu

Anthony G. Rud Jr. is Associate Dean in the School of Education at Purdue University. He did his undergraduate work at Dartmouth College, and holds a master's and PhD in philosophy from Northwestern University. In addition to his administrative duties, he teaches courses in philosophy of education, and a course in the cultural context of education for doctoral students in educational administration. Rud has also been heavily involved in teacher education curricular reform at Purdue. He is a founding member of the Editorial Board of this journal.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of

Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois— UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 334 times since August 1, 2000

Education Policy Analysis Archives

Volume 8 Number 40 August 1, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Advanced Placement: Access Not Exclusion

Wayne Camara
The College Board

Neil J. Dorans
Rick Morgan
Carol Myford
Educational Testing Service

Abstract

Lichten (2000) argues that increased access to AP courses in high schools has led to a decline in AP quality. He uses a mix of actual data, inaccurate data, and fabricated data to support this hypothesis. A logical consequence of his argument is that a reduction in the availability of AP courses will lead to an improvement in AP quality. In this paper, we maintain that his thesis is flawed because he confounds quality with scarcity. In contrast to his narrow conception of quality, quality in the AP context is subject- specific and multifaceted, embracing course content, the teacher, the student as well as the exam. Increased access will not diminish quality. Instead, increased access exposes students to college-level course material, encourages teachers to expand their knowledge domains, serves as a lever for lifting curriculum rigor, and provides students with the opportunity to experience the challenges associated with advanced placement in college.

Quality. What is quality? How do we measure quality? How do we improve quality? Lichten (2000), in his study "Whither Advanced Placement?," attempts to assess the quality of the Advanced Placement Program®. We believe he fails for several reasons, many of which revolve around his narrow, simplistic definition of

quality. We address these concerns in the following section, entitled "Quality."; Then we point out the many "Inaccuracies, Fabrications, and Leaps of Logic" in Lichten's study; indeed, he seems to use data the way an impassioned partisan would in fashioning an opinion piece for an op-ed page. We then explain in the section "AP® Grades" how AP grade levels are set, since Lichten's lack of understanding of the linkage between AP grades and college standards may have confused readers. Finally, we address the issue of ";Access and Elitism," contrasting Lichten's exclusionary ideal with the College Board's goal of widening the circle of students who have access to AP and its challenging curriculum.

Quality

Any effort to assess the quality of the AP Program must recognize its diversity and complexity, and the fact that each discipline has unique characteristics that must be taken into account. One size does not fit all. Some disciplines are more constant and well defined, which makes it easier to shape AP course descriptions and assess student capability. Other disciplines (such as computer science, for example) are continually evolving; the challenge is to be responsive to anticipated developments in an ever-changing field.

The diversity of students taking AP also adds to the complexity. They do not enter a course with the same level of preparedness for undertaking rigorous college-level course work. Some exam-takers come to the AP course with a head start. The advantage that native speakers of Spanish have in the AP Spanish Language and AP Spanish Literature courses is obvious. A similar, yet less apparent, advantage might be possessed by the children of physicists who might receive preparation for science courses through home-based experiences, when it comes to science courses. As AP offers opportunities to more and more students, the range of backgrounds of these students will increase commensurately.

Lichten ignores this diversity and complexity to promote his viewpoint. To him, quality can be captured in a simple operational definition: the ratio of the number of advanced placements made by colleges to the number of AP examinations taken, regardless of the subject area or the preparation of the students. By this standard, AP Spanish Language is a high quality examination because its many native Spanish speakers are very likely to receive advanced placement credit. Conversely, the AP Chemistry exam is lower in quality because the corresponding ratio is not as high as for AP Spanish Language.

This narrow, simplistic definition of quality is flawed for several reasons. First, the ratio is subject to many factors that have little or nothing to do with quality. For example, students vary with respect to the preparation they bring to the AP course, and their performance on the exam may reflect their varied backgrounds. This affects the top part of the ratio. External factors, such as certain legislative initiatives that provide payment for students' AP Examination fees, will increase the number of students who take AP exams, which in turn affects the bottom part of the ratio. Neither preexisting differences in preparation nor external initiatives affect the quality of the AP course or its examination (or the scoring or grade standards for the exams), yet they affect the ratio definition of quality Lichten uses.

Second, Lichten ignores the distinct nature of each AP course by aggregating results across all courses; for example, treating a 3 on the AP Spanish Language exam as if it means the same thing as a 3 on the AP Chemistry exam. Quality is a complex concept. Ignoring the fact that each course and exam is unique is akin to treating all elements as if they had the same atomic weight. Any serious scholarly treatment of the AP Program must recognize the uniqueness of each course.

Third, and most critical, Lichten's definition confuses quality with scarcity. Scarcity does not improve quality; it merely alters the context from which we judge it. He argues that access to AP must be restricted or limited in order to restore AP quality. This sounds like an OPEC argument with respect to oil production. Limit oil production (access to AP courses), and the price of oil will rise (Lichten's quality index will increase). Certainly, the price of oil will increase. But will its quality increase? Of course not. Likewise, restricting access to AP courses will make the number of qualified candidates smaller. But will it increase the quality of the AP courses and examinations?

Instead of viewing knowledge in disciplines as the exclusive domain of a selected few, the AP Program employs a model based on access. The more people know about math and the sciences, music and the arts, and languages, the more they and society will profit from this knowledge. AP is rooted in the meritocratic principles that led to the foundation of ETS by the College Board and other parties interested in tapping the potential that lay within America (Lemann, 1999). AP was never to be a barrier to access. Instead it should serve as an avenue for access. Students should be encouraged to maximize their capabilities. Quality, as AP defines it, should be measured by the number of students who have been positively influenced by taking AP courses, rather than by the ratio of the number of advanced placements to the number of exams administered.

The College Board states in its publication *A Guide to the Advanced Placement Program* (The College Board, 1999), "There are many benefits for students who take AP courses. They can study subjects they are interested in and challenge themselves with students who are similarly motivated. AP often helps steer students who are unsure about future plans toward college or advanced studies...AP prepares students for the future by giving them tools that will serve them well throughout their college career (p. 6)." The quality of the AP Program is multidimensional and rests on three pillars of quality: fair, valid, and reliable assessments; rigorous introductory college-level curricula; and exemplary teacher professional development. AP strives to ensure that the exam scoring and scaling are accurate and of high quality (as measured by statistical/psychometric indices of accuracy, reliability, and validity). Teacher quality and student preparedness are important factors that also influence quality.

Quality also manifests itself in the effects that AP has on students who take the courses but do not take the exam or who do take the exam but do not seek or receive college credit or advanced placement. By Lichten's standards, a student appears on the quality side of the ledger only if she receives advanced placement at the university she attends. Therefore a student who has a 3 on an exam will not receive advanced placement at a college that requires a 4, but will receive it at a college requiring a 3. If the student goes to the college requiring the 4, she is a debit on the quality ledger; if she goes to the other college, she is a plus on the Lichten index. From the AP perspective, the in-depth exposure to the discipline and quality instruction that the student received are the same regardless of which college she attends. She learned from the course; the existence of the course at her school enhanced the overall value of education at that school. While difficult to quantify, it is hard to argue that the existence of AP courses at more schools hurts quality, unless the definition of quality that one adopts confounds scarcity with quality.

Finally, AP quality is carefully monitored within each subject domain. AP, as a matter of course, strives to ensure that the exam, grading, scaling, and scoring are accurate and of high quality (as measured by statistical/psychometric indices of accuracy, reliability, and validity). Enhancing course quality is an important component of the AP process as well. Teacher professional development and student preparedness are important factors that also influence quality.

Inaccuracies, Fabrications, and Leaps of Logic

In addition to using a narrow, simplistic definition of quality, Lichten (2000) commits several serious errors in scholarship and makes erroneous assumptions about the use and utility of AP.

Table 6 is filled with inaccuracies. The number of exams is misreported by 10,000 in 1980 and by over 100,000 in the speculation for 2000. The basis for the percent of qualifying grades is never stated for any year and is thus left to the imagination of the reader. If one assumes that the author is using the percent of AP grades of 3 or higher, the percentage for 1960 is 49% rather than 75%. In 1970, 66% of AP grades were 3 or higher rather than the 75% Lichten reported. Likewise, the percentage for 1980 is off by 1% and the actual percentage for 1990 differs by 4%. The basis for any of the entries for 2000 and 2010 appears to be pure speculation, as are the percentages qualifying for earlier years. Due to the inaccuracies in the left-hand side of the table, the right-hand side errors are substantial (10% inaccuracy in the last column for 1980). The fabrications in the data throughout the entire paper

call to question the quality of the scholarship of the document and the inferences made from them.

Lichten creates a table of SAT and AP data from ETS and College Board sources. In preparing this table, he assumed that the college associated with each examinee was the college that the student attended. This is correct for students who sent grades to only one college. For those who sent grades to multiple colleges, the college in the Lichten data was the last one on the student's list of colleges. This reality calls into question the validity of his assumption (which would hold true only if every student went to the college that was last on their lists), and any inferences that depend on the validity of the assumption.

Table 2 is not only based on a questionable assumption, it also appears to involve unacknowledged estimation on the part of the author. He states that "55% of 3s pass." Unless Lichten contacted every college for their numbers of AP grades of 3, numbers of AP 4s, and their numbers of AP grades of 5 received, he is stating as fact something that he is fabricating. As discussed earlier, Table 6 shows that his estimations are often quite inaccurate.

The text indicates that the data in Table 5 were obtained from ETS. Standard practice is to cite where the data have been published before, and which colleges supplied data. It addition, it would have been helpful to know what constituted remedial classes to calculus. While focusing on the 24% (the paper incorrectly states 22%) of students with AP grades of 3 who took the second or third calculus as their first mathematics course, Lichten again misses the point about the benefits of AP. Exposing students to a rigorous college-level course at high school surely has many benefits.

It is clear that the study is unbalanced in its treatment of the issues. When there is competing evidence that refutes his assumptions, Lichten chooses not to cite it. Likewise, when there are alternative explanations for the findings he cites, those interpretations are not posited, even in a footnote. Selective citation may be acceptable in op-ed pieces, but it has no place in a scientific journal. Some examples follow:

- Lichten cites a lawsuit against the University of California as evidence against the AP Program. The plaintiffs argue that access to AP must be extended to all California high school students in order to make the admissions playing field more level. This increased access would actually damage quality as defined by the Lichten index. Thus, Lichten uses a lawsuit that advocates greater access to AP to argue against greater access to AP.
- The author uses a quotation from Bowen and Bok (1998) about the need for government to respect the autonomy of colleges as evidence that the College Board and Bowen and Bok disagree with respect to government involvement in AP. The author uses a leap in logic to infer that Bowen and Bok are opposed to government involvement in reducing student fees for the economically disadvantaged and in supporting governmental funding of teacher professional development. Is this what Bowen and Bok had in mind when they argued against government intervention in academic matters?
- The author claims "This disparity [between the College Board's grade equivalent recommendations and the cut points used by some colleges for advanced placement and/or college credit] is a sign of the remarkably poor communication between colleges and the College Board." As explained below in the section "AP Grades," the AP grade recommendations reflect empirical results from college comparability studies; when they differ from specific institutional cut points it is not based on lack of communication, but on different judgements by faculty about the level of performance they believe should be expected. Lichten bases his argument largely on his realization that colleges have their own admissions and placement policies. The College Board has no desire to tell any college what it should or should not require of students for admission or placement. Certainly, institutions vary in what they expect in terms of GPA, SAT, participation in extracurricular activities, as well as in AP requirements. These differences do not invalidate any of these measures or claims about general preparedness.
- Lichten cites Morgan and Ramist (1998) as having collected data from

colleges that receive large numbers of AP grades, but he ignores the conclusions of the study that support the awarding of advanced placement. Morgan and Ramist found that AP students performed well in upper-level courses after being placed out of the introductory courses. For the majority of these upper-level courses, students with AP grades of 3 had higher course-grade averages than those students who had taken an introductory course prior to the upper-level course.

- Lichten asserts that the majority of AP faculty consultants should come from colleges. Moreover, he dismisses college faculty who teach at community colleges and describes faculty from some four-year institutions as coming from "typically very low-level institutions." We wonder how Lichten arrived at his quality judgements of college faculty in all 32 AP subject areas. In addition, the author fails to report that the number of AP faculty consultants from four-year colleges is larger today than ever before.
- Lichten also fails to note that the curriculum for an AP course is based on curriculum surveys of the colleges who receive the most AP grades for that content area. Furthermore, college faculty members serve on the AP Developmental Committees that create each exam. The Chief Faculty Consultant, who is in charge of the free-response scoring, also serves as a very strong link to college faculty. In addition, when major changes are made to the AP curriculum (for example, graphing calculators being integrated into the teaching of calculus and computer languages changing), representatives from the disciplines' professional organizations participate in the development effort.

Finally, stating as truth something that is the author's opinion is a pervasive problem in the study. Several statements call for citations, but none are present. Here are some examples:

- "Some colleges, not all highly selective, will not accept a 5" for AP credit. Table 2 and the associated text provide no specifics.
- "A serious source of disagreement between College Board and higher education faculty is the increasing number of legal restrictions."
- "College faculty and deans cast a jaundiced eye on mandatory high school participation, which they view as dragging in schools that are not qualified to handle AP."
- "The College Board's qualification estimates (Table 1), backed by mandates in a growing number of states, would require acceptance into advanced courses of candidates with a score of '3'."
- "The pressure from mandates is on college faculty either to go along and lower quality or to misreport their AP policy."
- "With few exceptions, national and state standardized tests fail to cover abilities needed in college."

AP Grades

Lichten contends that the College Board's grade equivalents for AP courses are misleading because colleges use different standards for awarding college credit or advanced placement. There are flaws in this argument.

The AP grade equivalents are empirically established through research that compares student performance on AP Examinations with the grades students achieve in comparable introductory courses at college. Such grade equivalency studies are conducted with college students attending a range of colleges.

Typically, instructors at the 200 colleges receiving the largest number of AP grades for the AP Exam under evaluation are asked to have their students take portions of the appropriate AP Exam under motivated conditions. The lowest composite score that earns an AP grade of 5 is set to represent the average performance equivalent of college students who earn grades of A from their instructor on the AP Exam. The lowest composite score that earns an AP grade of 4 represents the average performance level equivalent of college students who earn grades of B from their instructor on the AP Exam. The lowest composite score that

earns AP grades of 3 and 2 represents those college students receiving grades of C and D, respectively, on the AP Exam. Thus, the AP grade scale reflects a consistent standard of student performance that is empirically related to college grades.

Lichten asserts that the AP grade scale is misleading and that a "yawning gap" is created between AP grades and college grading policies because some colleges and departments reject the AP recommendation for awarding credit and/or advanced placement to students with an AP grade of 3 as evidence that AP grades are misleading. Individual colleges, and often individual academic departments, establish their own policies for awarding college credit and/or advanced placement for a particular AP grade. It is the specific AP grades that individual colleges use and the course grades at these colleges that differ widely, as perhaps they should. The standard embodied in an AP grade level on a particular exam, e.g., AP Calculus, is the same across institutions; institutional use of AP grades varies across institutions.

Access and Elitism

The most disturbing aspects of the Lichten report are the repeated statements and inferences that the quality of the AP Program could only be maintained "as long as AP served a small, elite population chosen from selective schools (p.13)." Additional statements that minority students are not likely to succeed in AP and that better selection of students into AP courses is required to reestablish AP quality are equally troubling. AP data do illustrate that African-American students and Hispanic students generally perform less well on AP Exams than do Asian-American students and White students. Nevertheless, African-American students and Hispanic students can and do succeed in AP. For example, in the last year, there was a 23% increase over the previous year in the number of African-American students who received AP grades of 3 or higher in Charlotte-Mecklenburg, North Carolina.

In the 1999-2000 academic year, the AP Program consisted of 32 college-level courses delivered in approximately 13,000 schools to over 700,000 students who completed more than 1.25 million exams. The net impact of AP is that many more students are taking rigorous and challenging introductory college-level courses while in high school. Some of these students may elect not to take the AP Examination, others may take the Examination but not meet an individual college's requirement for advanced placement, and others may be entitled to advanced placement in a subject but not elect to place out of the introductory course. Yet most, if not all, of these students will have benefited from participating in AP. And, as more students complete AP courses, more teachers are completing AP professional development and mastering the teaching of challenging courses and preparing students in earlier grades to be ready for AP-level work in high school. The net effect is to raise academic standards throughout middle and high school and greatly expand the pool and diversity of students exposed to challenging AP courses.

In 1979, only 485 African-American and Hispanic students took Calculus AB. Forty-eight percent (236 of 495) of those students earned grades of 3 or higher. In 1999, the number of African-American and Hispanic students earning grades of 3 on the Calculus AB exam increased to 4,889 (a 2072% increase). Lichten may point out that the percentage of AP grades of 3 for these students decreased from 48% to 41%, but one should also note the increase in opportunity for African-American and Hispanic students. Nearly ten times more African-American and Hispanic students received AP grades of 3 or higher in 1999 than even took the AP Calculus AB Exam in 1979. In fact, in a recent publication, Lichten and Wainer (2000) state "...the PSAT-AP relation tells us that a major expansion of advanced placement achievement is possible in this country in all types of schools: inner city, high-performing suburbs, and just garden-variety schools. A doubling of the number of AP students is not only possible, but is likely within the next decade or so (p. 223)."

Yet in his study, the same author recommends reducing access to challenging courses such as AP to "only a small minority of above average high school students." The author is opposed to legislative efforts to prepare more students for success in AP and other rigorous courses through expanded teacher development and initiatives in the middle schools. Restricting access to only the highest ability students attending the most selective high schools is elitist and runs counter to the

goals and mission of AP and the College Board. The author attempts to construct a rationale for restricting access to AP and turning back the clock, based on half-truths, constructed data, and selective citations. He does not cite his sources and ignores research suggestive of alternatives. We believe his study does not meet even the minimal scholarly standards for a scientific publication and we reject the unsupported assertions made throughout.

Note

The order of authorship is alphabetical. The work was a collaboration. The views in this article represent the opinions of the authors and not those of the College Board or the Educational Testing Service. The paper was enhanced significantly by the authors following suggestions from Janet Cook, Drew Gitomer, Lee Jones, and Walter MacDonald.

References

Bowen, W.G. & Bok, D. (1998). *The shape of the river: Long-term consequences of considering race in college and university admissions*. Princeton, NJ: Princeton University Press.

Lemann, N. (1999). *The big test: The secret history of the American meritocracy* . New York, NY: Farrar, Straus, and Giroux.

Lichten, W. (2000, June 24). "Whither Advanced Placement?" *Education Policy Analysis Archives, 8,* (29). [Online] Available http://epaa.asu.edu/eppa/v8n29.html [Augiust 1, 2000]

Lichten, W. & Wainer, H. (2000). The aptitude-achievement function: An aid for allocating educational resources with an Advanced Placement Example. *Educational Psychology Review 12* - (2), 201-228.

Morgan, R. & Ramist, L. (1998). Advanced Placement students in college: An investigation of course grades at 21 colleges. (Statistical Report 98-13). Princeton, NJ: Educational Testing Service. Available http://collegeboard.org/ap/colleges/sr-98- 13.pdf.

About the Authors

Wayne J. Camara
Office of Research and Development
The College Board
45 Columbus Ave.
New York, NY 10023
212-713-8069
fax 212-649-8427

Email: wcamara@collegeboard.org

Wayne J. Camara is the Vice President for Research and Development at The College Board. He is responsible for monitoring, coordinating and conducting all research and product development associated with the range of College Board assessments, services, and programs. He has served as the Assistant Executive Director of Science at the American Psychological Association (APA) directing scientific involvement in policy and research activities. His principle areas of research are test validity, selection and admissions testing, standards and professional practice in testing, legal and regulatory issues relating to assessment, and public policy issues in assessment. Dr. Camara completed a Ph.D. in industrial-organizational psychology at the University of Illinois at Champaign-Urbana.

Neil J. Dorans is a Principal Measurement Statistician at Educational Testing Service. He is currently the statistical coordinator for the Advanced Placement Program. He has extensive experience in the statistical work associated with large-scale high-stakes testing programs, such as the SAT I. Dr. Dorans was the architect for the recentered SAT I and II scales. He also developed a flexible, easy-to-use method for assessing differential item functioning for selected choice and constructed response items. Dr. Dorans completed a Ph. D. in quantitative psychology at the University of Illinois at Champaign-Urbana.

Rick Morgan is a Program Administrator at Educational Testing Service for the Advanced Placement Program. During the 1990s he served as the statistical coordinator for several testing programs including AP. He has published research in the areas of exam validity, constructed response testing, and the impact of allowing examinee choice. Dr. Morgan completed his Ph. D at The Ohio State University in quantitative psychology and later was a post-doctoral fellow in measurement at Indiana University.

Carol Myford is a Senior Research Scientist in the Center for Measurement Models at Educational Testing Service. Her program of research at ETS focuses on scoring issues in performance and portfolio assessments. She has conducted studies related to rater training, designing scoring rubrics, quality control monitoring, improving rater performance, and detecting different types of rater errors. Dr. Myford received her doctoral degree from the University of Chicago.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger,
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 4739 times since August 19, 2000

Education Policy Analysis Archives

Volume 8 Number 41 August 19, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

The Myth of the Texas Miracle in Education

Walt Haney
Boston College

Abstract:
I summarize the recent history of education reform and statewide testing in Texas, which led to introduction of the Texas Assessment of Academic Skills (TAAS) in 1990-91. A variety of evidence in the late 1990s led a number of observers to conclude that the state of Texas had made near miraculous progress in reducing dropouts and increasing achievement. The passing scores on TAAS tests were arbitrary and discriminatory. Analyses comparing TAAS reading, writing and math scores with one another and with relevant high school grades raise doubts about the reliability and validity of TAAS scores. I discuss problems of missing students and other mirages in Texas enrollment statistics that profoundly affect both reported dropout statistics and test scores. Only 50% of minority students in Texas have been progressing from grade 9 to high school graduation since the initiation of the TAAS testing program. Since about 1982, the rates at which Black and Hispanic students are required to repeat grade 9 have climbed steadily, such that by the late 1990s, nearly 30% of Black and Hispanic students were "failing" grade 9. Cumulative rates of grade retention in Texas are almost twice as high for Black and Hispanic students as for White students. Some portion of the gains in grade 10 TAAS pass rates are illusory. The numbers of students taking the grade 10 tests who were classified as "in special education" and hence not counted in schools' accountability ratings nearly doubled between 1994 and 1998. A substantial portion of the apparent increases in TAAS pass rates in the 1990s are due to such exclusions. In the opinion of educators in Texas, schools are devoting a huge amount of time and energy preparing students specifically for TAAS, and emphasis on TAAS is hurting more than helping teaching and learning in Texas schools, particularly with at-risk students, and TAAS contributes to retention in grade and dropping out. Five different sources of evidence about rates of high school completion in Texas are compared and contrasted. The review of GED statistics indicated that there was a sharp upturn in numbers of young people taking the GED tests in Texas in the mid-1990s to avoid TAAS. A convergence of evidence indicates that during the 1990s, slightly less than 70% of students in Texas actually graduated from high school. Between 1994 and 1997, TAAS results showed a 20% increase in the percentage of students passing all three exit level TAAS tests (reading, writing and math), but TASP (a college readiness test) results showed a sharp decrease (from 65.2% to 43.3%) in the percentage of students passing all three parts (reading, math, and writing). As measured by performance on the SAT, the academic learning of secondary school students in Texas has not improved since the early 1990s, compared with SAT takers nationally. SAT-Math scores have deteriorated relative to students nationally. The gains on NAEP for Texas fail to confirm the dramatic gains apparent on TAAS. The gains on TAAS and the unbelievable decreases in dropouts during the 1990s are more illusory than real. The Texas "miracle" is more hat than cattle.

Click on items in list below for full text.

- Part 1: Introduction
- Part 2: Recent History of Testing in Texas
- Part 3: Evidence and Boosters of the Myth
- Part 4: Problems with TAAS
- Part 5: Missing Students and Other Mirages
- Part 6: Educators' Views of TAAS
- Part 7: Other Evidence on Education in Texas
- Part 8: Summary and Lessons from the Myth Deflated
- Notes and References

- Appendices

About the Author

Walt Haney
Center for the Study of Testing, Evaluation and Educational Policy
Campion Hall 323
Lynch School of Education
Boston College
Chestnut Hill, MA 02467
617-552-4199
617-552-8419 (Fax)
Email: haney@bc.edu
Home page

Walt Haney, Ed.D., Professor of Education at Boston College and Senior Research Associate in the Center for the Study of Testing Evaluation and Educational Policy (CSTEEP), specializes in educational evaluation and assessment and educational technology. He has published widely on testing and assessment issues in scholarly journals such as the *Harvard Educational Review, Review of Educational Research,* and *Review of Research in Education* and in wide-audience periodicals such as *Educational Leadership, Phi Delta Kappan, the Chronicle of Higher Education* and the "Washington Post." He has served on the editorial boards of *Educational Measurement: Issues and Practice* and the *American Journal of Education* and on the National Advisory Committee of the ERIC Clearinghouse on Assessment and Evaluation.

ACKNOWLEDGEMENTS

As I worked on this article over a period of more than two years, literally dozens of people helped me in numerous ways. At Boston College, graduate students Cathy Horn, Kelly Shasby, Miguel Ramos and Damtew Teferra helped on specific portions of the work reported here. Damtew Teferra has been especially helpful over a period of nearly two years in tracking down references and other source material and in checking accuracy of data input. Ed Rincon and Terry Hitchock both helped me more than once as I sought to build the data set on enrollments in Texas public schools over the last quarter century. Among many scholars who have answered repeated questions and kindly provided me with references and reviews of various portions of this article in many different versions, I thank James Hoffman, John Tyler, Jeff Rodamar, Angela Valenzuela, Bob Hauser, Duncan Chaplin, Richard Murnane, Linda McNeil, Dennis Shirley, Anne Wheelock, and Janet Baldwin. Diane Joyce, Lauren McGrath, Anna Geraty, Courtney Danley, and Genia Young helped me, as they do everyone else in our research center, in ways too many to mention here. Also, I thank Jane Hodosi of Market Data Retrieval who, under an extraordinarily short deadline, helped me get the mailing labels that allowed us to carry out two of the surveys recounted here. Thanks too to Chris Patterson for providing me with a great deal of useful information. Three electronic mail correspondents—Audrey Amrein, Craig Bolon and Alcin Jehlen also provided me with helpful suggestions and encouragement.

I also wish to express my appreciation to Judge Edward Prado. Though I think he may have erred in his ruling in the GI Forum case (as may be apparent from what is to follow), during the four days I was on the stand in his courtroom, he treated me with attention, respect and good humor. He even had the good sense to tell me simply to "answer the question," when the professor in me launched into discussions of literature on topics on which I was questioned. My wife, Kris, and daughter, Elizabeth, also deserve great appreciation for their tolerance in putting up with work that I told them many times would be done long before now. Thanks also to Gene V Glass who encouraged me to submit this work to *Education Policy Analysis Archives*. As a former editor, I know how hard it sometimes can be to pry manuscripts away from authors who know that there are always other nooks and crannies to explore. Thanks too to nine anonymous reviewers from the *EPAA* Editorial Board who commented generously on a previous version of this article.

More than anyone else, though, I wish to express my appreciation and respect for Al Kauffman. Over the past two years, on several occasions I have cursed him under my breath (and once or twice aloud), for getting me involved with TAAS and education reform in Texas. But after spending more than twice as long on this topic as I ever thought I would, I have developed enduring respect for Al, his integrity and good humor, and his quest for truth and justice. I regret that I was not able to complete all of the analyses reported here before the TAAS trial. But it will be a *long* time before I let Al talk me into working on another case, even if next time he tries to tell me I am not his second choice as an expert witness.

Any errors of fact or interpretation in this report are, of course, despite the enormous help of many good and generous people, entirely my responsibility. No corporations, foundations or anonymous donors have supported the research reported here. But I do owe an enormous debt of gratitude to Boston College for awarding me a sabbatical leave during the 1999-2000 academic year. Without the leave, there is no way I would have been able to complete this research. I did not do what I said I would when I applied for sabbatical leave, but I hope that the work reported here will win me, if not forgiveness, at least tolerance for being distracted from well-intentioned plans.

BEST COPY AVAILABLE

And on the topic of forgiveness, I am almost certain there are people I should have thanked here but could not remember. Forgive me, please, but I simply *had* to finish this work before returning to normal academic duties in September.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 451 times since August 21, 2000

Education Policy Analysis Archives

Volume 8 Number 42 August 21, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES.**
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Texas Gains on NAEP: Points of Light?

Gregory Camilli
Rutgers, The State University of New Jersey

Abstract:
The 1992-1996 gain in mathematics scores on NAEP from 4th to 8th grades in Texas is placed in perspective. The "miracle" in Texas looks much like the median elsewhere. Of 35 states and two districts (Guam and D.C.), the 52-point gain of Texas was good enough to earn Texas a rank of 17th or about the 46th percentile. Taking into consideration the wealth of states, Texas stands in the middle of the pack—no worse than most other states in delivering educational services to students.

Haney (2000) examined a number of aspects of the Texas record of educational progress. This brief response concerns one particular indicator: the 1992-1996 gain in mathematics scores from 4th to 8th grades as measured by the National Assessment of Educational Progress (NAEP). In terms of the NAEP scales scores—not the achievement level percentages—the Texas gain from 1992-1996 was about 49 points. In any metric, this represents a sizable gain. In order to give some perspective to this accomplishment, it is customary to compare states. Implicitly, the rationale for doing so is that some states do better than others, and through a process of competition and selection the level educational level of students can be bootstrapped. Since the Texas gain was the largest of any state, it could be argued that there is much merit in its methods and efficiencies.

However, Haney raised a number of questions about whether this was a gain in achievement or whether it could be attributed to a large degree to changed in grade retention and dropout rates. There is a study on the 4th-8th grade mathematics gains that Haney did not consider which is relevant to this point. The Math cohort study by Barton et al (1998) estimated gains in math for a cohort of students in 4th grade who attended 8th grade four years later. To those who look to statistics to support the educational record of Texas (and to those who would take credit for the miracle), there is good news and bad news in this study.

First, the good news. In the cohort study, Texas students gained about 52 points from 4th to 8th grade. Thus, unless students are retained in the 4th and 5th-8th grades disproportionately, there can be little question that the NAEP scores have gone up substantially. (Haney shows that for grades 2-8, the transition ratios are uniform. Questions arise in the 9th-10th grade transition.) But in regard to a comparison among states, the miracle in Texas looks much like the median elsewhere. Of 35 states and two districts (Guam and D.C.), the 52-point gain of Texas was good enough to earn Texas a rank of 17th or about the 46th percentile. Though Texas outranked four other states by less than one point, it should also be mentioned that six states outranked Texas by less than one point.

This latter finding brings up a central point in the NAEP mathematics results for 1992 and 1996. In fact, the states are pretty well bunched up in the middle. In terms of statistical significance, Texas is different only from Guam (with a 40-point gain), and is not significant from Nebraska (ranked 1st with a 57-point gain). Was there a miracle in NAEP gains from 1992 to 1996 in Texas? The answer very clearly is no. Texas was average.

One more simple representation helps to illustrate this latter point. In Figure 1, the state cohort gains are plotted against median state income (average across 1995-1997). Though a slight linear trend is evident (with Arizona and Hawaii being negative outliers), the story is relatively clear once more. With respect to wealth, which is one of the most reliable predictors of achievement, Texas stands in the middle of the pack—that is, no worse than most other states in delivering educational services to all students. Certainly, there is no criticism that can be leveled against Texas that cannot also be leveled against others states. However, within a paradigm that promotes healthy competition among states as a means of developing effective education policy, the points of light in Texas are not beacons.

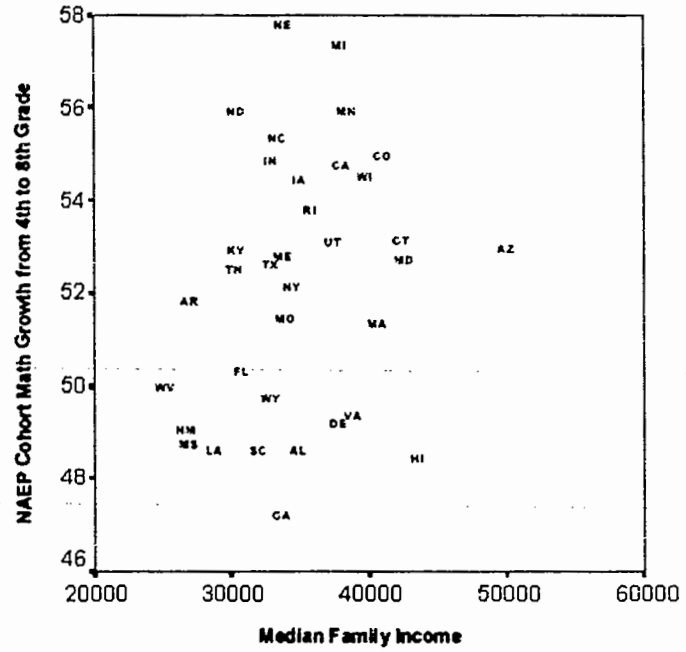


Figure 1. 1992-1996 NAEP cohort gains in mathematics plotted against median family income.

About the Author

Gregory Camilli
Email: camilli@rci.rutgers.edu
Gregory Camilli is a professor in the Rutgers Graduate School of Education, and former Chair of the Department of Educational Psychology. His interests include measurement, program evaluation, and policy issues regarding student assessment. Dr. Camilli teaches courses in statistics and psychometrics, structural equation

modeling, and meta-analysis. His research interets include efficacy studies of Head Start, implementation variability in cooperative learning and technology, and factors related to differential item functioning.

References

Haney, W. (2000). The Myth of the Texas Miracle in Education. *Education Policy Analysis Archives, 8*(41). Available online at http://epaa.asu.edu/epaa/v8n41/ (August 20, 2000)

Barton, P. and Coley, R. (1998). Growth in School: Achievement Gains From the Fourth to the Eighth Grade. Princeton, NJ: Educational Testing Service.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Andrew Coulson
a_coulson@msn.com

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Richard M. Jaeger
University of North Carolina— Greensboro

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
New York University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
Ann Leavenworth Center for Accelerated Learning

Jay D. Scribner
University of Texas at Austin

Michael Scriven

Robert E. Stake

scriven@aol.com
University of Illinois—UC

Robert Stonehill
U.S. Department of Education

Robert T. Stout
Arizona State University

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 223 times since August 21, 2000

prior vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 8 Number 43 August 21, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Consistency of Findings Across International Surveys of Mathematics and Science Achievement: A Comparison of IAEP2 and TIMSS

Michael O'Leary
Thomas Kellaghan
St Patrick's College, Dublin

George F. Madaus
Albert E. Beaton
Boston College

Abstract

The investigation reported in here was prompted by discrepancies between the performance of Irish students on two international tests of science achievement: the Second International Assessment of Educational Progress (IAEP2) administered in 1991 and the Third International Mathematics and Science Study (TIMSS) administered in 1995. While average science achievement for Irish 13-year-olds was reported to be at the low end of the distribution representing the 20 participating countries in IAEP2, it was around the middle of the distribution representing the 40 or so countries that participated in TIMSS at grades 7 and 8. An examination of the effect sizes associated with mean differences in performance on IAEP2 and TIMSS indicated that the largest differences are associated with the performance of students in France, Ireland and Switzerland. Five hypotheses are proposed to account for the differences.

Introduction

International comparative studies of student achievement have become part of

the educational landscape over the past four decades. In these studies, a number of countries (usually represented by research organizations) agree on an instrument to assess achievement in a curriculum area, the instrument is administered to a representative sample of students at a particular age or grade level in each country, and comparative analyses of the data obtained are carried out. The most frequently assessed areas have been reading, mathematics, and science at ages 9 or 10 and 13 or 14. The number of participating countries has grown from 12 in a pilot project conducted between 1959 and 1961 to over 40 for a survey of mathematics and science achievements in 1995 (see Goldstein, 1996; Husén & Postlethwaite, 1996; Kellaghan, 1996).

The potential of international studies to contribute to policy formation was made clear from the earliest studies (Husén, 1967; Lambin, 1995). Over the years, a range of purposes to which information derived from such studies might be put has been suggested. These include the pursuit of equity goals, setting priorities, assessing the effectiveness and efficiency of the educational enterprise and the appropriateness of curricula, evaluating instructional methods and the organization of the school systems, and providing a mechanism for accountability (Kellaghan & Grisay, 1995; Plomp, 1992). While we have relatively little information on the extent to which the findings of international studies have in fact been utilized, there is no doubt that they attract considerable media and public attention.

A variety of factors can affect the extent to which data obtained in an international study accurately reflects what students have learned in the participating countries, something that is necessary if valid comparisons between countries are to be made (see Brown, 1996, 1998; Goldstein, 1996; Kellaghan, 1996; Kellaghan & Grisay, 1995; Murphy, 1996; Nuttall, 1994). One relates to the adequacy of a uniformly administered assessment procedure to measure the outcomes of a variety of curricula. Since curricula differ from country to country, an assessment instrument will not reflect the curricula of all countries participating in an international study to the same degree.

The second factor relates to the extent that the populations and samples of pupils for whom data are obtained can be regarded as equivalent. Defined target populations may not be comparable across countries since exclusion practices may differ (e.g., relating to students with handicapping conditions/learning problems or when the language of the assessment instrument differs from the language of the school). Differences in participation rates of selected samples (due to lack of co-operation from schools, student absenteeism) will make matters worse.

Many commentators have considered how these problems impact on comparisons based on a single study. Additional problems arise when the findings of two different surveys are being compared. In the case of IAEP2 and TIMSS, instruments used to measure achievement differed in form and content sampled, age-based versus grade-based populations definitions were used, and different methods of data manipulation were utilized.

The investigation reported here was prompted by discrepancies between the performance of Irish students on tests of science in the Second International Assessment of Educational Progress in Mathematics and Science (IAEP2) (Lapointe, Askew & Mead, 1992) in 1991 and, four years later in the Third International Mathematics and Science Study (TIMSS) (Beaton, Mullis, Martin, Gonzalez, Kelly & Smith, 1996a; Beaton, Martin, Mullis, Gonzalez, Smith, & Kelly, 1996b). Initially, the intention was to focus on the Irish problem but, as the investigation proceeded, it became clear that discrepancies in performance between the two surveys were not confined to Irish students.

In this article, we first present brief descriptions of IAEP2 and TIMSS. We then select 12 countries that participated in both surveys for further analyses: Canada, England, France, Hungary, Ireland, Korea, Portugal, Scotland, Slovenia, Spain, Switzerland, and the United States. Our approach to assessing the consistency of countries' performances is based on an examination of the performance of each country relative to the performance of other countries in both surveys. If results are stable, differences in performance between countries should not vary very much from one survey to the next. To the extent that they do, findings may be regarded as unstable. Change in effect sizes between pairs of means on the two assessments were calculated to obtain an estimate of the magnitude of differences between

performance on the two occasions.

IAEP2 and TIMMS

In IAEP2, representative samples of 9 and 13-year-olds in 20 countries were assessed in mathematics and science in 1991 (Lapointe, Askew & Mead, 1992). In TIMSS, the mathematics and science achievements of students in grades 3, 4, 7, 8, and in the final grade of secondary education were assessed in 1995 (Beaton et al., 1996). Data are reported in our article for 13- year-olds in IAEP2 and for grades 7 and 8 students in TIMSS. However, the main focus is on grade 7 performance, since in all countries that had participated in both assessments, except Scotland, more 13 year-olds were in grade 7 than in grade 8 (Beaton et al, 1996a, p. A12).

The IAEP2 tests for 13-year olds were contained in two separate booklets, each of which had to be completed by students in four 15-minute segments (one hour testing time in all). The mathematics booklet contained 76 items and covered four content areas: Measurement, Geometry, Data Analysis/Statistics/Probability, and Algebra/Functions. The science test consisted of 72 items and covered four content areas: Life Sciences, Physical Sciences, Earth/Space Sciences, and the Nature of Science. Students completed either a mathematics or science test and were administered all items on the test.

Unlike IAEP2, the TIMSS test booklets contained both mathematics and science items. At grades 7 and 8, the mathematics test comprised 151 items and the science test 135 items. The TIMSS mathematics items covered six content areas: Fractions/Number Sense, Geometry, Algebra, Data Representations/Analysis/Probability, Measurement, and Proportionality. The science content areas were: Earth Science, Life Science, Physics, Chemistry, and Environmental Issues/Nature of Science. Items were rotated across eight test booklets and student performance was matrix-sampled using a modified Balanced-Incomplete-Block (BIB) spiraling design (Martin & Kelly, 1997). One and a half hours were allocated for the completion of each booklet. In both studies, performance on both tests was reported in the form of an average percentage correct score. In the case of TIMSS, an average scale score for each country was also reported. While scale scores were calculated for the IAEP2 study, they were not included in the published reports.

The Consistency of IAEP2 and TIMSS Science Results

In 1991, the average science performance of Irish 13-year-olds is significantly below the average performance of students in all but two of the 'common' countries (Portugal and the US) and also significantly below the international mean (Lapointe, Askew, & Mead, 1992). However, in 1995, the average performance of Irish students on the TIMSS test at grades 7 & 8 compares much more favorably with the 'common' countries and with the overall TIMSS means (Beaton et al, 1996b). This change of fortune is clearly evident in Table 1, in which countries are listed from highest achieving to lowest achieving, and are categorized according to whether their means were statistically significantly above, below, or did not differ from, the Irish mean.

Table 1
Science and Mathematics Means of Countries that Participated in IAEP2 and TIMSS
(Categorised in Terms of the Significance of Difference of Each Mean from the Irish Mean)[a, b]

IAEP2 13-year-olds			TIMSS Grade 7			TIMSS Grade 8		
Science								
	M	SE		M	SE		M	SE
Overall[c]	66.9			49.8	(0.1)		55.5	(0.1)
Kor	77.5	(0.5)	Kor	61.4	(0.4)	Kor	65.5	(0.3)
Swi	73.7	(0.9)	Slo	57.2	(0.5)	Slo	61.7	(0.5)
Hun	73.4	(0.5)	Hun	55.5	(0.6)	**Eng**	**61.3**	(0.6)
Slo	70.3	(0.5)	Eng	55.6	(0.6)	**Hun**	**60.7**	(0.6)
Can	68.8	(0.4)	**US**	**54.0**	(1.1)	**Can**	**58.7**	(0.5)
Eng	68.7	(1.2)	**Can**	**54.0**	(0.5)	**Ire**	**58.4**	(0.9)
Fra	68.6	(0.6)	**Ire**	**52.0**	(0.7)	**US**	**58.3**	(1.0)
Sco	67.9	(0.6)	**Swi**	**50.1**	(0.4)	**Swi**	**56.3**	(0.5)
Spa	67.5	(0.6)	Spa	49.3	(0.4)	Spa	55.6	(0.4)
US	**67.0**	(1.0)	Sco	48.2	(0.8)	**Sco**	**55.3**	(1.0)
Ire	**63.3**	(0.6)	Fra	46.1	(0.6)	Fra	53.7	(0.6)
Por	**62.6**	(0.8)	Por	41.3	(0.5)	Por	49.9	(0.6)
IAEP2 13-year-olds			**TIMSS Grade 7**			**TIMSS Grade 8**		
Mathematics								
	M	SE		M	SE		M	SE
Overall	58.3			49.3	(0.1)		55.1	(0.1)
Kor	73.4	(0.6)	Kor	67.0	(0.6)	Kor	71.7	(0.5)
Swi	70.8	(1.3)	**Hun**	**53.8**	(0.8)	**Swi**	**62.0**	(0.6)
Hun	68.4	(0.8)	**Swi**	**53.1**	(0.5)	**Hun**	**61.5**	(0.7)
Fra	**64.2**	(0.8)	**Ire**	**53.3**	(1.0)	**Fra**	**61.3**	(0.8)
Can	**62.0**	(0.6)	**Slo**	**52.5**	(0.7)	**Slo**	**61.2**	(0.7)
Eng	**60.6**	(2.2)	**Can**	**51.6**	(0.5)	**Ire**	**58.7**	(1.2)
Sco	**60.6**	(0.9)	**Fra**	**51.0**	(0.8)	**Can**	**58.7**	(0.5)
Ire	**60.5**	(0.9)	US	47.7	(1.2)	Eng	53.1	(0.7)
Slo	**57.1**	(0.8)	Eng	47.2	(0.9)	US	53.0	(1.1)
Spa	55.4	(0.8)	Sco	44.3	(0.9)	Sco	51.6	(1.3)
US	55.3	(1.0)	Spa	42.4	(0.6)	Spa	51.0	(0.5)
Por	48.3	(0.8)	Por	36.6	(0.6)	Por	42.9	(0.7)

[a] In TIMSS, overall scale scores rather than overall average percents correct were used to report the outcomes of statistical tests.

[b] Average performance in countries whose data appear in bolded type is not statistically significantly different from that in Ireland. Average performance in countries above the bolded entires is statistically significantly above that in Ireland. Average performance in countries below the bolded entries is statistically significantly below that in Ireland.

[c] The international averages in the table are for all participating countries and educational systems in each of the studies. The standard errors for the IAEP averages were not published.
Source. For IAEP2: Lapointe, Askew, & Mead (1992), Lapointe, Mead, & Askew (1992), ETS, (1992). For TIMSS: Beaton et al. (1996a; b), Center for the Study of Testing, Evaluation and Public Policy (n.d.).

Compared to their performance on the IAEP2 science assessment, four countries maintain their superiority over Ireland on the TIMSS assessment at grade 7 (Korea, Slovenia, Hungary, England). Two, having performed at a superior level on IAEP2, achieve at levels comparable to Ireland in TIMSS (Canada, Switzerland), while three that were superior on IAEP2 record a significantly poorer performance on TIMSS (France, Scotland, Spain). Comparisons between IAEP2 performance and performance at grade 8 on TIMSS reveal a somewhat similar pattern in which only two countries (Korea and Slovenia) maintain their superior position.

BEST COPY AVAILABLE

It is apparent that the relative performances of countries other than Ireland also change between IAEP2 and TIMSS (e.g., France and Switzerland). It could be argued that the same phenomenon occurs in mathematics (compare, for example, English and Scottish performances in the two surveys). However, changes in position are less frequent in mathematics, a finding that is reflected in the magnitude of the correlations between scores in the two assessments (Table 2).

Table 2
Correlations Between the Performances of Countries that Participated in Both IAEP2 and TIMSS (n=12)

	TIMSS Grade 7 Mean Scale Score	TIMSS Grade 7 Mean Percent Correct
Mathematics		
IAEP2 Mean Scale Score	.83	
IAEP2 Mean Percent Correct		.83
Science		
IAEP2 Mean Scale Score	.55	
IAEP2 Mean Percent Correct		.66

In considering the consistency of scores from one assessment to another, data on statistical significance from the published reports could have been used (as they were in Table 1). However, since our interest is in the extent to which the size of differences between pairs of country means changed across the assessments, we chose to use an effect-size index.

Effect Size Differences

The effect size is a measure of the magnitude in numerical terms of a difference of interest (in the present case, mean differences between countries) (Hair, Anderson, & Black, 1995; Wolf, 1986). The measure chosen for the present analysis is Cohen's *d* which is a measure of standardized differences between means, expressed in terms of standard deviation units (Cohen, 1977). The measure provides a scale-invariant estimate of the magnitude of an effect and involves dividing the value of the difference between two group means by the pooled standard deviation, using the formula,

$$d = (M_1 - M_2)/s_{pooled}$$ in which,

d is the effect size index for differences between means in standard units;
M_1 and M_2 are the sample means in original measurement units; and
s_{pooled} is the pooled standard deviation for both samples and is calculated as
$[(n_1 - 1)s_1 + (n_2 - 1)s_2]^{1/2} (n_1 + (n_2 - 2)^{-1/2}$

The effect size measure is now in the common metric of standard deviation units. Thus, an effect size of 0.3 indicates that one country scored 0.3 of a standard deviation higher (or lower) than the comparison country. Guidance for interpreting effect sizes is equivocal. It has been suggested that effect sizes around 0.2 are small, those around 0.5 are medium, and those around or above 0.8 are large (Cohen, 1977). However, the significance of an effect size will depend on the context in

which it is obtained (Durlak, 1995).

Table 3
Effect Sizes Observed in Science for IAEP2

	Can	Eng	Fra	Hun	Irl	Kor	Por	Sco	Slo	Spa	Swi	US
Can	.00	+.01	+.04	-.27	+.39	-.54	+.45	+.08	-.03	+.15	-.31	+.16
Eng	-.01	.00	+.03	-.27	+.34	-.53	+.41	+.06	-.04	+.12	-.28	+.15
Fra	-.04	-.03	.00	-.30	+.32	-.57	+.39	+.03	-.07	+.09	-.32	+.12
Hun	+.27	+.27	+.30	-.00	+.60	-.26	+.67	+.32	+.23	+.42	-.01	+.43
Ire	-.39	-.34	-.32	-.60	.00	-.89	+.07	+.28	-.39	-.26	-.65	-.21
Kor	+.54	+.53	+.57	+.26	+.89	.00	+.96	+.60	+.50	+.69	+.25	+.69
Por	-.45	-.41	-.39	-.67	-.07	-.96	.00	-.35	-.45	-.33	-.70	-.28
Sco	-.08	-.06	-.03	-.32	+.29	-.60	+.35	.00	-.10	+.06	-.36	+.09
Slo	+.03	+.04	+.07	-.23	+.39	-.50	+.45	+.10	.00	+.18	-.27	+.19
Spa	-.15	-.12	-.09	-.42	+.26	-.69	+.33	-.06	-.18	.00	-.46	+.03
Swi	+.31	+.28	+.32	+.01	+.66	-.25	+.70	+.36	+.27	+.46	.00	+.44
US	-.16	-.15	-.12	-.43	+.21	-.69	+.28	-.09	-.19	-.03	-.44	.00

Note: Reading across the row and comparing performance with country listed in heading: Positive effect sizes reflect higher average performance; negative effect sizes reflect lower average performance.

Table 4
Effect Sizes Observed in Science for TIMSS Lower Grade

	Can	Eng	Fra	Hun	Irl	Kor	Por	Sco	Slo	Spa	Swi	US
Can	.00	-.14	+.61	-.21	+.04	-.39	+.83	+.34	-.35	+.26	+.17	-.09
Eng	+.14	.00	+.72	-.06	+.17	-.24	+.89	+.44	-.18	+.39	+.28	+.04
Fra	-.61	-.72	.00	-.88	+.58	-1.01	+.31	-.23	-1.06	-.34	-.44	-.57
Hun	+.21	+.06	+.88	.00	+.25	-.19	+1.12	+.54	-.13	+.50	+.39	+.10
Ire	-.04	-.17	+.58	-.25	.00	-.44	+.86	+.29	-.39	+.22	+.13	-.12
Kor	+.39	+.24	+1.01	+.19	+.44	.00	+1.20	+.73	+.05	+.66	+.56	+.26
Por	-.83	-.89	-.31	-1.12	-.86	-1.20	.00	-.51	-1.39	-.63	-.75	-.77
Sco	-.34	-.44	+.23	-.54	-.29	-.73	+.51	.00	-.68	-.11	-.18	-.38
Slo	+.35	+.18	+1.06	+.13	+.39	-.05	+1.39	+.68	-.00	+.66	+.55	+.21
Spa	-.26	-.39	+.34	-.50	-.22	-.66	+.63	+.11	-.66	.00	-.09	-.30
Swi	-.17	-.28	+.44	-.39	-.13	-.56	+.75	+.18	-.55	+.09	.00	-.23
US	+.09	-.04	+.57	-.10	+.12	-.26	+.77	+.38	-.21	+.30	+.23	.00

Note: Reading across the row and comparing performance with country listed in heading: Positive effect sizes reflect higher average performance; Negative effect sizes reflect lower average performance.

The effect sizes associated with country differences in the IAEP2 and TIMSS surveys are contained in Tables 3 and 4 respectively and are based on the weighted *n*s, scale scores, and standard deviations (see Appendix A and B). Scale scores for IAEP2 were taken from the public use data file. Changes in effect sizes between pairs of means on the assessments are the absolute values of the difference between the effect size for the IAEP2 assessment and the effect size for TIMSS, i.e.,

$$d_{change} = |d_{IAEP2} - d_{TIMSS}|.$$

These absolute values are presented in Table 5.

Table 5
Absolute Value of the Differences Between the Effect Sizes Observed in Science for IAEP2 and TIMSS Lower Grade

	Can	Eng	Fra	Hun	Ire	Kor	Por	Sco	Slo	Spa	Swi	US
Can	.00	.15	.56	.06	.35	.15	.38	.26	.31	.11	.48	.25
Eng	.15	.00	.69	.21	.17	.29	.48	.38	.14	.27	.57	.11
Fra	.56	.69	.00	.58	.90	.44	.08	.25	.99	.43	.12	.69
Hun	.06	.21	.58	.00	.35	.08	.45	.22	.36	.08	.40	.33
Ire	.35	.17	.90	.34	.00	.46	.79	.58	.01	.48	.77	.08
Kor	.15	.29	.44	.08	.46	.00	.24	.12	.45	.02	.31	.43
Por	.38	.48	.08	.45	.79	.24	.00	.16	.93	.30	.05	.49
Sco	.26	.38	.25	.22	.58	.12	.16	.00	.57	.17	.18	.47
Slo	.31	.14	.99	.36	.00	.45	.93	.57	.00	.48	.82	.02
Spa	.11	.27	.43	.08	.48	.02	.30	.17	.48	.00	.37	.33
Swi	.48	.57	.12	.40	.78	.31	.05	.18	.82	.37	.00	.67
US	.25	.11	.69	.33	.08	.43	.49	.47	.02	.33	.67	.00

Note: Slight differences between the absolute values in this table and the values in Tables 3 and 4 on which they are based result from rounding error.

Reading across the columns or down the rows gives the effect size differences for a country compared to all other countries. For example, the difference between the effect sizes for Canada and England in the two assessments is 0.15 standard deviation units – a small difference reflecting the fact that the mean achievement in both countries is not significantly different in either assessment.

Most of the largest effect size differences are associated with France, Ireland, and Switzerland (Table 5). Large effect size differences are evident at the intersection of France and Ireland (0.90) and at the intersection of Ireland and Switzerland (0.77). This is a reflection of the fact that while Ireland's standing relative to these countries was poor in IAEP2, Ireland scored higher than these countries in TIMSS. The intersection of France and Switzerland shows a small effect size difference (0.12) and confirms that these countries maintained their position relative to each other on both occasions. However, effect sizes at the intersection of France and countries such as England (0.69), Hungary (0.58), Slovenia (0.99) and the US (0.69) are large. The Swiss change of fortune is clearly reflected in the effect size differences between it and England (0.57), Slovenia (0.82), and the US (0.67).

Moderate to large effect sizes are also associated with comparisons involving Portugal, Scotland, Slovenia, and the US. For example, the effect size difference at the intersection of Portugal and Slovenia is 0.93. In both assessments, Portugal scored significantly lower than Slovenia. However, the large value results from the fact that while the effect size was in the order of 0.45 in IAEP2, it increased to 1.39 in TIMSS. Indeed, most of the other large effect sizes associated with Portugal reflect that country's very poor performance in TIMSS. Other moderately large effect sizes worth noting are those at the intersections of Scotland and Slovenia (0.57), Scotland and the US (0.47), Korea and Slovenia (0.45), Slovenia and Spain (0.48), and Korea and the US (0.43). Other analyses, not reported here, show that the absolute value of differences between effect sizes observed for mathematics, though large in some cases, are generally much smaller than for science (O'Leary, 1999).

Conclusion

The dilemma that our findings give rise to for policy makers seems straightforward enough. Do the findings (for more countries at any rate) indicate a change in level of science achievement over time? And if not, which results are to be taken as a 'true' reflection of its nation's achievement? Careful consideration now needs to be given to the task of trying to explain why performance in the two assessments seems to be so different for some countries. At least five hypotheses can be suggested (see Beaton et al., 1990 for a description of efforts to disentangle the 1985/86 reading anomaly in the National Assessment of Educational Progress in the United States). These, each of which will be briefly considered, relate to population definitions, survey implementation, approaches to data analysis, the possibility of real gains or losses in the achievement of students in some countries during the period between the two surveys and measuring instrument issues.

Firstly, differences in population definitions might account for differences in the relative performance of students in IAEP2 and TIMSS science. In IAEP2 a sample of students who were 13 years old was tested. In TIMSS the students were in grades 7 and 8. While there is some overlap between these two populations, there are differences between them that need to be taken into account when comparing performance. For example, it is noteworthy that for TIMSS science the estimated median scale score for Irish 13-year olds (486) is lower than the mean scale score for Irish seventh graders (495) and that the median score for Swiss 13-year-olds is exactly equivalent to the Irish mean at the seventh grade (see, Beaton et al., 1996b, pp. 26 and 37).

(A median scale score rather than a mean scale score was calculated for 13-year-olds in TIMSS due to the fact that students were sampled by grade and not by age. Not all 13- year-olds were in the grades sampled and, as a consequence, an estimate of the median was thought to be more reliable.) Ramseier (1997, personal communication) claims that a large part of the change in Swiss performance between IAEP2 and TIMSS can be explained by the fact that 44% of Swiss 13-year olds are in grade 8. He argues that comparing Swiss grade 8 performance to the performance of grade 7 students in Ireland (where most 13-year olds are) provides evidence that Swiss IAEP2 and TIMSS performances may not be all that different. However, taking the sampling variability of both medians into account, it must still be argued that, as the scores for both sets of 13-year olds suggest, Switzerland did not perform significantly better than Ireland in TIMSS. (The standard errors of the Irish and Swiss medians were 3.1 and 2.2 respectively).

Secondly, populations with exclusions and low participation rates in some countries may also account for some of the differences in outcomes across the two studies. Exclusions were caused by countries modifying the internationally agreed definition of the population to be tested. Low participation rates were caused by having combined school and student participation rates below an agreed cut-off mark (70% in IAEP2 and 75% in TIMSS). A few examples will suffice to illustrate the point. In IAEP2, Spain excluded students in Cataluna but included them in TIMSS. In IAEP2, Switzerland tested in only 15 of the 26 Cantons whereas 22 Cantons were involved in TIMSS. In IAEP2, England had a final participation rate of only 48% while in TIMSS it was closer to 80% after replacement. Indeed, a particularly vexing question in international assessments (or any large- scale assessment for that matter) is the extent to which exclusions and participation rates affect overall performance (see Linn & Baker, 1995).

Thirdly, differences in approaches to data analysis may account for differences in the relative performance of students in IAEP2 and TIMMS science. Both IAEP2 and TIMSS use complex procedures for estimating average percentage correct and average proficiency scale scores. Technical reports that were published in conjunction with the assessments indicate that the technologies differed for the two surveys. For example, approaches to handling missing data when calculating average percents for items differed across the two studies (not reached items were treated as not administered in IAEP2 while they were treated as incorrect in TIMSS). Moreover, in IAEP2, average scale scores were calculated using a 3-parameter Item Response Theory model, while in TIMSS a modified Rasch model was used (see Adams, Wilson & Wang, 1997). The fact that TIMSS items were matrix sampled

(using a BIB design) and that a plausible values technology was used makes it a very different kind of survey to the more straightforward IAEP2.

Fourthly, between 1991 and 1995, levels of science achievement for students around 13 years of age may have increased or decreased, accounting for differences in the relative performance of students in IAEP2 and TIMSS science. We do not, however, have any evidence to support the view that substantial change occurred in the achievement of Irish 13- year old students during the four years between IAEP2 and TIMSS. Comparing outcomes from the two assessments, all we can say is that, in a normative sense, Irish performance in TIMSS improved. Comparison with the Swiss is important here. Ramseier (1997, personal communication) suggests that age, instruction time and curriculum issues affected Swiss performance in TIMSS. Was Ireland's favorable comparison with the Swiss in TIMSS merely an artifact of poor Swiss performance? Of course Ireland's performance relative to more than one country improved and this suggests that achievement in a real sense may have improved. But we cannot say for sure. While the time-span between the two assessments is probably not long enough to allow for the kind of gains that might help explain the improved relative performance in TIMSS, the matter of how performance in IAEP2 can be equated with performance in TIMSS in an absolute sense is a substantial matter and one that is of the utmost importance to an accurate interpretation of national performance in the two surveys.

Fifthly, differences in measuring instruments might account for differences in the relative program of students in IAEP2 and TIMSS science. As noted above, there were differences in the content areas of the IAEP2 and TIMSS tests. TIMSS had a section entitled Environmental Issues which IAEP2 did not. There were also differences in the proportion of items assigned to common content areas. For example, while 17% of the IAEP2 items were devoted to the Nature of Science, the figure for TIMSS was 6%. In addition, more of the TIMSS test (5%) was devoted to Physics. Hence, differences in performance may be a function of differences in the nature of the achievement that was assessed. However, an interesting issue arising in this context is worth raising here. The fact is that while the instruments measuring mathematics achievement also differed in content coverage, the mathematics performance of countries across the two studies was more consistent. The question arises: In international studies do particular factors impinge much more strongly on science achievement than mathematics achievement?

Finally, and as an extension of the last point, what seems reasonably clear is that underlying the reporting of results of international studies in the popular media and in many reports emanating from government ministries is an assumption that 'science,' 'mathematics,' 'reading' and the like are clearly understood. But is this the case? Can we say that there is real consensus about the nature of these domains and the underlying psychological constructs implied by "achievement" in these subjects? Or could it be that at the international level an understanding of what constitutes achievement in mathematics, for example, is at a more advanced level than the understanding of what constitutes science achievement? It is noteworthy that some support for this hypothesis is contained in our finding that country rank orderings were more stable in mathematics than in science across two distinct international assessments. Moreover, in the United States the analysis by Hamilton and her colleagues (1995) of a large scale national test (NELS:88) provides further food for thought in suggesting that "achievement patterns in science were much more heterogeneous than in math" and that "[i]n science, a far greater number of factors was required to account for student performance differences" (p. 577). Such findings raise critical questions about the science tests used in international comparative studies.

Note

The poor performance of Irish students in science was also a feature of the First International Assessment of Educational Progress in Mathematics and Science (IAEP1) test in 1988 (Lapointe, Meade, & Phillips, 1989).

References

Adams, R. J., Wilson, M., & Wang, W-C. (1997) The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement,* 21, pp. 1-23.

Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996a) *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study.* Chestnut Hill. MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Smith, T. A., & Kelly, D. L (1996b) *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study.* Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Beaton, A. E., Zwick, R., Yamamoto, K., Mislevy, R. J., Johnson, E. G., & Rust, K. F. (1990) *Disentangling the NAEP 1985-86 Reading Anomaly.* Princeton, NJ: Educational Testing Service.

Brown, M. (1996) FIMS and SIMS: the first two IEA international mathematics surveys, *Assessment in Education: Principles, Policy & Practice*, 3, pp. 193- 212.

Brown, M. (1998) The tyranny of the international horse race. In R. Slee & G. Weiner with S. Tomlinson (Eds) *School Effectiveness for Whom? Challenges to the School Effectiveness and School Improvement Movements*, pp. 33- 47 (London, Falmer Press).

Cohen, P. (1977) *Statistical Power Analysis for the Behavioural Sciences.* New York: Academic Press.

Durlak, J. A. (1995) Understanding meta analysis. In L. G. G. P. R. Yarnold (Eds.), *Reading and Understanding Multivariate Statistics*, pp. 319-352 Washington, D.C.: American Psychological Association.

ETS (1992) *IAEP technical report: Volume 1*. Princeton, NJ: Educational Testing Service.

Goldstein, H. (1996) Introduction. *Assessment in Education: Principles, Policy and Practice*, 3, pp. 125- 128.

Hair, J. F., Anderson, R. E., & Black, R. L. T. C. (1995) *Multivariate Data Analysis* (4th ed) Englewood Cliffs, NJ: Prentice Hall.

Hamilton, L. S., Nussbaum, E. M, Kupermintz, H., Kerkhoven, J. I. M., & Snow, R. S. (1995) Enhancing the validity and usefulness of large-scale educational assessments: NELS:88 science achievement. *American Educational Research Journal*, 32, pp. 555-581.

Husén, T. (1967) *International Study of Achievement in Mathematics*, 2 vols. Stockholm: Almquist & Wiksell.

Husén, T., & Postlethwaite, T. N. (1995) A brief history of the International Association for the Evaluation of Educational Achievement (IEA). *Assessment in Education: Principles, Policy and Practice*, 3, pp. 129- 141.

Kellaghan, T. (1996) IEA studies and educational policy. *Assessment in Education: Principles, Policy and Practice*, 3, pp. 143-160.

Kellaghan, T., & Grisay, A. (1995) International comparisons of student achievement: Problems and prospects. In *Measuring What Students Learn*, pp. 41-61 Paris: Organization for Economic Co-Operation and Development.

Lambin, R. (1995) What can planners expect from international quantitative studies?, in W. Bos & R. H. Lehman (Eds) *Reflections on Educational Achievement. Papers in honour of T. Neville Postlewaithe*, pp. 169-182 New York: Waxman.

Lapointe, A. E., Askew, J. M., & Mead, N. A. (1992) *Learning Science.* Princeton, NJ: Educational Testing Service.

Lapointe, A. E., Mead, N. A. & Phillips, G. W. (1989) *A World of Differences: An International Assessment of Mathematics and Science.* Princeton, NJ: Educational Testing Service.

Linn, R. L., & Baker, E. L. (1995). What do international assessments imply for world-class standards? Implications of international assessments. *Educational Evaluation and Policy Analysis,* 17, *4*, pp. 405-418.

Martin, J. O., Hickey, B. L., & Murchan, D. P. (1992) The second international assessment of educational progress: mathematics and science findings in Ireland. *Irish Journal of Education*, 26, pp. 3-146.

Martin, M. O., & Kelly, D. L. (1996) *Third International Mathematics and Science Study. Technical report. Volume 1: Design and Development* (Chestnut Hill, MA, Center for the Study of Testing, Evaluation, and Educational Policy, Boston College).

Murphy, P. (1996) The IEA assessment of science achievement, *Assessment in Education: Principles, Policy & Practice*, 3, pp. 213-232.

Nuttall, D. (1994) Choosing indicators, In *Making Education Count. Developing and using international indicators*, pp. 79-96 (Paris, Organisation for Economic Co-operation and Development)

O'Leary, M. (1999) *The validity and consistency of findings from international comparative studies of student achievement: A comparison of outcomes from IAEP2 and TIMSS.* (Unpublished doctoral dissertation, Boston College, Massachusetts).

Plomp, T. (1992) Conceptualizing a comparative educational research framework. *Prospects*, 22, pp. 278-288.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991) *Statistical Principles in Experimental Design.* New York: McGraw Hill.

Wolf, F. M. (1986) *Meta Analysis: Quantitative Methods for Research Synthesis.* Newbury Park, CA: SAGE.

About the Authors

Michael O'Leary
St. Patrick's College
Drumcondra, Dublin 9, Ireland
Telephone +353-1-8842000
Fax +353-1-8376197
Email: Michael.OLeary@spd.ie
Michael O'Leary is a member of the Education Department at St.Patrick's College, Dublin, Ireland. He holds a Ph.D. from Boston College in the area of educational research, measurement and evaluation. He has served as Ireland's representative on the Board of Participating Countries for the OECD's Programme for International Student Assessment (PISA) project.

Thomas Kellaghan
Thomas Kellaghan is Director of the Educational Research Centre at St. Patrick's

College, Dublin. He is a member of Academia Europaea and a Fellow of the International Academy of Education. He is currently serving as President of the International Association for Educational Assessment.

George Madaus
George Madaus is Boisi Professor of Education and Public Policy at Boston College. He has served as Director of the National Commission on Testing and Public Policy, and as Vice President of AERA Division D, and as President of NCME. He is a a member of the National Academy of Education and a Senior Research Associate of the National Board on Educational Testing and Public Policy.

Albert Beaton
Albert Beaton is a professor in Boston College's Graduate School of Education. He was international study director of the Third International Mathematics and Science Study (TIMSS) and is a former director of design, research, and data analysis for the National Assessment of Educational Progress (NAEP). He is a member of the International Academy of Education and an honorary member of the International Association for the Evaluation of Educational Achievement (IEA).

Appendix A
Average Science Scale Scores for 13-year-olds in IAEP2

	n	Weighted n	Scale Score	se	sd
Can	4980	182312	534	1.5	61
Eng	929	504590	533	3.9	71
Fra	1787	672764	531	2.5	69
Hun	1623	149647	552	2.3	72
Ire	1657	63791	509	2.5	72
Kor	1635	671867	570	2.3	68
Por	1520	149228	504	3.8	72
Sco	1584	55398	529	2.8	69
Slo	1598	26640	536	2.2	65
Spa	1609	440322	525	2.3	61
Swi	3653	52726	553	3.4	63
US	1404	3028386	523	4.4	68

Source: International Assessment of Educational Progress (IAEP2), 1991-1992.

Appendix B
Average Science Scale Scores at Grade 7 in TIMSS

	n	Weighted n	Scale Score	se	sd
Can	8219	377731	499	2.3	90
Eng	1803	465457	512	3.5	101
Fra	3016	860657	451	2.6	74
Hun	3066	118727	518	3.2	91
Ire	3127	68477	495	3.5	91
Kor	2907	798409	535	2.1	92
Por	3362	146882	428	2.1	71
Sco	2913	62917	468	3.8	94
Slo	3600	28049	530	2.4	86
Spa	3741	549032	477	2.1	80
Swi	4085	66681	484	2.5	82
US	3886	3156847	508	5.5	105

Source: IEA's Third International Mathematics and Science Study (TIMSS), 1994-1995.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 470 times since August 27, 2000

Education Policy Analysis Archives

Volume 8 Number 44 August 27, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES.**
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Information Needs in the 21st Century: Will ERIC Be Ready?

Lawrence M. Rudner
ERIC Clearinghouse on Assessment and Evaluation
University of Maryland, College Park

Abstract
Ubiquitous for 35 years, the Educational Resources Information Center (ERIC) is known for its database and recently for its range of web-based information services. I contend that federal policy with regard to ERIC must change and that ERIC will need massive restructuring in order to continue to meet the information needs of the education community. Five arguments are presented and justified: 1) ERIC is the most widely known and used educational resource of the US Department of Education, 2) senior OERI and Department of Education officials have consistently undervalued, neglected, and underfunded the project, 3) ERIC's success is due largely to information analysis and dissemination activities beyond ERIC's contracted scope, 4) information needs have changed dramatically in the past few years and ERIC cannot keep up with the demands given its current resources, and 5) the ERIC database itself needs to be examined and probably redesigned.

Introduction

The Educational Resources Information Center (ERIC) has been the most visible source for education information since its inception in 1966. As a system of 16 clearinghouses and 3 support contractors, ERIC collects, abstracts, and indexes education materials for the ERIC database; responds to requests for information in subject specific areas; and produces special print and electronic publications on current research, programs, and practices. As we enter into the 21st century and the

Information Age, the question to ask is: "Will ERIC be ready?" Taking a hard look at what ERIC has been and what ERIC is today relative to user information needs, I conclude that ERIC will need massive restructuring in order to continue to meet the information needs of the education community.

I base my conclusion on five basic arguments

1. ERIC is the most widely known and used educational resource of the US Department of Education.

2. While ERIC staff, including Office of Educational Research and Improvement monitors, have long appreciated ERIC, senior OERI and Department of Education officials have consistently undervalued, neglected, and underfunded the project.

3. ERIC's success is due largely to information analysis and dissemination activities that go beyond ERIC's contracted scope.

4. Information needs have changed dramatically in the past few years and ERIC cannot keep up with the demands given its current resources.

5. The ERIC database itself needs to be examined and probably redesigned.

In this article, I justify these arguments. In my summary, I look at the federal role in education and conclude that unless ERIC is restructured, the U.S. Department of Education will fragment the nation's already frail educational information infrastructure. Educational research and practice will lose because neither will be able to readily build on past findings.

ERIC is the most widely known and used educational resource of the U.S. Department of Education

In its early years, ERIC was primarily an archive of the education literature. Its main activity was the development of its databases, *Resources in Education (RIE)* and *Current Index to Journals in Education (CIJE)*. Its primary users were researchers; the primary mode of access was through expert intermediaries - typically, reference librarians.

While these two databases continue to be a major cornerstone for all clearinghouses, the rapid advancements of information technology have prompted ERIC to evolve into a much more powerful and useful resource. With the explosive growth of the Internet and CD-ROM products, ERIC as a system is now widely recognized as the central source for educational information.

ERIC's user base has also changed. The majority of ERIC users today are teachers and other education practitioners. The mode has also changed—most users access ERIC themselves. And the nature of ERIC's work has changed—we are now more heavily involved in providing direct user services for many different audiences. All clearinghouses are heavily involved in providing a strong value-added service, i.e., information adapted to local need. Today, ERIC Clearinghouses

- prepare syntheses on topics within their scopes,
- provide easy access to quality material,
- respond to an ever-growing number of user inquiries, and
- serve as centers for scope-related activities.

ERIC has always been the leader in providing useful information to teachers and other educators.

- In 1972, when Lockheed established the DIALOG on-line retrieval system,

ERIC was its first file. ERIC continues to be one of its most frequently searched databases, and it retains its position as the first file in the system.

- In the mid 1970's, ERIC became one of the first databases on CD-ROM.
- In the early 1990's, the ERIC Digest File became one of the most popular items on the Internet, in any field.
- In 1996, ERIC's new Internet question answering service was recognized for innovation and excellence in use of the "Information Highway."
- In 1997, ERIC became the first to offer a thesaurus as a front end for searching its database on the Internet.
- In 2000, ERIC became one of the largest full-text repositories on the Internet.

These are major firsts for both information science and education. Each of these innovations and accomplishments enhanced the usefulness and availability of information for ERIC's end-users, i.e. teachers and practitioners as well as researchers and policymakers. That these results are appreciated is readily evident:

- ERIC received nearly 180,000 letters and toll-free telephone inquiries in 1998.
- The ERIC Clearinghouses responded to over 100,000 user questions in 1999.
- The ERIC database is the third most frequently used database in any field (*Computers in Libraries*, February 1995).
- Nearly 1000 organizations buy the expensive ERIC microfiche collection.
- The last time the topic was investigated, ERIC was the most widely known OERI program (Stalford and Stern, 1990).
- More than 600 organizations have formal partnerships with ERIC.
- The ERIC Document Reproduction Service (EDRS) now fills individual orders for more than 35,000 copies of documents annually.
- ERIC Clearinghouses maintain more than 80 electronic discussion groups serving more than 37,000 education policymakers, administrators, teachers, parents, and library/media specialists.
- ERIC web sites are heavily used: In June 2000, the ERIC Clearinghouse on Assessment and Evaluation web site received 80,000 users per week; the ERIC Clearinghouse on Reading, English, and Communication received 140,000 users (users, not hits). Alexa.com gathers data on page-views and provides a popularity ranking with a ranking of 1 corresponding to the most popular site on the Internet. On August 10, 2000, the mean ranking of ERIC web sites was 128,000. In contrast, the mean rank of OERI's laboratories was 236,000 and the mean rank for OERI Centers was 296,000 (see Table 1). The ranks were comparable in March 2000 during the school year.

Table 1
Popularity Rank of OERI ERIC, Regional Laboratories, and Research Center websites as rated by Alexa 8/10/2000

ERIC Clearinghouses		OERI R&D Laboratories		OERI Centers	
Reading, English, & Communication	1,891	Northwest Regional Education Laboratory	41,021	Center for the Study of Teaching & Policy	18,886
* Information & Technology	5,630	Mid-continent Regional Education Laboratory	41,620	National Center for Early Development & Learning	52,336
Assessment & Evaluation	9,512	North Central Regional Edueation Laboratory	42,519	Center for Improvement of Early Reading Achievement	171,770
Urban Education	58,764	Southwest Educational Development Laboratory	82,744	National Center for the Study of Adult Learning & Literacy	200,195
Social Studies/Social Science Education	67,902	WestEd	102,588	Center for Research on the Education of Students Placed At-Risk	210,687
Disabilities & Gifted Education	97,825	* Appalachia Regional Education Laboratory	167,103	National Center for Improving Student Learning & Achievement in Mathematics & Science	218,904
Community Colleges	99,033	Southeastern Regional Vision for Educators	220,079	Center for Research on Evaluation, Standards, & Student Testing	357,558
* Elementary & Early Childhood Education	157,034	Northeast & Island Regional Education Laboratory	411,025	National Research Center on the Gifted & Talented	402,967
Teaching & Teacher Education	181,268	Pacific Region Education Laboratory	1,020,475	National Research & Development Center on English Learning & Achievement	545,177
Educational Management	209,587			Center for Research on Education, Diversity, & Excellence	782,396
Higher Education	210,740				
Adult, Career, & Vocational Education	271,478				
Science, Mathematics, & Environmental Education	298,924				
mean	128,430	mean	236,575	mean	296,088
median	99,033	median	102,588	median	214,765

Rankings by Alexa are based on page visits by Alexa users. With millions of users, Alexa claims to have the largest, most geographically and demographically diverse sample o overall web usage currently available. Organizations that do not have their own domain name are not ranked and are not shown in the table.

* ERIC/Early Childhood and ERIC/Information & Technology operate multiple websites with multiple domain names. Shown are just the rankings for the main clearinghouse website. More than half of AEL page visits are from the ERIC/Rural Clearinghouse.

Department of Education officials have consistently undervalued, neglected, and underfunded the ERIC program.

This is a bold statement. It reflects 19 years of personal observation. I preface

my remarks with a recognition that senior Department of Education officials arrive with large agendas and limited time. ERIC is a program that appears to be working and not causing problems. Hence it is a program that doesn't require much attention. However, ERIC has suffered both from efforts to politicize it and from benign neglect.

One of the first OERI Assistant Secretaries formed an ERIC Recompetition Design Panel involving government and non-government representatives. Inserting politics rather than informed judgment, that Assistant Secretary then claimed that the panel advocated changes that were part of his agenda and that had nothing to do with the deliberations of the Design Panel. Historically, Assistant Secretaries and other senior U.S. Department of Education officers had so many misconceptions that the Director of the ERIC program authored a paper entitled "Myths and Realities about ERIC" (Stonehill, 1995). ERIC has received few invitations to participate in various OERI panels and advisory meetings. Until recently, the ERIC program office within OERI has been severely understaffed.

For the past 10 years, the federal government has spent approximately $9 million yearly on ERIC. The funding goes to pay for the clearinghouses, a central processing facility, GPO printing of *Resources in Education*, and ACCESS ERIC which serves as a contact point for the ERIC system and produces many reports previously produced by the central processing facility. Most users think we have a much bigger budget. During ERIC's lifetime, federal support for education nearly quadrupled (Hoffman, 1995). In constant dollars, funding for ERIC, however, is now less than one-half what it was 20 years ago. In the last ERIC recompetition, Clearinghouses were each level funded while required to provide support for AskERIC and to devote $30,000 toward web development.

Notably absent are funds for research and development. Until this year, the US Department of Education's Office for Educational Research and Improvement has spent zero dollars for study and systematic evaluation of its most visible project. In FY 2000, four papers were commissioned at $10,000 each. When one considers that ERIC has been level funded for 20 years and that virtually no money has been allocated for research and evaluation in support of the ERIC project, ERIC's accomplishments appear even more amazing. Credit goes to the ERIC Directors for being in tune with their content areas and to the ERIC program office for gently guiding ERIC without the benefit of hard data. However, the assumptions that have guided ERIC so well in the past, no longer hold. Information needs have changed dramatically and, more than ever, the ERIC program office needs to be guided by data rather than by intuition and to have the benefit of adequate resources to allocate.

ERIC has always taken pride in its ability to leverage resources. The ERIC Document Reproduction Service, which prepares microfiche of ERIC documents and distributes paper and electronic copies on demand, is a no-cost-to-the-government contract. It is paid for by standing orders for ERIC microfiche, fees collected for on-demand papers and electronic copies, and more recently subscriptions to the on-line, on-demand file. Central processing and quality control for the *Current Index to Journals in Education* was handled by Oryx Press at no charge to the government in exchange for the right to print *CIJE*. The private sector disseminated the ERIC database by mounting it as part of electronic information services (e.g., Dialog, BRS) or CD-ROM. Again these activities occur at no cost to the government.

Consistent with this minimal funding level, the scope of work for the individual clearinghouses has changed little over the past 30 years. Clearinghouses are charged with

- Acquiring documents
- Selecting documents for the ERIC database
- Preparing citations (about 1500-3000 per clearinghouse each year)
- Preparing Digests (about 10 per clearinghouse each year)
- Preparing major publications (about 2 books per clearinghouse each year)
- Giving workshops (about 2 per clearinghouse each year)
- Responding to user quest ons

The Request for Proposals used to compete the ERIC Clearinghouses has not changed significantly in the past 20 years. In fact, the scopes of work for the individual clearinghouses have not changed. In the 1970s, career and adult education were hot topics. Approximately 12% of the documents put into the ERIC database during that time were put in by the ERIC Clearinghouse on Adult, Career, and Vocational Education. This was more than twice the average of the other clearinghouses. Despite today's interest in bilingual education, assessment, higher education, and reform, the ERIC Clearinghouse on Adult, Career, and Vocational Education continues to be contractually obligated to supply some 12% of the ERIC documents while the clearinghouses responsible for these other topics contribute at the same levels they did 25 years ago—an average of approximately 6.0% (See Table 2). The activities of the ERIC clearinghouses should be guided by the ebb and flow of contemporary issues, contributions to knowledge, and user demand. It should not be basically static for 30 years.

Table 2
Distribution of RIE entries by Clearinghouse over time

	1976-1980	1990-1998	p-ratio
Elem Ed	4.2%	7.3%	1.73
Reading	7.2%	9.5%	1.32
Foreign Lang	4.5%	5.8%	1.28
Test Measure	5.0%	6.0%	1.20
Cmmnity Col	4.1%	4.7%	1.16
Disab/Gifted	6.2%	6.5%	1.05
Higher Ed	7.2%	7.5%	1.04
Inform Reso	7.1%	7.2%	1.02
Social Stud	4.9%	4.9%	1.01
Teacher Ed	5.4%	5.3%	0.97
Educ Manage	6.9%	6.7%	0.96
Career/Adult Ed	12.4%	11.4%	0.92
Counsel Guid	5.5%	4.7%	0.86
Rural Sch	4.1%	3.4%	0.84
Urban Sch	5.1%	4.2%	0.82
Science Math	5.9%	4.6%	0.79

p-ratios between .8 and 1.25 indicate that the percentages are practically equivalent.

A final example of ERIC's apparent failure to be appreciated within the Department of Education has to do with the creation of an Internet presence. When it became clear that educators at all levels were expecting to see Federally produced documents on the Internet, OERI provided supplemental funding to its Regional Labs to post their materials. The Labs responded with wonderful web pages, great collections of useful material. The ERIC Clearinghouses did not get any of this supplemental funding. ERIC's web presence is mostly the result of dedicated professio ials staying up late at night. The irony is that the Labs and Centers receive a great deal of funding to disseminate their own research, yet, as shown in Table 1, ERIC web sites have been much more effective. As the *national education dissemination system* (Mathtech, 1998a), ERIC is responsible for disseminating all quality material related to education and, even without sufficient funds, has been far more successful in serving the education community. I argue later that ERIC cannot maintain that level of service any longer.

Part of the problem stems from the nature of the program. ERIC is best known for its archiving of educational materials. ERIC gathers the literature and prepares the microfiche. From one point of view, ERIC is a fairly uninteresting project. It doesn't provide research breakthroughs. It does not generate headlines. It does not provide political mileage. It is not known outside of education and information science. Further, it appears to do its job adequately at the current funding level.

What senior Department of Education officials apparently have not appreciated is that, to be a quality archive, ERIC had to be a quality information center. ERIC has established formal relationships with every major organization that produces and consumes educational resources and information. To build these relationships, ERIC has to be an appreciated provider of information services.

ERIC's success is due largely to many marginal activities beyond ERIC's contracted scope

The success of ERIC is clearly not due solely to its efforts to gather papers and build a database. Rather, ERIC's success is due, to a great extent, to its value-added services. ERIC excels at identifying what will be helpful to its clients, identifying what is relevant and of high quality, and organizing and presenting information. In other words, ERIC is successful because it blends information science with subject matter expertise.

Some ERIC activities that are beyond the basic scope of clearinghouse work are:

- Mounting and maintaining the ERIC database on the web
- Most responses to Frequently Asked Questions
- Pathfinders
- Newsletters
- Journals (print and electronic)
- Newsletter and journal columns
- Workshops (beyond the first 2 each year)
- All printing activities
- All research activities
- Bookstores
- Major publications and books (beyond the first 2 each year)
- Development of lesson plans
- Compilations of reference materials
- Writing state-of-the-art search software for the web
- Test Locator
- Most web activities beyond simply establishing an Internet presence

The magnitude of these out-of-contract activities is evident in the wide range of on-line services offered at ERIC Clearinghouse websites, especially the more popular ERIC websites—those of the Reading, Information Resources, Assessment, Social Studies, Urban, and Disabilities/Gifted Clearinghouses. These are massive websites with many special features. However, they are marginal relative to what could be accomplished with a concerted, well-planned, and well-supported effort.

Lynch (2000) points out that ERIC needs to be concerned with database services in addition to database building. The Clearinghouses undertake these activities because this is what is necessary to be a viable clearinghouse. The time to create these products comes as volunteer time, either contributed by individuals or by their host institutions. Several ERIC Clearinghouses actually view the ERIC contract as a franchise license (Colker, 2000) and put a great deal of effort into selling and making money from books with the ERIC label. They then use this money to support the necessary Clearinghouse efforts not adequately funded by the government. Senior Department staff appear to be oblivious to these activities. They are paying primarily for the creation of the database; to them, everything else appears to be viewed as tangential. The Directors, however, view these activities as critical to clearinghouse success.

Information needs have changed dramatically in the past few years

For thirty-five years, the ERIC database has been built around well-established information science principles. Abstracts are developed following a set of standards. Citations draw upon authority lists so publication types, journals, and organizations are always presented the same way. The *ERIC Thesaurus* is used to identify

appropriate major and minor descriptors. The ERIC procedures manual takes more than a foot of shelf space. The quality of the ERIC database in terms of its structure is well appreciated in the information science community.

About 10 years ago, most ERIC searching was conducted by expert intermediaries. Reference librarians familiar with the ERIC database and trained in information retrieval would conduct searches rather than the end user. Once information needs were clearly identified, the intermediary would often present a highly relevant set of references. In my experience, I usually received 30 to 100 citations that were of potential interest. I would then spend hours in the library looking up and obtaining appropriate articles. The process would take weeks.

That type of searching has changed. Today the end user conducts his or her own search. When reference services are provided, the end user is often given 10 to 15 potentially relevant citations. End users today would like to obtain the most current information and they want it immediately. ERIC has responded by now offering the full-text of RIE documents since 1994, on-demand (For more information read about the E*subscribe program at www.edrs.com). Efforts are underway to make ERIC more timely.

To underscore that information needs have changed, let me ask a set of questions.

Which would you prefer to search?

a. National Academy of Science full-text of their books on-line
b. OCLC First Search of full-text journals
c. ERIC—Abstracts only

Twenty years ago, there were few options. Five years ago, ERIC was still basically the only education database. University Microfilms International (UMI) provided access to most of the journal articles in ERIC. The ERIC Document Reproduction Service provided access to the documents in RIE. Today, there are multiple education databases. For most people, the first preference will be high-quality materials they can get immediately. OCLC, EBSCOHost, JSTOR, CatchWord, the American Psychological Association and others are creating fee-based databases linked to the full text of peer-reviewed articles. ERIC's CIJE database has no such set of links, and UMI no longer provides reprint services. However, documents in ERIC's RIE database that were prepared in 1994 and later are now available on-demand, on-line. Should ERIC continue to abstract journal articles if it can't make them readily available?

Which would you prefer?

a. Packages with an Introduction to an issue and carefully selected full-text resources
b. An annotated bibliography
c. Search for yourself

Obtaining an answer to an education question is often not a trivial task. The literature is full of high- and low-quality articles; it is often difficult to identify potentially relevant articles, yet alone key articles. Ten years ago, there were few information analysis packages, and those that existed were often difficult to find. A lengthy annotated bibliography was considered a great starting tool. Today, there is a growing number of expertly prepared responses to Frequently Asked Questions. These make excellent starting points when one is interested in search a topic. Today, any FAQ is a blessing. In five years, however, the demand will be for *quality* FAQs. In a watch-dog role, the researchers in the content area will want to be sure novices are led to the best resources. Novices will want the best resources. Quality FAQs, with expert introductions to each topic's special problems and key references identified, require reference librarians working in conjunction with subject experts, as well as peer review and periodic updating. Today's ERIC can develop some FAQs, but not enough, not at the quality ERIC is capable of, and not with the ongoing maintenance FAQs require.

You need to make a policy decision, which do you prefer?

a. Carefully edited briefing papers presenting all sides of an issue
b. A selected collection of abstracts that summarize papers
c. Large collection of abstracts that summarize papers
d. Short abstracts that indicate without summarizing.

This question illustrates several points. First, a search of the ERIC database may be the end product desired of researchers, but it is generally a long way from the information desired by policymakers. Researchers may be willing to wade through indicative abstracts. Unless the policymaker has the luxury of time and is a researcher, the policy maker would prefer informative abstracts that summarize a paper. Ten years ago, the policymaker would have been happy with a large collection of informative abstracts, or better yet, a carefully selected collection.

Today, when information is required, the need is for greater depth and for immediate answers or at least viewpoints. ERIC's Digest Series fill that role nicely. Some 80,000 digests are distributed each month by www.ed.gov and cricac.net. But, will Digests be adequate, yet alone optimal, five years from now? I don't think so. The clearinghouses are told to budget approximately $1,200 for each Digest title. This amount does not provide the resources for an analysis of policy decisions, for the commissioning of papers, or even for assuring that the Digests are of the highest possible quality. While the education community has been very supportive of the ERIC Digest series and most expert authors are willing to volunteer to write Digests, something that is designed to introduce topics and possibly help guide decision making, should not be funded at the lowest possible level.

Which do you prefer to help you search for resources?

a. An expert in your field who is also an expert reference librarian
b. Expert librarian to search for you
c. A graduate student to search for you
d. Search for yourself

Ten years ago, one often used an expert librarian to help locate resources. There was often some tension as the expert librarian often did not have the subject-matter expertise. With the growth of on-line services, such as Dialog and the Internet, many have searched for themselves and have become frustrated (Rudner, 2000). The Clearinghouses now provide on-line reference services in response to those needs. In theory, we have subject-matter experts within the ERIC system and they respond with a set of relevant ERIC and Internet resources. In many ways, this has been a major success. Most patrons have been delighted with the service. However, ERIC cannot provide reference services as it does for the next five years. The clearinghouses are told to budget approximately $10.00 to respond to questions and it typically takes 30 to 45 minutes to provide a response. At this rate, most questions are answered by junior staff and graduate students. At that funding level, we cannot provide the quality and systematic evaluation that we would like and patrons should receive. The problem will get worse as the number of questions are increasing rapidly each year and the current ERIC contracts only allows for minor increments.

You are a researcher or practitioner, which do you prefer?

a. Search a carefully constructed pathfinder of the best resources
b. Search the entire Internet by yourself

Of course, ten years ago, the Internet was not an option. Perhaps last year, many were content to search the Internet themselves. But the Internet has become massive and overwhelming. Using the major search engines often yields many irrelevant links. Typically, the user enters a word or two and the engines provide a crude ranking and relevancy match based on all the text appearing on each web page. Improvements in this area will be marginal at best. An alternative is a carefully constructed pathfinder that identifies, organizes and annotates resources within a

given field. The Argus Corporation (www.clearinghouse.net) maintains an impressive list of such pathfinders. Many ERIC Clearinghouses have developed such tools and they are well-received. But, pathfinders must be maintained. URLs change; new resources become available; the pathfinder categories need to evolve; and resources should be continuously evaluated. Five years from now, the Clearinghouses will not be able to maintain their pathfinders as volunteer activities given increasing demand and the sheer growth in the knowledge base.

The ERIC database itself needs to be examined and probably redesigned

The ERIC system has always sought to be a comprehensive database by including virtually everything that has been written about education. The idea was that if the database is comprehensive, then with the right search strategy, a person could find everything that is important to them. With constant level funding, however, the reality is that ERIC is no longer comprehensive. Several education-related journals are not routinely put into the database. Acquisition of conference papers is often not aggressive. Many high quality, state and federal reports do not get into the database.

There is a real question whether the mix of documents being put into the ERIC database is optimal. To address this question, I looked at the demand and supply of ERIC citations. On the demand side, I analyzed characteristics of two datasets: 1) 56,073 ERIC citations retrieved by web patrons of the ERIC Clearinghouse on Assessment and Evaluation during three days in September 1999, and 2) all 35,433 documents ordered from the ERIC Document Reproduction Service in 1999. I looked at the target audience, publication type, clearinghouse codes, descriptors and publication years within each of the ERIC citations. I evaluated demand in terms of the absolute number and percent of retrieved citations with the addressed characteristics. I evaluated supply using the percent of documents in the ERIC database from 1985 with the addressed characteristics. Supply for the first data set included both CIJE and RIE documents; for the second data set, just RIE documents.

A major problem with retrieval percentage as a demand indicator is that it is heavily influenced by supply. If nearly all the documents in the database were of a certain type, for example, then we would expect nearly all the retrieved documents to be of that type. To gauge the relationship of demand and supply, I computed a probability ratio by dividing the percent of retrieved documents with the addressed characteristic by the percent of documents in the ERIC database with that characteristic. A ratio of 1.0 would indicate that supply exactly equals demand. A ratio greater than 1.25 is accepted as indicating that there is greater demand than supply. A ratio less than .80 indicates that the supply is greater than demand. Because the sample sizes are so big, all ratios are significantly different from 1.000. One should concentrate on practical significance. Table 3 shows supply and demand by target audience; Table 4 shows supply and demand by publication type.

This evaluation of supply and demand is in terms of quantity, not quality. While there may not be many documents of a certain type in the database, the few that are in the database may address the patron questions and completely meet the demand. Further, low demand does not necessarily indicate that a document type should not be sought. Demand may be low because patrons don't know that a certain type of document may be in the database. Other documents should be archived, such as publications from the National Center for Educational Statistics, and hence belong in the database even if they are in low demand. Nevertheless, ERIC acquisitions needs to be rethought.

Table 3
Supply and Demand of ERIC Citations by Target Audience

	On-line citations			Reproduced documents		
	Demand	Supply	Ratio	Demand	Supply	Ratio
Community	0.7%	0.5%	1.49	1.6%	0.7%	2.43
Practitioners	50.2%	18.3%	2.75	43.2%	18.9%	2.29
Counselors	0.3%	0.4%	0.91	0.8%	0.5%	1.56
Parents	1.3%	0.7%	1.79	2.5%	1.6%	1.54
Support Staff	0.1%	0.1%	0.41	0.1%	0.1%	1.21
Administrators	3.2%	3.8%	0.84	4.4%	3.9%	1.13
Researchers	2.5%	5.1%	0.49	2.2%	2.1%	1.07
Students	1.3%	1.6%	0.81	2.9%	2.7%	1.06
Teachers	14.6%	9.9%	1.48	11.0%	11.4%	0.97
Policymakers	2.3%	2.8%	0.83	3.0%	3.3%	0.92

p-ratios between .8 and 1.25 indicate that the percentages are practically equivalent.

Table 4
Supply and Demand of ERIC Citations and Documents by Publication Type

	On-line citations			Reproduced documents		
	Demand	Supply	Ratio	Demand	Supply	Ratio
ERIC Product	0.9%	0.9%	1.03	3.8%	2.3%	1.68
Thesis	0.6%	0.3%	2.27	1.4%	0.8%	1.65
Review Literature	9.5%	7.5%	1.26	10.5%	6.4%	1.64
Dissertation	0.4%	0.3%	1.29	0.9%	0.6%	1.42
Research Report	31.4%	30.6%	1.02	30.7%	25.9%	1.19
Conference Paper	9.5%	12.6%	0.76	31.7%	28.5%	1.11
Practicum Paper	0.5%	0.4%	1.50	1.3%	1.2%	1.09
Position Paper	14.4%	19.1%	0.75	9.5%	9.7%	0.98
Test, Questionnaire	2.1%	2.7%	0.75	6.0%	6.4%	0.93
Evaluative Report	11.4%	8.6%	1.33	10.0%	11.5%	0.87
Project Description	18.7%	20.9%	0.90	13.3%	16.8%	0.79
Bibliography	1.2%	1.7%	0.69	1.7%	2.2%	0.76
Non-clssrm Material	9.2%	7.5%	1.22	8.5%	11.3%	0.75
General Report	1.1%	2.3%	0.48	0.8%	1.1%	0.70
Teaching Guide	8.9%	9.2%	0.97	5.8%	8.7%	0.67
Confer. Proceedings	0.6%	1.1%	0.52	1.3%	2.2%	0.59
Historical Material	0.5%	1.2%	0.38	0.5%	1.0%	0.53
Directory	0.3%	0.6%	0.51	0.6%	1.2%	0.50
General Reference	0.1%	0.2%	0.65	0.1%	0.3%	0.47
Legal Material	0.6%	1.3%	0.45	0.7%	1.7%	0.40
Statistical Material	1.0%	2.2%	0.47	1.8%	4.6%	0.40
Instructional Material	0.5%	1.5%	0.36	0.8%	2.1%	0.39
Book	4.2%	2.1%	2.05	1.7%	8.0%	0.21
Audiovisual Material	0.1%	0.1%	0.53	0.0%	0.3%	0.09

p-ratios between .8 and 1.25 indicate that the percentages are practically equivalent.

Based on this analysis, the most popular types of documents are those flagged as written for practitioners and teachers; demand for these types of documents exceeds the supply in the database. Documents written expressly for researchers are also in demand; however, there appears to be an adequate supply of such documents. There is very little demand, however, for historical materials, directories, general reference material, legal material, and audio-visual material. Of special interest is that there is very little demand for instructional material. Right now, patrons do not come to ERIC in search of materials to use in their classroom. Yet, a significant portion of documents are selected for inclusion in the database on the grounds that a teacher

may find the materials useful. The data suggest that either ERIC markets the availability of these types of documents or puts much less effort into their acquisition.

Another read of these data is that demand exceeds supply for comprehensive materials such as literature reviews, books, theses and dissertations as well as evaluative materials. One reviewer pointed out that ERIC needs a better policy with regard to books. One one hand, there are databases for books and one could flood the database with textbooks. On the other hand, books providing insights into policy issues and books summarizing scholarly research are sorely needed and are not adequately being identified by ERIC.

I noted earlier that the scopes of work for the ERIC Clearinghouses have not changed significantly in the past 25 years. As shown in Table 5, this lack of change may be becoming problematic. Five clearinghouses are putting in significantly more documents than people seem to be demanding. Further, these clearinghouses supply about one-third of the documents in the ERIC database yet account for only one-fifth of the demand. This is not to say that the mix of documents in the ERIC database should be determined by demand, but rather the mix of clearinghouse activities needs to be periodically re-examined.

The ERIC database is composed of a documents database, *RIE*, and a journal article database, *CIJE*. While the documents in *RIE* are not peer reviewed, the *RIE* database has many advantages. It serves as a pre-print service for many papers originally presented at conferences. It serves as an archive for on-line journals, such as *Education Policy Analysis Archives*. And it contains state and federally produced reports. Most importantly, ERIC can make most of these documents available, either though the microfiche collection, or on-line for documents submitted after 1994. Thus, people can search the *RIE* database and usually obtain the documents.

The same is not true for *CIJE*. Patrons finding articles in *CIJE* need to go to an academic library, or if it is in one of a limited number of journals, order the document through a reprint service. Thus, *CIJE* presents additional work for the patron and there are alternatives. As mentioned earlier, OCLC, EBSCO, and the American Psychological Association provide on-line access to a growing number of journal articles. H.R. Wilson's Education Abstracts database covers many of the journals covered by *CIJE*. Perhaps, ERIC should drop *CIJE* in light of these other databases or perhaps index only those journals it can archive in *RIE*.

Table 5
Supply and Demand of ERIC Citations and Documents by Clearinghouse

	On-line citations			Reproduced documents		
	Demand	Supply	Ratio	Demand	Supply	Ratio
Ed Manage	9.6%	6.4%	1.52	9.2%	6.7%	1.38
Teacher Ed	7.7%	5.0%	1.53	6.5%	5.3%	1.24
Disab/Gifted	16.3%	8.2%	1.99	7.5%	6.5%	1.16
Early Child	9.6%	5.6%	1.71	7.9%	7.3%	1.09
Reading	9.4%	8.2%	1.15	10.3%	9.5%	1.08
Assessment	4.8%	4.6%	1.04	6.4%	6.0%	1.07
Commn Col	1.7%	2.8%	0.59	4.7%	4.7%	1.00
Urban	4.4%	4.0%	1.10	4.1%	4.2%	0.98
Counsel	6.1%	6.4%	0.96	4.6%	4.7%	0.97
Foreign Lang	3.2%	5.0%	0.65	5.2%	5.8%	0.89
Rural	3.6%	3.0%	1.18	2.9%	3.4%	0.87
Sci Math	5.7%	7.5%	0.76	3.2%	4.6%	0.70
Higher Ed	3.8%	7.4%	0.51	5.2%	7.5%	0.69
Info Resou	4.9%	8.0%	0.61	4.9%	7.2%	0.68
Career/Adult Ed	5.1%	10.0%	0.51	6.4%	11.4%	0.56
Soc Stud	4.0%	6.1%	0.65	2.8%	4.9%	0.56

p-ratios between .8 and 1.25 indicate that the percentages are practically equivalent.

Summary and Conclusions

ERIC's value lies is its ability to make educational information relevant to a wide range of consumers. ERIC does this by identifying resources, organizing information, applying information science, using literature, synthesizing information, developing new information tools, and developing special information products. While building the database has been its central activity, the most visible and useful ERIC accomplishments are not part of the core ERIC contract. They do, however, stem from the database and the process of building the database.

I have argued that ERIC will not be able to provide its current level of services much longer because demand is outpacing institutional and personal capacity. If ERIC maintains the low levels of service the government currently funds, without any effort to redirect and expand resources to meet demonstrated need, the education community will lose. ERIC is the information infrastructure for American education. While operating at a fraction of its capacity, it has effectively provided access to the wide range of information and information services produced across the country. The need to build this education information infrastructure is increasing. Perhaps more than ever, the education community needs to use information to inform decision-making at all levels. The daily instructional activities of America's 3,000,000 elementary and secondary school teachers should be guided by sound educational practices. Administrators and policymakers should benefit from the management decisions made by their colleagues. Research is a cumulative science and should be built on the methods and findings of other researchers with built-in mechanisms for dissemination and feedback from practitioners.

The need to build and maintain the education information infrastructure exists and the responsibility falls squarely on the U.S. Department of Education. Historically, there have been two criteria in determining the appropriateness of government interventions (programs):

1. limit the intervention of all governments to undertaking only those activities whose purposes are unattainable in the desired amount or quality through private action and where the public benefits equal or exceed the public costs of production
2. remand the public intervention to the lowest level (local, state, federal, or some combination) where the function can be effectively performed Mathtech (1998b).

By these criteria, providing information to the education community is clearly an appropriate federal role. Federal involvement in this area prevents needless duplication of effort, can assure better quality, can assure a range of products, and is cost effective.

ERIC could be doing a great deal more in its quest to provide information to the education community. I have mentioned several things ERIC is not doing:

- systematically gathering and analyzing patron satisfaction information
- systematically analyzing queries and search strategies to identify user community training needs and topics of interest
- designing benchmarks and systematically evaluating and improving the quality of reference services
- producing management resources to be shared across the 16 clearinghouses
- gathering and analyzing high-quality usage statistics
- vigorously pursuing acquisitions
- vigorously acquiring and cataloging web resources
- providing access to the journal literature
- marketing and disseminating itself to a broader audience
- preparing articles about the project

I have also mentioned some things ERIC is doing, but should do more of:

- developing a wide range of content-oriented training material
- disseminating information about itself
- establishing on-line electronic journals

- creating access to full-text documents
- posting quality materials on the Internet as they are acquired
- providing more syntheses and information products

ERIC has amply demonstrated the need to infuse information science in the various educational subject matter disciplines, and its ability to do so. ERIC needs to expand if it to institutionalize its current level of service and respond well to information requests of the 21st century. Properly funding the volunteer activities will allow for more concentrated effort and inevitably higher quality and usability. Just as educational practice and advances should be based on research, ERIC also needs a program of research into ways of being more responsive to user needs.

The ERIC of today is confronted with a vastly different user base, mode of access, mix of services and set of demands. No, ERIC is not ready for this new environment. It has the ability, but not the resources and not the guidance. In my view, this will hurt not only the research community, but more importantly, teachers and practitioners who have neither the time, desire or ability to sift through today's overwhelming volumes of potential resources.

Notes

1. . Based on a paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA April 24-28, 2000.
2. This study did NOT receive any funding from the U.S. Department of Education.

Endorsements

The Directors of the following ERIC Clearinghouses have indicated that they concur with most, but not necessarily all, of the points raised in this article:

- ERIC Clearinghouse on Higher Education,
- ERIC Clearinghouse on Counseling and Student Services,
- ERIC Clearinghouse on Educational Management,
- ERIC Clearinghouse on Elementary and Early Childhood Education,
- ERIC Clearinghouse on Languages and Linguistics,
- ERIC Clearinghouse on Urban Education,
- ERIC Clearinghouse on Reading, English and Communication,
- ERIC Clearinghouse on Disabilities and Gifted Education,
- ERIC Clearinghouse on Information and Technology,
- ERIC Clearinghouse on Social Studies/Social Science Education,
- ERIC Clearinghouse on Community Colleges, and
- ERIC Clearinghouse on Rural Education and Small Schools,
- and I am the Director of the ERIC Clearinghouse on Assessment and Evaluation.

References

Colker, Laura J. (2000). *Reminiscences from the Field: The Continuing Story of ERIC*. Springfield, VA: Dyntel Corporation.

Eisenberg, Mike; Henson, Jane; Howley, Craig; Cawley, Nancy; Ramirez, Bruce; Rothenberg, Dianne (1997). Rising Expectations: A Framework for ERIC's Future in the National Library of Education. Report of the ERIC Operations Framework Task Force. (ERIC Document Reproduction Service No. ED410969)

ERIC Annual Report 1999. Available. Online: http://www.accesseric.org/resources/annual99/index.html

Hoffman, Charlene M (1995). Federal Support for Education. Fiscal Years 1980 to 1995. Washington, DC: National Center for Educational Statistics. (ERIC Document Reproduction Service No. ED392853).

Lynch, Clifford A. (2000). Technology and the ERIC System: New Opportunities and New Impacts. Discussion Draft for March 8, 2000 ERIC Director's Meeting.

Mathtech (1998a). Shaping the Future of Educational Research. Development, and Communication. OERI Reauthorization Working Papers Prepared for the National Educational Research Policy and Priorities Board (ERIC Document Reproduction Service No. ED428105).

Mathtech (1998b). A Review of the Federal Role and the Department of Education Structure. OERI Reauthorization Working Papers Prepared for the National Educational Research Policy and Priorities Board (ERIC Document Reproduction Service No. ED428105).

Rudner, Lawrence M. (1997). ERIC and NLE: Past, Present, and Future. Paper prepared for the National Educational Research Policy and Priorities Board. May 30, 1997.

Rudner, Lawrence M. (2000). Who Is Going to Mine Digital Library Resources? And How? *D-Lib Magazine*, 6(5) May 2000. On-line: http://www.dlib.org/dlib/may00/rudner/05rudner.html.

Stalford, Charles B.; Stern, Joyce D. (1990). Major Results of a Survey on the Use of Educational R&D Resources by School Districts. Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990). (ERIC Document Reproduction Service No. ED329212).

Stonehill, Robert M. 1992. "Myths and Realities About ERIC." ERIC Digest. Washington, DC: Office of Educational Research and Improvement, National Center for Education Statistics. (ERIC Document Reproduction Service No. ED345756).

Trester, Delmer J. (1981). *ERIC: The First Fifteen Years: 1964-1979.* Columbus, OH: SMEAC Information Reference Center.

About the Author

Lawrence M. Rudner, Director
ERIC Clearinghouse on Assessment and Evaluation
Department of Measurement, Statistics and Evaluation
1129 Shriver Laboratory
University of Maryland, College Park
College Park, MD 20742

Email: rudner@ericae.net

Lawrence Rudner is with the College of Library and Information Services, University of Maryland, College Park. He has been involved in quantitative analysis for over 30 years, having served as a university professor, a branch chief in the U.S. Department of Education, and a classroom teacher. For the past 12 years, he has been the Director of the ERIC Clearinghouse on Assessment and Evaluation, an information service sponsored by the National Library of Education, U.S. Department of Education. Dr. Rudner holds a Ph.D. in Educational Psychology (1977), an MBA in Finance (1991), and lifetime teaching certificates from two states.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina— Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State Univer ty

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 655 times since August 30, 2000

Education Policy Analysis Archives

Volume 8 Number 45 August 30, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the EDUCATION POLICY ANALYSIS ARCHIVES.
Permission is hereby granted to copy any article
if EPAA is credited and copies are not sold.

Articles appearing in EPAA are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

The Social Construction of School Failure: Leadership's Limitations

Merylann J. Schuttloffel
The Catholic University of America

Abstract
A case study highlights barriers encountered by an urban school principal in implementing reforms within the context of the Kentucky Educational Reform Act. By comparing the competing expectations of Miller's (1995) five capitals and Ianneconne and Lutz's (1970) dissatisfaction theory, the case study dramatizes that Site-Based Decision-Making councils exemplify a policy decision that ignores the practical realities of distressed schools. The lack of congruence between policies and the school reality makes implementation of school reform predictably unsuccessful.

Introduction

Widespread press coverage of the march for civil rights in the 1960's opened the public's eyes to center city poverty and rural regions with third world living conditions. These images made believers in the American tenets of justice and equality attack the status quo (Sergiovanni, Burlingame, Coombs, & Thurston, 1999). Social activism compelled idealistic reformers to the optimistic assumption that public policy could dictate a more just society (Kantor & Lowe, 1995; Spring, 1998, 1997, 1976). Public schools became the laboratory to experiment in the social reconstruction of society (Corbleth & Waugh, 1995; Levine, Lowe, Peterson, & Tenorio, 1995; Fullan, 1993; Steele, 1992, 1990).

During the intervening years many educational reformers have attempted to translate their social justice assumption into policies that impact practice. Unfortunately, at the same time, the urban community reality frustrated reform progress. The failure of numerous reforms left dismal images of urban life that continued to march across the television screen or create a mental picture

with grim statistical data (Sarason, 1997, 1995, 1990). As recently as the 1998-99 school year, well-intentioned policy mandates continued to fall short of a real solution to the social construction of failure that plagues too many students in urban public schools (Clark, 1999; Comer, 1998). These same schools house the majority of America's poor and minority students.

Kentucky Educational Reform Act and Site-Based Decision-Making Councils

On June 8, 1989, the concluding opinion of the Supreme Court of Kentucky ordered the state's school system dismantled. Justices expanded the case from an examination of the state's school-finance distribution to the public school system's limits. At a recent celebration, Former Chief Justice Robert Stephens recalled, "I realized as I was writing that we weren't talking about a few things that needed to be fixed; we were talking about the whole thing." The shock wave that followed the court's ruling inspired the 1990 Kentucky Education Reform Act. The impact of KERA continues to shape policy for public schooling and education in Kentucky into the next decade.

Too often the very policies created to improve urban schools and their educational possibilities prevent school improvement. Site-Based Decision-Making councils are such a policy example (David, 1995-1996). An SBDM council consists of teachers, administrators, parents and community members. The limitations of Site-Based Decision-Making councils and their contribution to the unrelenting failure of some urban schools, ties directly to policy mandates created by state policy makers with little understanding of the urban school reality (Fraser, 1997).

The argument that parent involvement is a necessary component for school improvement has been generally accepted since Coleman's report introduced the concept of social capital. Many others have expanded this concept to confirm their position that parent involvement is the key to school improvement. Those policy makers who included the SBDM council requirement in KERA believed in the engagement of parents and community members in school improvement. Students in high achieving schools seem to affirm their belief and proponents enumerate the parents' contributions to the schools. However, fairness also requires proponents to delineate the characteristics those parents bring with them to the school: moderate to affluent income, advanced education, productive community ties, and an understanding of the political elements of the district's school system.

The opposing argument builds a case proposing that a difference exists between a general plea for parent involvement and the benefits implied in particular parent- school-community relationships. Including positions for parents and community members on a Site-Based Decision-Making council does not insure school improvement. The urban school reality is more complex than that approach considers. Comer and Haynes (1991) suggest that schools alienate low income parents from school involvement by ignoring their pressing basic needs. When parents feel ill-equipped for informal volunteerism it is not likely these same parents are candidates for high-stakes governance positions (Cavaretta, 1998; Gismondi, 1999).

Guskey and Peterson (1995-1996) enumerate the weaknesses inherent in the site-based decision-making model to include:

- the power problem,
- the implementation problem,
- the ambiguous mission problem,
- the time problem,
- the expertise problem,
- the cultural constraints problem,
- the avoidance problem, and
- the motivation problem.

Each of these problems contributes to the external pressures principals experience as they initiate change within their building by developing a capable parent and community constituency. Unfortunately, these caveats received little consideration within the Kentucky model for Site-Based Decision-Making councils.

By the beginning of the 1998-99 school year sufficient evidence had accrued to demonstrate that the KERA reforms were not taking hold at the anticipated pace. Kentucky had already committed ten years to implementation. Although the results were unimpressive, reformers continued to believe that modifications of the plan and more time invested would lead to the intended improvements. By postponing deadlines for the schools' assessment until 2014, a new cycle begins in 2002.

Research Framework

Five community capitals: Miller's argument.

In his text, *An American Imperative,* Miller (1995) builds a theoretical argument for the social construction of minority student failure. According to Miller, the lack of specific parent and community resources, which he defines as human capital, social capital, health capital, financial capital, and polity capital aggravates the urban school reality. Human capital is the knowledge and skills required to function in a technologically complex society like the United States in the twenty-first century. Social capital is "the norms, the social networks, the relationships between adults and children that are of value for the children's growing up"(Coleman, 1990, p. 36). Health capital is the ability to sustain good health through nutrition and preventative care. Financial capital is the income and savings that provide the ability to purchase other resources and advantages. And polity capital refers to the benefits that the community at large provides for all its members. Polity capital acknowledges the interdependent nature of society today. Grounding his theoretical rationale in the non-school urban reality, Miller intends to impact school practice.

Miller argues that due to weak economic expansion and multiple social hardships, the urban school community requires the school to be a conduit of the five capitals for its children and their families. Miller emphasizes the school's role in developing parent-school-community relationships within the urban school community that are "capital-adding" for students. His capital resources, existing as they do outside the student, demonstrate benefits beyond the student's control that further motivate students to achieve. The practical implication of Miller's theory is that individual student effort, while necessary and important, is not a sufficient contribution to dramatically raise en masse student underachievement. Capitals that rest outside the student are also integral for student success.

Clearly, distressed urban schools suffer from their lack of success and spiraling failure. Disappointing student performance results fuel the metaphorical autopsy of the urban school (Shirley 1997, p.4). The public's perception of the urban school portrays a place to be fixed, restructured, or perhaps even abandoned. This negative perception of the urban school reality has changed little in thirty years with urban schools lagging behind in nearly all quantitative assessments of educational reform progress.

Dissatisfaction theory: The Ianneconne and Lutz argument

Like many state reform policies, the central character in charge of KERA's school reform is the building principal. Principals are often credited with the successful reform of their school (Blase *et al.*, 1995; Goldring & Rallis, 1993; Murphy & Louis, 1994; Peterson & Valli, 1994; Speck, 1999). From this leadership assumption the individual school site has emerged as the crucible of educational reform. This scenario places the building principal in a position of dwindling legal authority, diminishing traditional power, and increasing academic and social responsibility for students. Principals who have successfully improved their school may provide a model, but improvement models do not easily transfer within a locally driven educational system. Reforms that might prove successful in one school or district may confront multiple restrictions within another school, such as an incompatible school culture, a reluctant parent community, or minimal teacher support. Within these inconsistent settings, it seems that each principal builds school reform with little anticipation of success until it transpires within that very building.

In the current school reform environment, crediting successful change to the action of a building principal may be as misleading as the assignment of failure solely to the same principal. Ianneconne and Lutz (1970) pointed to the profound effects external forces exerted upon school change in their dissatisfaction theory. Their dissatisfaction theory states that members of a school community initiate change based on their dissatisfaction with the school's performance. The dissatisfaction theory implies a level of political sophistication on the part of the school community. Informed parents and community members must know what school services are potentially available to them. Too often a parent's tacit beliefs and personal experiences with schooling and learning drive their expectations.

Weakening the dissatisfaction theory for urban schools, those parents whose negative experiences as students color their school activism as adults. Evaluation of curriculum, extra-curriculum, and leadership qualities are typically outside the experiences of most urban school constituents. Parents who are aware of possibilities for school improvement may not know how to manipulate the system to make their expectations for the school a reality. Further, those parents who are more politically proficient routinely withdraw to another school.

Ianneconne and Lutz's proposal that superintendents can only function as change agents within

a cast of supportive external players points to the ineffectiveness of school reform that fails to acknowledge the school's external environment (Peshkin, 1978; Smith *et al.* 1971, 1986, 1987, 1988). With site-based management, the urban principal's role is a political role, more similar to that of a superintendent under the traditional local school board.

Summary

Miller argues that the sources of support students require for achievement are fundamentally lacking with the urban school community. He proposes that the urban school will continue to fail to raise student achievement unless an expansive support system prevails within the school community. Successful inner city Catholic schools provide evidence that supports Miller's theory (Bryk, Lee, & Holland, 1993).

Iannaconne and Lutz's dissatisfaction theory rests on the premise that community members are capable of becoming change agents within the school. Dissatisfaction with the school requires knowledge of a school's potential and the skills to initiate the needed change. As Miller suggests too often parents in disadvantaged communities do not have the five capitals within their adults so that parents are not capable of providing these capitals for their children.

Detailed descriptions of a distressed urban school help to illustrate the difficulties with school reform, within a single district under state mandated reforms, that ignore the arguments of Miller, Iannaconne, and Lutz. The following case study provides a window to view assumptions made about school leadership and policy implementation in an urban school (Ashbaugh, 1991; Hamel *et al.*, 1993; Kowalski, 1991; Salter & Tapper, 1985).

Johnny Flynn (pseudonym), principal of a Kentucky public middle school, plays the central character in this case study that portrays the urban school reality. His school, John Adams Middle School (pseudonym), represents distressed urban schools operating under reform guidelines. Through his willingness to share the details of his school's context and his personal dilemmas with school improvement, Flynn hopes to influence the public's perception of the urban school reality. He further believes that by shaping public perceptions, he ultimately helps his students to receive the capitals they require to improve their academic performance. As Flynn's case unfolds, the significant connection between the public's perceptions of the urban school reality and the impact of these perceptions on his school's reform efforts becomes clearer.

The Case of Johnny Flynn and John Adams Middle School

The current reality.

Like many southern cities in the 1970s, the urban site of John Adams Middle School desegregated by a court ordered ruling. Socially painful and financially costly, busing students still balances the African-American and "other" racial categories within the district's schools. Today these two categories simplistically betray the many enrolled minority groups. Principals acknowledge that some past district programs were instituted to slow earlier "white flight" trends. In the current reality, poverty and class issues often displace previous racial barriers, but John Adams Middle School still reflects the public's perception that a low performing school links poverty *and* race.

Johnny Flynn has been principal of John Adams Middle School throughout the decade of state reform implementation. He questions numerous policies designed to reform schooling. Flynn admits that his school has been unable to meet performance goals, in part, due to policies that allow schools and classrooms to re-segregate by race and class (Orfield & Yun, 1999).

Accountability and school choice are features of Kentucky's state reform. These two very public items interact to complicate life for Johnny Flynn. Test scores at John Adams flutter below their goal just as the recruiting environment within the district reaches a competitive frenzy. The district's modified choice plan allows parents to seek out the most appropriate school program for their students. The result is that individual schools use a variety of marketing strategies to attract students. Flynn readily admits that recruitment time amplifies his awareness of the school's problem with public perceptions. Publicized information about John Adams's test results certainly constrains recruitment of high achieving students. Some parents openly discuss their reluctance to enroll their students in John Adams due to low test score results and the school's negative reputation for performance.

Public perceptions and recruitment.

The district's arrangement of specialty programs, magnet schools, and traditional schools, places a neighborhood school, such as John Adams Middle School, at a distinct recruitment disadvantage. Specialty programs and magnet programs (e.g. Science, Math & Technology) are open to neighborhood minority children, but are routinely filled with white middle and upper class students who have parents with the knowledge to maneuver their way through the district's application process. Typically, any parent who takes advantage of the choice options enrolls a student who meets grade level achievement expectations, and the parent is actively involved with the student's education. Losing these students is a particularly excruciating drain on John Adams Middle School. The enrollment situation wreaks double jeopardy as the top students are lost as contributors to the school's overall assessment scores and as positive role models to the rest of the student body. The parent is also lost as a contributor and a positive role model within the school community (Cavaretta, 1998). These enrollment incidents multiply, making recruitment extremely frustrating for Flynn and his staff. There exists a certain cynicism at an urban school like John Adams that their enrollees are "what's left over." This situation creates low morale that ripples through the school's faculty, staff, and students.

When Principal Flynn responds to questions about his "choice or specialty" program at John Adams, he jokes that he is the "special education magnet." Flynn does not intend his comment to be disrespectful to these students, he simply acknowledges that John Adams has a high proportion of special education students. John Adams enrolls the second highest percentage of special education students in the district (2nd out of 24 middle schools). The school with the highest percentage of special education students is an equally distressed school.

The school categories in Table 1 include an urban school (John Adams), a neighborhood/home school and a traditional school. A neighborhood or "home" school is the school where the district assigns a student by home address. A magnet school attracts students district-wide with a special program. Traditional schools offer a program espousing enhanced home-school partnerships, regular homework, appropriate behavior, and high academic performance. The popularity of the traditional programs caused the district to increase the number of these schools in recent years. Option or specialty programs, traditional, and magnet programs are open to all students within geographical attendance zones.

The data in Table 1 indicate the discrepancies in special education enrollment between the various categories of schools. John Adams represents the distressed urban school as the data in Table 2 will help verify. The percentages of students assigned to the"resource" or "self-contained" category significantly impact the disbursal of resources. Special education students who are in the "resource" category are able to attend regular classes but receive supplemental special education services.

Table 1
Placement Rates for Special Education (Resource) and Regular (Self-Contained) Classrooms

	Total %	%Black Resource	%Other Resource	Total % Resource	%Black Self-Contained	% Other Self-Contained	% Total Self-Contained
John Adams	**17.3%**	**3.3%**	**6.3%**	**9.6%**	**3.7%**	**4.0%**	**7.7%**
Neighborhood	11.4%	2.0%	7.2%	9.2%	1.2%	1.0%	2.2%
Traditional	1.5%	0.5%	1.1%	1.5%	0.0%	0.0%	0.0%

By comparison, those students who are assigned to self-contained special education classrooms require more intense services. A self-contained special education classroom has a limited number of students per teacher and requires a teacher licensed in special education. There is no clear explanation why John Adams has a higher percentage of these self-contained classrooms, but one possible reason is the available space. Often district decisions about a program's location reflect the availability of space rather than consideration of other factors. The numbers dramatically illustrate the difference in student population between the selective traditional program, the home-neighborhood school, and the distressed urban school.

Principals readily admit that special education programs are high maintenance, demanding attention to the legal requirements, teacher and aide licenses, and parent communication/meetings. A public perception in the district that the students at John Adams were unusually "bad" aggravates a

difficult recruitment situation that includes all personnel: teachers, aides, cafeteria, and custodial staff. Flynn admits his frustration with having too many substitute teachers or aides in the special education classrooms or, even worse, long term substitute teachers who might lack the appropriate training.

Flynn's situation is not unique and unfortunately reflects national trends. On June 24, 1999, the Education Commission of States, a non-profit group that helps policy makers work to improve student learning, announced the group's upcoming focus on the need to attract competently qualified teachers for special education classrooms in "hard-to-staff" schools. The organization received a grant from the DeWitt-Wallace Reader's Digest fund to finance the initiative, *Focusing State Policy on High-Quality Teachers for Hard-to-Staff Schools*. Wyoming Governor Jim Geringer, the 1999-2000 ECE chairman, states," Common sense tells us, and research confirms that the number one factor in determining how well students do in school is the teacher" (McElhinney 1999, p.1).

Time that Flynn invests wrestling with special education issues is time taken away from other dimensions of school reform. His colleagues at the traditional or even the neighborhood schools designate that time to building the curriculum, supervising teachers, working with community leaders, or developing parent leadership. Flynn's daily reality is not the same.

Principals of a distressed school, like John Adams Middle School, deal with a student population that arrives at school with life experiences from a reality far distant from preschool and elementary school experiences that assist in academic preparation. Flynn describes his students and his school with care.

> I think the most challenging thing would be the things that our kids----what they come with, baggage that they bring with them primarily. They come from single parent homes, coming from homes where the parents are not involved that much with the schools, coming from homes, there's not a whole lot of money in homes, and also I would say their academic achievement is low at the time in which they come to you and you have to turn all those around.

Table 2
Percent of Students on Free & Reduced Lunch

	1996-97	1997-98	1998-99
John Adams	80.35 %	79.28 %	80.36 %
Neighborhood	56.67 %	57.91 %	57.96 %
Traditional	15.42 %	15.62 %	21.30 %

Data on Free and Reduced Lunches serves as a standard indicator of poverty within a given school population. The data could be even more accurate if "Free" and "Reduced" were disaggregated. This would enable a clearer distinction between the John Adams public housing population and that of the predominately working class neighborhood school.

Public perceptions and accountability.

Forty-five years after the *Brown v. Board of Education* ruling, the 1999 Civil Rights Project report for Harvard University, "Resegregation in American Schools," points to accountability measures, such as high stakes testing, that "punish students in inferior segregated schools, or even sending more children to such schools while simultaneously raising sanctions for those who do not achieve at a sufficiently high level" (Orfield and Yun, 1999). John Adams Middle School reflects this trend with its loss of performing students to other schools while the student body assigned to John Adams sinks into deeper poverty and social disarray.

Measurable disparities in income do not completely capture the disadvantages of the urban school. Miller's description of the non-school-based disadvantages of urban minority students that resonate with the John Adams' student population. These disadvantages profoundly affect student potential before students enter school. These disadvantages are almost impossible for the school to remedy alone. To further illustrate Flynn's point about the students that John Adams enrolls, Flynn shares the results of the sixth grade reading placement test. "We only had 14 out of 300 some odd 6th graders that were reading on level. Urban principals recognize that reading is the fundamental skill that must be improved. Reaching grade level performance appears to be an overwhelming task

considering the number of students that require assistance. These students' success on the state's assessment test looms near impossibility.

Table 3
KRIS Assessment Scores

	Baseline	Goal	Index
John Adams	**27.2**	**34.5**	**27.6**
Neighborhood	30.2	37.2	33.8
Traditional	53.6	58.2	56.3

KIRIS has been Kentucky's version of a high stakes assessment test. The test results over the years of KERA reform have been disappointing. During this anniversary, the assessment tools and processes underwent examination for revisions, including the subsequent evaluative rankings. The data in Table 3 reflect a system used prior to the revisions. A school's testing performance is public news, but often remains a source of confusion to the public. Parents question how a school ranks "in decline" while their academic teams hold high honors in state competitions. Principals are weary of explaining that ranks were determined solely by the KIRIS assessment. The school's scores must be moving toward the goal score to be considered improved.

Intertwined with the testing debate are special education issues. Marking the current anniversary, some Kentucky legislators promote the increase in fourth grade reading scores as a sign of KERA's impact. Critics counter that in 1998 fewer special-education students were tested than in 1994, making the gains an illusion if the testing population has changed. Mark Musick, the chairman of the National Assessment governing board, believes Kentucky students performed better this year even with the testing population adjustments. Others have remained critical stating that there will never be any way to know the real results. Musick reminds state officials that no test is incontrovertible, in spite of careful monitoring. During the decade of KERA, Congress changed federal law to mandate the testing of students with disabilities as a condition for federal aid for special education. Under these conditions district pressure for improved testing performance increases for Flynn and his teachers. Again, the high numbers of special education students at John Adams weigh heavily on Flynn's efforts for school improvement.

In spite of state and district efforts to funnel supplemental programs and extra funds into distressed schools, assessment tests still fail to demonstrate adequate progress. John Dornan, executive director of the Public School Forum of North Carolina, a Raleigh-based group for school reform, states that, "It's possible very accurately to predict the schools most likely not to succeed in high-stakes tests." Dornan explains further that in significant school reform the school provides a value-added environment. In other words, the school does bring an effect to achievement. The challenge for urban schools is that considerable value must be added, or considerable disadvantage alleviated, for students to experience a substantive benefit from their educational experiences.

One area that highlights the disconnection between reform expectations at John Adams Middle School and life in the urban community is the suspension rate. The suspension- rate and distribution display the contradiction between the context of schooling and the reality of the urban student's life. Principal Flynn believes that one of the chief barriers to successful student achievement that he regularly encounters is the lack of student self-discipline:

> The kids seem to not show a lot of self-discipline so I think that is one of the major issues that we deal with.

Flynn implies that self-discipline impacts student performance in a variety of ways including their ability to learn to read. Self-discipline is an example of a skill that students must have to be successful in school behavior and academic performance. Unfortunately, the urban community environment does not assist students to appropriate structure and discipline into their lives. This lack of self-discipline then handicaps the student at school.

The suspension rate of John Adams in 1996-97 was nearly the equivalent to the suspension of every student in the school (student enrollment = 921). The 1997-98 figures show a drop of about 30 % at John Adams and the neighborhood school (Table 4).

Table 4

Suspensions

	White Male	White Female	Black Male	Black Female	Total
John Adams	**187**	**70**	**202**	**118**	**577**
Neighborhood	97	18	73	9	197
Traditional	8	3	7	1	19

Suspensions add to the inconsistent academic preparation some students receive. And in turn, these students are unable to reach an appropriate score on the state's assessment. Behavior that requires a suspension adds to a chaotic classroom environment that does not support learning for classmates either. Too often young African- American male students consider a suspension a sign of defiance to a white establishment. Too often school personnel fear a suspension serves as preparation for more extreme forms of antisocial behavior including crime. The alternative, the in-house suspension, also accounts for time lost from the classroom, but an in-house suspension is the school's attempt to keep students within the building where there might be some positive influence.

Site-Based Decision-Making as Tool to Assist Reform Efforts

Within the urban school reality, how does the Site-Based Decision-Making model assist the principal to improve the school's accountability results? The descriptions of John Adams Middle School and the principal's daily life attempt to connect the urban school reality with theoretical rationales for the policy on Site-Based Decision-Making (SBDM). Flynn speaks about his difficultly in facilitating a SBDM council to meet its intended purpose within his school community.

> Also, we don't get the community leaders involved with the schools, I'd say in school which they have in the suburbs, and then the attitude of some of the parents. Maybe they weren't that successful in school. School left a bad taste in their mouth so they tend to think the same way and that attitude is displayed in their kids when they come to the urban school.

Flynn's word ring similar to Burns' position that some parent's previous negative experiences in school impacts their interactions with the school and contaminates their child's viewpoint of school and learning. Just as the John Adams' students suffer from their school's negative public image, the parents also bear the burden of the public's negative perception of adults who wallow in poverty, single parents who receive welfare checks, reside in public housing projects, and are unemployable. Many of John Adams' parents feel intimidated by school personnel with their "school speak" and some parents are openly hostile, shaped by their own negative experiences with teachers and schooling.

Flynn must organize the SBDM council, fill the positions, train the members, and then administer the policies created by his local Site-Based Decision-Making council. Urban principals struggle to develop more sophisticated interactions within the school's Site-Based Decision-Making council members but they are often thwarted by the sheer lack of resources. Johnny Flynn's daily tasks at John Adams Middle School demonstrate the gap between good intentions as policy and the reality of the urban school. Site-based decision-making councils are the practical venue for parents to become involved with the policy decisions for John Adams Middle School.

Closing Reflections

Supporters and critics of Site-Based Decision-Making muster convincing arguments. On one side, the concept of Site- Based Decision-Making councils remains a worthy element of school reform. Community leader and parent participation in policy decisions for their local school seems reasonable.

On the other side are urban schools like John Adams, with principals like Johnny Flynn, who add his Site-Based Decision-Making council to a long list of activities that take his time and energy and are not easily implemented within the urban school community.

Side-Based Decision-Making councils are predicated on the assumption that the parent and

community membership will provide the means to acquire non-school resources that advance student performance. The urban school, due to its inherent characteristics including poverty, minority membership, and lack of political acumen diminishes the power of the SBDM council to assist the urban school improve achievement. This flaw in the Site-Based Decision-Making model remains over-looked due to the apparent success of the model within other socioeconomic strata. The naive assumption remains that by manipulating (because they are not necessarily increased) resources at the school level, the urban school will catapult to a competitive level.

An understanding of the urban school reality makes it clear that non-school capitals also require enhancement. In order for the SBDM's contribution to reach the maximum, the public's perception of the urban school must be expanded to include its capital deficient community. These augmented capitals will develop the requisite conditions to dramatically improve student academic performance.

The Site-Based Decision-Making model generates its power and strength from the various capital-resources parents, community members, and school personnel bring to the school(Cavaretta, 1998; Gismondi, 1999; Comer & Haynes, 1991). The flaw in the Site-Based Decision-Making model for the urban school is the very lack of these capital-resources within the community's membership.

Related considerations.

Several side issues emerge from an observation of the effects of the Site-Based Decision-Making model on a distressed urban middle school. First, there is the issue of school leadership. A local Site-Based Decision-Making council lacks the broad view of the district. Local SBDM council members seldom consider the advantages of changing the school's principal since they are so closely bound to the current leadership themselves. This is particularly true in distressed urban schools where parent, and perhaps novice teacher participants, often lack experience in assessing leadership quality. Members are often suspicious of a new individual from outside their community.

In turn, under the current SBDM model, a principal is unlikely to attempt to force a change in leadership by applying to another school. A principal bears the same image difficulties that students carry. Consequently, a principal is reluctant to risk credibility with their current school by applying for another position. Should a principal make application to another school, and if the principal was unsuccessful during the hiring process and had to return to the current school, the faculty, staff and parents might interpret those actions as disloyal, contaminating future interactions. Under the SBDM model, seeking a new principal position is a very difficult situation for any principal to politically finesse. Typically, the urban principal is left to await some other cue, perhaps from the central district office, for any possibility of changing schools. Ultimately, the instigator of principal change is the superintendent. Oftentimes a building level leadership change is a necessary requirement for school change.

Second, within the SBDM council, energy and interest focuses on the members' local school. This myopic approach handicaps distressed schools that require input in resources and expertise from other schools or the broader district community. Challenging a local SBDM to feel social responsibility for other children in the district, not enrolled in their local school, is a difficult endeavor. But, if students in distressed communities must rely on local resources, their plight seems an inevitable social construction of school failure.

Third, other policies such as the modified in-district choice plan further disadvantage distressed urban schools by allowing positive contributors to the school to move on to healthier settings. The distressed school loses not only a positive role model in the student, but typically a parent who is a capable partner with the school. This "capital drain" creates problems similar to "white flight" in its effect on the urban school. Parents who are aggressive about their children's welfare should not be penalized for wanting to improve their situation, but the message is clear that a schools must be made effective or closed.

Policy implications

Returning to the arguments of Miller, Iannaconne, and Lutz, an analysis of John Adams Middle School reveals that the defect in the dissatisfaction theory for the urban school rests with the community's deficiency in Miller's five capitals. The assumption that the constituents of a distressed urban school will conclude that their SBDM council's membership is ineffective, or their principal is incompetent, or the district inadequately represents their interests, is improbable. It is unlikely that this dissatisfying situation will motivate community members to become politically active or initiate

a change in leadership.

Site-Based Decision-Making councils as the centerpiece of community participation in urban school improvement legislation like KERA require modification. Two issues impact the effectiveness of the Site-Based Decision-Making model on reform efforts at urban schools.

First, the dissatisfaction theory implies a level of political sophistication on the part of the school community. Parents and community members must recognize the lack of quality in their school's performance. Then, parents and community members must know how to manipulate the school system to provide services to increase the quality. Too often the urban school community lacks business and industry leaders capable of exerting power and political influence that produces positive results for their local school. Those parents who are aware of possibilities for school improvement, but do not know how to manipulate the system to make their expectations a reality, routinely withdraw.

A second impediment to school reform at an urban school comes from the larger district community's lack of polity capital. Outsiders are reluctant to initiate the substantive reforms necessary to dramatically improve urban schools. The perception that improvement at urban schools like John Adams will require a sacrifice from their school community is not attractive to those outside the urban school community. Most outsiders lament the state of affairs at urban schools, but this lamentation accompanies stated relief that their children do not attend such a distressed school. Too many district constituents do not consider distressed urban schools their school community's responsibility. This lack of commitment to the common good seriously handicaps urban school improvement. The more politically savvy constituents of Flynn's colleague principals have left John Adams Middle School alone to maneuver out of its situation.

At the core, the lack of political acumen by the insiders at John Adams Middle School, and the fundamental lack of polity capital contributed by the outsiders in the district community, perpetuates the current situation. The lack of polity capital, an acknowledgment of the interdependent nature of the community, diminishes the urban principal's ability to accelerate urban school improvement. Autonomous Site-Based Decision-Making councils aggravate the development of the requisite polity capital by sustaining an "us/them" mentality.

School autonomy, which was propagated as a virtue by KERA's school reform movement, has become a destructive vice. School reform has become so idiosyncratic that an individual principal must compete for students, generate supplemental funding, develop community relationships, preferably with generous businesses, and provide leadership for the school in the political arena of district politics. Principals from even modestly affluent school communities have multiple means to attack this situation. The reservoir of parent resources (i.e. volunteer time, fund raising, political connections) make their Site-Based Decision-Making council appear successful. The public perception of a school like John Adams includes an implicit assumption that its deficient performance rests within the people living in the school community rather than within the negative capitals present in the school community. The incriminating evidence might extend to beliefs in racial inferiority, "their" lack of effort and willingness to improve, or simply the obvious characteristics of the community (i.e. minorities, single parents, low SES). The SBDM model requires the distressed urban school community to generate resources it does not have, and holds no one outside the school community responsible for the social construction of failure for urban students.

Kentucky's Site-Based Decision-Making council attempts to assemble parents and community members together for the improvement of public schooling. The concept of school-parent-community involvement intends to generate the positive attributes of Miller's capitals and bring them to the schoolhouse. Unfortunately, the flaw in applying the Site-Based Decision-Making council model to the distressed urban school is less with the concept than with a deceptive perception of the urban school reality.

KERA's tenth anniversary and the on-going national attention to its reform initiatives provide an opportunity to modify or supplement the SBDM model for the distressed school context. The benefits of parent and community involvement should not be abandoned, but capital development requires a broader community responsibility for distressed schools. A comprehensive community focus that develops the capitals within the entire district, or perhaps even statewide, increases student improvement in all schools.

School reform legislation that fails to take into consideration the distressed urban school reality creates a paradoxical environment for school change.

References

Ashbaugh, C. R., & Kasten, K. L. (1991). *Educational leadership: Case studies for reflective*

practice. New York: Longman Publishing Group.

Blase, J., Blase, J., Anderson, G. L., & Dungan, S. (1995). *Democratic principals in action: Eight pioneers*. Thousand Oaks, CA: Corwin Press, Inc.

Burns, R. (1993). *Parents and schools: From visitor to partners*. Washington, DC: The National Education Association.

Cavaretta, J. (1998). Parents are a School's Best Friend. *Educational Leadership*, *55*, (8), 12-15.

Clark, J. L. (1999). "KERA has failed across the board." http://www.kentuckyconnect.com/heraldlead...899/commentarydocs/ 618clark anti-Kera.htm.

Coleman, J. (1990). Families and schools. *Foundations of Social Theory*. Cambridge: Harvard University Press, 36.

Coleman, J. (1966). *Equality of educational opportunity*. Washington, DC: Office of Education.

Comer, J. P. (1998). *Waiting for a miracle*. New York: Plume Books.

Comer, J., & Haynes, N. (1991). Parent involvement in schools: An ecological approach. *The Elementary School Journal*, *91*, (3), 271-277.

Cornbleth, C., & Waugh, D. (1995). The great speckled bird: Multicultural politics and education policy making. New York: St. Martins Press.

David, J. L. (1995-96). The who, what, and why of site-based management. *Educational Leadership*, *53*, (4), 4-9.

Evans, R. (1999). The great accountability fallacy. http://www.edweek.org.ew/vol- 18/2lrevans.h18

Fraser, J. W. (1997). *Reading, writing, and justice: School reform as if democracy matters*. Albany: State University of New York Press.

Fullan, M. (1993). *Change forces: Probing the depths of educational reform*. New York: The Falmer Press.

Giannetti, Charlene C. , & Sagarese, Margaret M. (1998). "Turning parents from critics to allies." *Educational Leadership, 55*,(8) 40-42.

Gismondi, D. M. (1999). *Beyond the Bakesale*. Unpublished doctoral dissertation, The Catholic University of America, Washington, DC.

Golding, E. B., & Rallis, S. F. (1993). *Principals of Dynamic Schools: Taking charge of change*. Thousand Oaks, CA: Corwin Press, Inc.

Greenwood, G. E. , & Parkay, F. W. (1989). Case studies for teacher decision making. New York: Random House.

Guskey, T. R., & Peterson, K. D. (1995-96). The road to classroom change. *Educational Leadership*, *53*, (4), 10-14.

Hamel, J., Dufour, S., & Fortin, D. (1993). *Case study methods*. London: Sage Publications.

Harp, L. (June 9, 1999). "Landmark education ruling turns 10." The Courier-Journal. http://www.courier-journal.com/localnews/1999/9906/09//0609kera.html

Hoy, W. K., & Tarter, C.J. (1995). Administrators solving the problems of practice: Decision-making concepts, cases, and consequences. Boston: Allyn and Bacon.

Iannaconne, L. , & Lutz, F. W. (1970). *Politics, policy and power: The governance of public*

schools. Columbus, OH: Charles E. Merrill Publishing Co.

http://www.lrc.state.ky.us/RECARCH/96RS/BILLS/SCR094.htm

http://www.lrc.state.ky.us/RECARCH/92RS/BILLS/SCR078.htm

http://www.lrc.state.ky.us/RECARCH/94RS/BILLS/HB256.htm

http://www.lrc.state.ky.us/RECARCH/98RS/HC131.htm

http://www.lrc.state.ky.us/RECARCH/94RS/BILLS/SRC081.htm

http://www.lrc.state.ky.us/KRS/PAR71/007-00/410.PDF

Kantor, H., & Lowe, R. (1995). Class, race, and the emergence of federal education policy: From the new deal to the great society. *Educational Researcher, 24*, (3) 4-19, 21.

Kowalski, T. J. (1991). Case studies on educational administration. New York: Longman Publishing Group.

Kozol, J. (1991). *Savage inequalities: Children in America's schools (p. 54)*. New York: Crown.

Levine, D., Lowe, R., Peterson, B., & Tenorio, R. (Eds.). (1995). *Rethinking schools: An agenda for change*. New York: The New Press.

Lutz, F. W., & Merz, C. (1992). *The politics of school-community relations*. New York: Teachers College Press, 7.

McElhinney, Christie. (1999). ECE Receives Dewitt-Wallace Reader's Digest Grant to Help Place Quality Teachers in Hard-to-Staff Schools." *http://www.ecs.org/ecs/ecsweb.nsf*

Miller, J. G. (1997). African American males in the criminal justice system." *Phi Delta Kappan, 78*, (10), k1-k12.

Miller, L. S. (1995). An American imperative: Accelerating minority educational advancement. New Haven: Yale University Press.

Murphy, J., & Louis, K. S. (Eds.). (1994). Reshaping the principalship: Insights from transformational reform efforts. Thousand Oaks, CA: Corwin Press.

Orfield, G., & Yun, J. T. (1999). "Resegregation in American Schools," The Civil Rights Project, Harvard University. http://www.law.harvard.edu/civilrights/publications/resegre gation99.html

Patterson, J. L. (1993). Leadership for tomorrow's schools. Alexandria, VA: ASCD.

Peshkin, A. (1978). *Growing up American*. Chicago: The University of Chicago Press.

Peterson, K. D., & Valli, D. W. (1994). "Changes in School Governance and Principals' Roles: Changing jurisdictions, new power dynamics, and conflict in restructured schools." In J. Murphy & K. S. Louis (Eds.), *Reshaping the Principalship: Insights from transformational reform efforts (* 219-236). Thousand Oaks, CA: Corwin Press, Inc.

Salter, B., & Tapper, T. (1985). Power and policy in education: The case of independent schooling. London: The Falmer Press.

Sarason, S. B. (1997). *How schools might be governed and why*. New York: Teachers College.

Sarason, S. B. (1995). *Parental involvement and the political principle: Why the existing governance structure of schools should be abolished*. San Francisco: Jossey- Bass Publisher.

Sarason, S. B. (1990). *The Predictable Failure of School Reform*. San Francisco: Jossey-Bass.

Shirley, D. (1997). *Community Organizing for Urban School Reform*, quote p. 4.

Smith, L. M., Dwyer, D. C., Prunty, J. P., & Kleine, P. F. (1988). *Innovation and change in schooling: History, politics, and agency*. New York: The Falmer Press.

Smith, L. M., Prunty, J. P., Dwyer, D. C., & Kleine, P. F. (1987). *The fate of an innovative school*. New York: The Falmer Press.

Smith, L. M., Prunty, J. P., Dwyer, D. C. (1986). *Educational innovators: Then and now*. New York: The Falmer Press.

Smith, L. M., & Keith, P. F. (1971). *Anatomy of educational innovation: An organizational analysis of an elementary school*. New York: John Wiley & Sons, Inc.

Speck, M. (1999). *The principalship: Building a learning community*. Upper Saddle River, NJ: Prentice-Hall.

Spring, J. (1998). *Conflicts of interests: The politics of American education*. Boston: McGraw-Hill Companies, Inc.

Spring, J. (1997). *The American school 1642-1996*. New York: The McGraw-Hill Companies, Inc.

Spring, J. (1976). *The sorting machine: National educational policy since 1945*. New York: David McKay, Company, Inc.

Steele, C. M. (1992). "Race and the Schooling of Black Americans," *Atlantic Monthly, (April)*, 68-78.

Steele, S. (1990). *The content of our character: A new vision of race in America*. New York: St. Martin's Press.

Acknowledgment
This research was supported by the Edna McConnell Clark Foundation.

About the Author

Merylann J. Schuttloffel

Dr. Merylann J. "Mimi" Schuttloffel is Assistant Professor of Educational Administration and Policy Studies at the Catholic University of America. She has her Ph. D. from the University of Tulsa in Educational Administration and Research. Her research interests include reflective practice in leadership, innovation and school change, and the transformation of educational beliefs and practice. Comments may be sent to:
Email: schuttloffel@cua.edu

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina - Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois - UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 2273 times since September 17, 2000

Education Policy Analysis Archives

Volume 8 Number 46 September 17, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES.**
Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Should Achievement Tests be Used to Judge School Quality?

Scott C. Bauer
University of New Orleans

Abstract
This study provides empirical evidence to answer the question whether student scores on standardized achievement tests represent reasonable measures of instructional quality. Using a research protocol designed by Popham and the local study directors, individual test items from a nationally-marketed standardized achievement test were rated by educators and parents to determine the degree to which raters felt that the items reflect important content that is actually taught in schools, and the degree to which raters felt that students' answers to the questions would be likely to be unduly influenced by confounded causality. Three research questions are addressed: What percentage of test items are considered suspect by raters as indicators of school instructional quality? Do educators and parents of school-age children differ in their ratings of the appropriateness of test items? Do educators and parents feel that standardized achievement test scores should be used as an indicator of school instructional quality? Descriptive statistics show that on average, raters felt that the content reflected in test questions measured material that is important for students to know. However, for reading and language arts questions, between about 20% to 40% of the items were viewed as suspect in terms of the other criteria.

Introduction

Since publication of *A Nation at Risk* in 1983, issues associated with accountability have been at the forefront of educational reform in the United States. Kirst (1990) estimated that in the 1980's alone, 40 states created or amended their accountability systems. Stecher and Barron (1999) note that the number of states with a mandated student testing program rose from 29 in 1980 to 46 in 1992. Presidents Bush and Clinton both proposed the creation of a voluntary national test that would allow the reporting of student performance in relation to national standards (Carnevale & Kimmel, 1997).

The emergence of high-stakes accountability policies has intensified the debate over whether state-mandated assessment is a useful instrument for changing educational practice (Firestone, Mayrowetz, and Fairman, 1998; Ginsberg and Berry, 1998; Sheldon and Biddle, 1998). Proponents of high-stakes testing assume that poor performance in American schools results from a lack of attention to

school performance. "To solve such problems, according to this view, we need to set high standards for students, assess students' performance with standardized tests, and reward or punish students, their teachers, and their schools, depending on whether those standards are met" (Sheldon and Biddle, 1998, p. 165):

> Forty-nine states and a number of urban districts have set standards for what students should know and be able to do at various points in their school careers. Half the states hold schools accountable and apply sanctions to those whose students fail to meet the standards. At least a third – with more soon to follow – require students to score at designated levels on tests to get promoted and/or graduate. (Wolk, 1998, p. 48)

A recent survey by the Council of Chief State School Officers (1998) shows that while the states are increasingly introducing less traditional performance measures like portfolios into their assessment programs, 31 states use norm-referenced tests to measure student achievement in language arts, reading and mathematics. Tests are generally a part of the accountability system because they are inexpensive and quick to implement, and they are considered socially accepted as indicators of student performance (Linn, 1999).

At the heart of the debate over the use of high-stakes testing policies as a reform is the assumption that introducing new assessments will result in changes in teacher behavior in the classroom. As Firestone, Mayrowetz and Fairman (1998) observed, there is in fact a good deal of evidence that testing changes patterns of teaching, "if only by promoting 'teaching to the test'" (p. 96). There is evidence that school-based performance and reward programs such as Kentucky's produces desired results (Kelley and Protsik, 1997), and research supports the notion that school leaders take high-stakes testing very seriously (Mitchell, 1995). However, research also suggests that high- stakes testing programs do not necessarily provide valid data on students and schools (Stecher & Barron, 1999), and these systems tend to produce a high level of stress for teachers and principals. Critics argue that high-stakes testing may encourage teachers to consider test scores as ends in themselves:

> Evidence...reveals various perils associated with rigid standards, narrow accountability, and tangible sanctions that can debase the motivations and performances of teachers and students. Teachers faced with reforms that stress such practices may become controlling, unresponsive to individual students, and alienated. Test- and sanction-focused students may lose intrinsic interest in subject matter, learn at only a superficial level, and fail to develop a desire for future learning. (Sheldon and Biddle, 1998, p. 164)

Opponents of these measures conclude that they result in dumbing-down the curriculum (e.g., Corbett and Wilson, 1991), while others argue that they deny the reality of the situation faced by students, particularly those in urban districts, who are not well prepared to meet harsh standards (Wolk, 1998). Still others question whether policy is an effective instrument for shaping instructional practice at all (e.g., Cohen, 1995). Newmann, King and Rigdon argue that high-stakes accountability programs are doomed to failure because insufficient attention is paid to increasing schools' capacity for change, and Mayer (1998) raises the question of whether pursuing standards-based reform while leaving testing policy largely unchanged undermines reform. Wallace (2000, p. 66) concludes, "Provincial achievement exams create undue pressure on students, teachers, and schools. Even worse, the tests fail to assess what students will need to know in the next century."

Nevertheless, rating school performance based on the results of state testing programs has become an increasingly popular feature of state accountability programs (Watts, Gaines & Creech, 1998). The CCSSO survey referenced earlier indicates, in fact, that standardized achievement tests generally serve as summative indicators of elementary, middle, and high school performance, at least in part. For instance, in my home state of Louisiana, the new testing program is used to produce a school performance score that includes scores from the state's criterion-referenced test (60% of score), a nationally-marketed norm-referenced test (30% of score), and student attendance and dropout rates (10 percent of score). The school performance score will be used to establish 10-year goals, and schools will be held accountable for reaching two-year targets that represent progress toward these goals. A series of corrective actions are spelled out for schools that fail to meet their targets (Louisiana's School and District Accountability System, 1999).

At the 1998 Annual Meeting of the Mid-South Educational Research Association, W. James Popham raised the following question: Is it *appropriate* to use norm-referenced tests to evaluate instructional quality? Specifically, he challenged participants to consider whether norm-referenced tests measure knowledge that is taught and learned in schools. Popham then invited researchers to participate with him in a study to answer the question: Should student scores on standardized achievement tests be used to evaluate instructional quality in local schools?

In a subsequent paper, Popham (1999) laid out the basic argument that frames this study. While standardized achievement tests are useful tools to provide evidence about a specific students' mastery of knowledge and skills in certain content domains, "Employing standardized achievement tests to ascertain educational quality is like measuring temperature with a tablespoon" (p. 10). There are several difficulties with using aggregate measures from norm-referenced tests to judge the performance of a school. First, there is considerable diversity across states and school systems with regard to content standards, and therefore test developers produce "one-size-fits-all

assessments" which do not adequately align with what's supposed to be taught in schools. Second, because norm- referenced tests must provide a mechanism to differentiate between students based on a relatively small number of test items, test developers select "middle difficulty" items. As Popham put it,

> As a consequence of the quest for score variance in a standardized achievement test, items on which students perform well are often excluded. However, items on which students perform well often cover the content that, because of its importance, teachers stress. Thus the better the job that teachers do in teaching important knowledge and/or skills, the less likely it is that there will be items on a standardized achievement test measuring such knowledge and skills (p. 12).

Finally, scores on standardized achievement tests may not be attributable to the instructional quality of a school. Student performance may be caused by any number of factors, including what's taught in schools, a student's native intelligence, and out-of-school learning opportunities that are heavily influenced by a students' home environment. Popham terms this last issue the problem of "confounded causality."

Here we report the results of one of several local studies designed to provide empirical evidence to answer the question of whether student scores on standardized achievement tests represent reasonable measures of instructional quality. Using a research protocol designed by Popham and the local study directors, individual test items from a nationally-marketed standardized achievement test were rated by educators and parents to determine the degree to which raters felt that the items reflect important content that is actually taught in schools, and the degree to which raters felt that students' answers to the questions would be likely to be unduly influenced by confounded causality. Three research questions are addressed:

1. What percentage of test items are considered suspect by raters as indicators of school instructional quality?
2. Do educators and parents of school-age children differ in their ratings of the appropriateness of test items?
3. Do educators and parents feel that standardized achievement test scores should be used as an indicator of school instructional quality?

Methods

The investigation consisted of a series of three separate item-review studies designed to secure evidence regarding the appropriateness of using students' scores on standardized achievement tests as evidence of instructional quality. All sections of a nationally-marketed standardized achievement test was studied at the third grade level. The test covers mathematics, reading and language

arts content areas. The test was secured by the local study director, who also took responsibility for security.

Participants

Participants were solicited from two sources. First, principals associated with the School Leadership Center of Greater New Orleans (SLC-GNO) were invited to put together teams of teachers and parents to host an item-rating session. Two principals were able to put together groups of ten and eleven raters. From these 21 participants, 10 were parents and 11 were educators. These rating sessions were held at the participant's schools after school hours. Additionally, nine teachers enrolled in a graduate level course dealing with testing and measurement at the University of New Orleans formed a third group. This rating session was held on campus. In sum, then, 30 reviewers served as item raters, including two principals, 18 teachers, and 10 parents of elementary school children.

Procedures

Reviewers were provided with a description of the goals and procedures associated with the study prior to the actual rating session. In addition to signing a standard human subjects protocol outlining the responsibilities and risks associated with participation, reviewers signed a test- confidentiality form prior to their participation, and the item reviews were carried out under the scrutiny of the local director so that no security violations could occur. All test booklets were retained by the study director. Data were recorded on forms that do not reveal the specific test reviewed or any test questions.

Reviewers were asked to make their item-by-item judgments individually on summary rating sheets (see Exhibit 1 for a sample of the rating sheet), without group discussion, using a protocol that asked them to examine test items and judge their appropriateness in terms of five criteria:

1. IMPORT: Is the skill or knowledge measured by this item truly important for children to learn?
2. TAUGHT: Is the skill or knowledge measured by this item likely to be taught if teachers follow the prescribed curriculum?
3. SES: Is this item *free* of qualities (form or content) that will make the likelihood of a student's answering correctly be dominantly influenced by the student's socioeconomic status?
4. INHERITED CAPABILITIES: Is this item *free* of qualities (form or content) that will make the likelihood of a student's answering correctly be dominantly influenced by the student's inherited academic capabilities?
5. VALIDITY: Will a student's response to this item contribute to a valid inference about the student's status regarding whatever the test is supposed to be measuring?

Exhibit 1. Sample item rating sheet

Item	Import?	Taught?	SES?	IQ?	Validity?
1	Y ? N	Y ? N	Y ? N	Y ? N	Y ? N
2	Y ? N	Y ? N	Y ? N	Y ? N	Y ? N
3	Y ? N	Y ? N	Y ? N	Y ? N	Y ? N
4	Y ? N	Y ? N	Y ? N	Y ? N	Y ? N
5	Y ? N	Y ? N	Y ? N	Y ? N	Y ? N

Exhibit 2. The five item-review questions

1. IMPORT: Is the skill or knowledge measured by this item truly important for children to learn?

2. TAUGHT: Is the skill or knowledge measured by this item likely to be taught if teachers follow the prescribed curriculum?

3. SES: Is this item *free of qualities* (form or content) that will make the likelihood of a student's answering correctly be dependent on the student's socioeconomic status? **WOULD A STUDENT FROM A WELL-OFF HOME BE UNLIKELY TO GET THE ITEM CORRECT JUST BECAUSE HE OR SHE IS MORE "ADVANTAGED?"**

4. IQ: Is this item *free of qualities* (form or content) that will make the likelihood of a student's answering correctly be dependent on the student's inherited academic capabilities? **WOULD A STUDENT WITH GREATER NATIVE INTELLIGENCE (IQ) BE UNLIKELY TO GET THE ITEM CORRECT JUST BECAUSE OF THIS INBORN QUALITY?**

5. VALIDITY: Will a student's response to this item contribute to a valid conclusion about the student's ability relating to whatever the test is supposed to be measuring? **IS THIS ITEM A VALID MEASURE OF THE ABILITY THE TEST IS MEASURING IN THIS SECTION OF THE TEST?**

During an orientation phase, prior to item-review, the local study director practiced reviewing a selection of test items from a test-booklet's sample items and/or from a different test to clarify item-reviewers' understanding of the five item-review questions. During a pre-test of the procedure, it became clear that respondents may have difficulty with the questions related to SES, IQ, and validity, thus some clarifying language was added and a summary sheet was provided to raters which allowed them to access the definitions as they

performed the ratings. (Exhibit 2 shows the summary sheet.)

Each rating session was held in the afternoon, and took approximately three hours. Because of the time of day and the considerable investment of time and energy, participants were provided with a light dinner after each rating session. They also participated in a short debriefing session, during which they answered questions about the methodology and their ability to sensibly rate the test items.

Analysis

Response sheets were collected and numbered after each session. The number of items rated yes, no, or with a question mark (not sure) were tallied for each content area of the test, and the number of no and "not sure" (question mark) ratings were entered into an SPSS 9.0 for Windows system file. To address the question of what percentage of test items raters considered suspect as indicators of school instructional quality, the mean percentages of items rated "no" or "not sure" were computed for each of the rating criteria and for each content area of the test. Descriptive statistics related to the raters' judgments of items in each content area of the test and for each of the criteria are presented. Additionally, a summary statistic indicating the mean percentage of items rated as suspect on at least one criterion was computed. For purposes of discussion, the percentage of items rated as either a "no" or "not sure" are combined; given the high-stakes involved in the state accountability programs, if raters cannot determine if an item meets the criteria used in this study, we will consider it suspect. The full breakdown of ratings are presented in the Appendices.

To see if educators and parents of school-age children differ in their ratings of the appropriateness of test items, analysis of variance was computed to test whether the mean ratings are statistically significant. Eta-squared is also reported; Stevens (1996) recommends that to interpret the effect size, an eta-squared of .01 should be treated as a small effect, .06 a medium effect, and .14 a large effect.

To address whether educators and parents feel that standardized achievement test scores should be used as an indicator of school instructional quality, the frequency distribution is reported for a summary question which asked respondents to answer yes, no, or "not sure" in regard to this question. Chi-square was computed to see if there is a statistically significant association between the answer to this summary question and group membership.

As a final portion of the study, answers to questions posed during debriefing sessions were analyzed to determine whether raters felt confident in their ability to assess test items on these criteria. In an exploratory study such as this, rater's sense of their ability to render reliable judgments in terms of these criteria is an important question. These data may shed some light on whether the methodology provides a valid assessment of the usefulness of the test to judge school quality.

Results

Table 1 displays the mean percentage of test items rated as suspect by respondents. As mentioned earlier, the percentage reflects the number of items rated as either a "no" or "not sure" on each of the five criteria for each content area of the test. Overall, the mean percentage of items rated as suspect varies widely; only 2% of the items were rated as suspect in importance for math procedures, whereas 41% of the vocabulary items were rated as suspect because the likelihood seemed great that student's answering correctly would be dependent on the student's inherited academic capabilities (IQ). Overall, raters felt that the items dealing with reading and language arts were more often suspect as indicators of school quality, especially in terms of the likelihood that students' answering these items correctly would be unduly influenced by native intelligence (IQ) or socio-economic status (SES). Raters were somewhat more comfortable with measures relating to mathematics problem-solving and reasoning, and considerably more comfortable with the items measuring mathematics procedures.

Table 1
Mean percentage of items rated as "suspect" for each content area

Content Area	Important?	Taught?	SES?	IQ?	Valid?
Vocabulary	11%	26%	38%	41%	26%
Reading comprehension	14	26	38	40	26
Grammar & language	8	24	37	38	21
Math problem solving & reasoning	11	24	19	33	21
Math procedures	2	7	11	22	10

Viewing the data in Table 1 in terms of criteria instead of content area, one sees that from among the various criteria used to rate test items, raters judged the test items more likely to be suspect in terms of SES and IQ. That is, from among the five possible reasons a test item might be inappropriate to assess school quality, raters felt the greatest threat to validity was the likelihood that a student might answer the item correctly because of socio- economic advantage or because of native intelligence rather than because of what he or she learned in school. In fact, for the reading and language arts content areas, between 30 and 40% of the items were rated as suspect in these regards. Considerably fewer items were rated as suspect because they were deemed unimportant for students to know, and for most content

areas between 20 and 30% of the items were deemed unacceptable because raters felt that the material was not a part of the standard curriculum at that grade level.

The above-mentioned data show the mean percentage of items rated as suspect on each of the five criteria; a final summary statistic was computed to show the mean percentage of items in each section of the test that was rated as suspect on *at least one* of the five criteria. Table 2 shows that for all areas of the test, approximately 50% of the items were deemed inappropriate as indicators of instructional quality on at least one criterion. The table also shows that the range of ratings is considerable – for most areas, at least one rater felt that nearly all of the items were alright as indicators of instructional quality on all criteria, and at least one rater felt that all items were suspect on at least one of the five criteria.

Table 2
Mean percentage of items deemed suspect on *at least* one criterion

Content area	Mean %	High	Low
Vocabulary	57%	100%	15%
Reading comprehension	52%	100%	3%
Grammar & language	55%	100%	13%
Math problem solving & reasoning	48%	100%	3%
Math procedures	46%	100%	0%

To address the question of whether educators and parents rated the test items differently, analyses of variance were computed to test the null hypothesis that the mean percentages do not differ between the two groups of respondents. These data, presented in Table 3, show that the only statistically significant differences between the mean percentage of items rated as suspect by parents and educators exist for the criteria dealing with whether the content measured by the test item is taught in the regular school curriculum (taught). Parents consistently felt that a greater percentage of the items on the test covered material that would not be a part of the standard curriculum. An examination of eta-squared shows that for most of the content areas, the effect size of the difference in means for this criterion (taught) is large (eta^2 for vocabulary=.16, for reading comprehension=.16, for math problem-solving=.19) or moderate (eta^2 for grammar and language=.10, for math procedures=.11).

Table 3
Mean ratings by respondent group

Vocabulary

Respondent	IMP	TAU	SES	IQ	VAL
Educator	.09	.20	.43	.44	.21
Parent	.14	.39	.29	.37	.35
F (1,28)	.72	5.40*	1.88	.47	2.89
Eta-squared	.03	.16	.06	.02	.09

Reading comprehension

Educator	.11	.20	.43	.38	.21
Parent	.20	.39	.29	.46	.35
F (1, 28)	1.99	5.40*	1.88	.50	2.89
Eta-squared	.07	.16	.06	.02	.09

Grammar and language

Educator	.07	.18	.38	.38	.22
Parent	.10	.35	.35	.37	.20
F (1, 28)	.69	2.95	.04	.02	.08
Eta-squared	.02	.10	.01	.00	.00

Math problem-solving & reasoning

Educator	.11	.17	.19	.35	.18
Parent	.10	.37	.17	.29	.25
F (1, 28)	.01	6.36*	.04	.25	1.08
Eta-squared	.00	.19	.00	.01	.04

Math procedures

Educator	.02	.05	.08	.24	.12
Parent	.01	.12	.16	.19	.07
F (1, 28)	.00	3.39	.59	.11	.85
Eta-squared	.00	.11	.02	.00	.03

* $p<.05$

Table 4 shows the results for the summary item that asked raters to judge whether they would recommend using standardized achievement tests as an indicator of instructional quality. Results show that approximately a quarter of the educators and 30% of the parents felt that standardized achievement tests ought to be used as an indicator of school quality, whereas about two-thirds of the educators

and 40% of the parents felt that they should not. Another 30% of the parents and 11% of the educators were not sure, and one respondent left the question blank. The chi-square test of association showed that there is not a statistically significant association between the answer to this question and role [X^2 (2, n=29) = 2.11, $p<.05$].

Table 4
Should standardized tests be used to measure instructional quality?

	Yes	Not sure	No
Educator	5 (26%)	2 (11%)	12 (64%)
Parent	3 (30%)	3 (30%)	4 (40%)
Total	8 (28%)	5 (17%)	16 (55%)

The final data collected in this study had to do with the methodology itself. A formal debriefing was held after each item rating session. Respondents were asked a short series of questions in writing about their ability to rate test items and about the kinds of factors they felt influenced their ratings. Raters also discussed their experiences and any difficulties they perceived with the rating process. These data provide us with some sense of the threats to validity present in the ratings.

Respondents were asked to rate how easy they felt it was to make judgments about the test items, on a scale of 1 = "very easy" to 10 = "very difficult." On average, these data show that respondents felt that it was relatively easy to assess whether an item measured import material for students to know (2.1) and whether the item was likely to be taught as a part of the regular curriculum (2.9). Raters found it most difficult to rate whether an item would be more likely to be answered correctly because of a child's inherited capabilities (IQ) or socio-economic status (5.0 and 4.5, respectively). Respondents also found it relatively more difficult to judge whether an item was a valid measure of the skill it was intended to measure (4.7). Overall, then, on a ten-point scale raters found their job moderately easy (i.e., lower than the midpoint between very easy and very difficult), though some criteria were more difficult to apply than others.

Respondents also answered open-ended questions that probed into the kinds of factors that they felt might threaten their ability to render reliable judgments about the test items. These answers show that most of the parents felt at least a bit unsure about what was in the regular or "official" curriculum, thus they were not sure about the reliability of their judgments on the criterion labeled "taught." One respondent pointed out that SES and IQ were tough to assess because these relate to a subjective assessment of the fairness of an item, and several other respondents noted that SES was likely influenced by their own socio-economic status. That is, they questioned whether relatively

well-off parents or teachers could render a valid judgment on this criterion. Some teachers questioned whether their beliefs about teaching would "get in the way" of their ability to rate the items, and several raters simply said that they found it tough – "speculative" – to assess the degree to which a students' answer on a test item would relate more to native intelligence than knowledge gained in school.

Summary and Conclusions

The purpose of this study was to attempt to amass credible evidence concerning whether student scores on standardized achievement tests should be used to evaluate instructional quality in local schools. Using a framework developed by Popham (1999) and a research protocol collaboratively devised by Popham and local study directors, educators and parents of school-age children rated all items contained on a commercially-marketed standardized achievement test that covered third grade content in reading, language arts, and mathematics. Descriptive statistics show that on average, raters felt that the content reflected in test questions measured material that is important for students to know. However, for reading and language arts questions, between about 20% to 40% of the items were viewed as suspect in terms of the other criteria. Raters saw fewer problems with questions dealing with mathematics problem-solving and reasoning, and they felt the fewest problems existed with questions on mathematical procedures. Overall, though, raters felt that about half of all items they appraised were suspect on at least one of the criteria used to assess the test. Educators and parents did not differ statistically on their ratings on most criteria, though about two-thirds of the educators felt that tests should not be used to judge instructional quality whereas only 40% of the parents felt this way. The range of ratings across respondents was considerable for all content areas and for each of the rating criteria; some respondents saw very few problems with any questions, while others felt that the vast majority of items were suspect on at least one criterion.

This study was prompted by the realization that while standardized achievement tests are useful tools to provide evidence about students' mastery of knowledge and skills in tested content domains, it does not logically follow that they should be useful as indicators of school performance. As reflected in the rating scheme used in this study, student performance on standardized tests may be caused by any number of factors, including what's taught in schools, a student's native intelligence, and out-of-school learning opportunities that are heavily influenced by a students' home environment.

The question that follows, then, is whether this confounded causality poses a problem in terms of using standardized test scores as measures of instructional quality. In a critique of Popham's argument regarding confounded causality, Schmoker (2000) argues that it does not. What happens in classrooms can "significantly mitigate and even overcome environmental and genetic factors" (p. 64), and standardized tests give schools focus and empower teachers by providing specific data on students' needs. "Standardized test results have provided the

essential focus and urgency for schools to improve and refine instructional programs in reading, writing, and math practices" (p. 64).

This argument misses the point. There is no question that norm-referenced tests are exceedingly valuable in their intended purpose: to identify knowledge and skills that individual students need to improve, thus providing professional educators with essential data with which they can craft programs and practices. It does not follow, however, that using aggregate average scores on standardized tests serves as a good indicator of *school quality*. To say that norm-referenced tests can help teachers identify areas in need of attention does not rely on an assumption that school programs alone caused a deficiency; instead, as Schmoker observed, this relied on the belief that schools can do something to overcome the deficiency regardless of cause.

The notion that aggregate scores on standardized tests should serve as an indicator of school quality relies on an assumption of causality. The underlying logic is that the scores are *predominantly* caused by something the school does or has some control over. For this assumption to hold, at a minimum we must be willing to believe that student performance on standardized tests is related to school quality, that the tests measure the skills and abilities stressed in school programs, and that there are no antecedent factors that might otherwise explain aggregate student performance on the tests. If the data presented here are credible, the soundness of this assumption must be questioned. On average about half of the items on the rated test suffer from "confounded causality" on at least one of these criteria.

The question of whether the data presented here are, in fact, "credible," deserves attention. The data collected from debriefing presented earlier barely scratch the surface of the potential threats to validity. Perhaps the biggest issues stem from the fact that the study was purposefully constructed to include both educators and parents. The fact that parents felt less knowledgeable about what *should be* in the regular school curriculum may have resulted in an exaggeration of the percentage of items that were deemed suspect on this criterion. Additionally, some respondents felt it difficult to judge whether items might be unduly influenced by a students' native intelligence (where do you draw the line between native intelligence and knowledge learned in school?) and some felt that their own social standing made it hard for them to determine if a students' socio-economic background would greatly influence the likelihood of answering a test item correctly.

Regardless of criterion, the rating process asked for a judgment, that is, the subjective assessment of an item's appropriateness. These are difficult conclusions to make. Yet, in terms of the message to policy-makers, this is precisely be the point. Aggregate average scores on standardized tests are at best a gross approximation of the instructional quality of a school, and any number of factors may have more to do with the production of this number than the quality of educational services delivered. We should be questioning what these numbers mean, especially considering the fact that in many states the numbers are being used to reward or punish school staff and students.

By design, policy makers have raised the stakes. As this analysis shows, though, when you get beneath the summary number and ask whether the test items that go into producing that number are sensible measures of knowledge and skills learned in school, the answer is far from clear. This would suggest, at a minimum, that policy-makers should consider eliminating or de-emphasizing their use of norm-referenced achievement tests as a barometer of how well a school is doing.

Note

Research presented here was supported by a grant from the School Leadership Center of Greater New Orleans. An earlier version of this work was presented at the Annual Meeting of the Mid-South Educational Research Association, Point Clear, AL, November 1999.

References

Carnevale, A. & Kimmel, E. (1997). *A national test: Balancing policy and technical issues*. Princeton, NJ: Educational Testing Service.

Cohen, D. (1995). What is the system in systemic reform? *Educational researcher, 24*, 11-17.

Corbett, H. and Wilson, B. (1991). *Testing, reform and rebellion.* Norwood, NJ: Ablex.

Council of Chief State School Officers. (1998). Key state education policies on K-12 education: Standards, graduation, assessment, teacher licensure, time and attendance - a 50- state report. Washington, DC: Author.

Firestone, W., Mayrowetz, D., and Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational evaluation and policy analysis, 20*, 95-113.

Ginsberg, R. and Berry, B. (1998). The capability for enhancing accountability. In R. MacPherson (Ed.), *The politics of accountability: Educative and international perspectives* (pp. 43-61). Newbury Park, CA: Corwin.

Kelley, C. and Protsik, C. (1997). Risk and reward: Perspectives on the implementation of Kentucky's school- based performance award program. *Educational Administration Quarterly, 33*. 474-505.

Kirst, M. (1990). Accountability: Implications for state and local policymakers. Washington, DC: OERI.

Linn, R. (1999). Standards-based accountability: Ten suggestions

(CRESST Policy Brief). Los Angeles: National Center for Research on Evaluation, Standards and Student Testing.

Mayer, D. (1998). Do new teaching standards undermine performance on old tests? *Educational evaluation and policy analysis, 20,* 53-74.

Mitchell, K. (1995). Reforming and conforming: NASDC principals talk about the impact of accountability systems on school reform (Technical Report 143). Santa Monica, CA: Rand Corp.

Popham, W. J. (1999, March). Why standardized tests don't measure educational quality. *Educational Leadership, 57,* 8-15.

Schmoker, M. (2000, February). The results we want. *Educational Leadership, 58*, 62-65.

Sheldon K. and Biddle, B. (1998). Standards, accountability and school reform: Perils and pitfalls. *Teachers' College Record 100 (1)*: 164-180.

Stecher, B. & Barron, S. (1999). Test-Based Accountability: The perverse consequences of milepost testing. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada, April 1999.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Wallace, D. (2000, February). Results, results, results? *Educational Leadership, 58*, 66-68.

Watts, J., Gaines, G. & Creech, J. (1998). Getting results: A fresh look at school accountability. Atlanta: Southern Regional Education Board.

Wolk, R. (1998). Education's high-stakes gamble. *Education Week, 18 (15)*: 48.

About the Author

Scott C. Bauer
348J Bicentennial Education Bldg.
University of New Orleans
New Orleans, LA 70148
504-280-6446 (voice)
504-280-6453 (fax)
Email: sbauer@uno.edu

Scott C. Bauer (Ph.D., Cornell University) is an assistant professor in the Educational Leadership, Counseling, and Foundations Department in the College of Education at the University of New Orleans, and

Director of Research at the School Leadership Center of Greater New Orleans. His research and teaching focuses on the application of the principles of organizational behavior and development to the study of school leadership, organizational change and restructuring. Dr. Bauer's most recent work deals with designing and implementing site-based decision making systems in schools.

Appendix A
Descriptive statistics: mean percentages, standard deviations and range of all ratings

Skill area	Criteria	Rating	Mean	sd	high	low
Vocabulary	importance	not sure	.07	.11	.40	0
		no	.04	.06	.20	0
	taught	not sure	.21	.22	1.00	0
		no	.06	.09	.40	0
	SES	not sure	.14	.15	.75	0
		no	.24	.23	1.00	0
	IQ	not sure	.15	.12	.35	0
		no	.27	.24	1.00	0
	Validity	not sure	.12	.14	.45	0
		no	.14	.15	.75	0
Skill area	Criteria	Rating	Mean	sd	high	low
Reading comprehension	importance	not sure	.07	.12	.53	0
		no	.07	.10	.37	0
	taught	not sure	.16	.23	1.00	0
		no	.07	.11	.30	0
	SES	not sure	.08	.09	.30	0
		no	.20	.25	.93	0
	IQ	not sure	.13	.19	.93	0
		no	.28	.27	.97	0
	Validity	not sure	.12	.13	.57	0
		no	.15	.17	.77	0
Skill area	Criteria	Rating	Mean	sd	high	low
Grammar and Language	importance	not sure	.05	.07	.23	0
		no	.03	.05	.23	0
	taught	not sure	.17	.21	1.00	0
		no	.07	.16	.77	0
	SES	not sure	.13	.14	.57	0
		no	.24	.27	1.00	0

	IQ	not sure	.12	.15	.47	0
		no	.26	.32	1.00	0
	Validity	not sure	.10	.11	.40	0
		no	.12	.11	.40	0
Skill area	Criteria	Rating	Mean	sd	high	low
Math problem-solving and reasoning	importance	not sure	.06	.11	.50	0
		no	.05	.06	.27	0
	taught	not sure	.17	.22	1.00	0
		no	.07	.14	.57	0
	SES	not sure	.05	.07	.27	0
		no	.13	.25	.97	0
	IQ	not sure	.07	.08	.33	0
		no	.27	.30	.97	0
	Validity	not sure	.10	.10	.30	0
		no	.11	.14	.67	0
Skill area	Criteria	Rating	Mean	sd	high	low
Math Procedures	importance	not sure	.01	.03	.15	0
		no	.01	.03	.15	0
	taught	not sure	.05	.10	.40	0
		no	.02	.04	.15	0
	SES	not sure	.01	.02	.05	0
		no	.10	.27	1.00	0
	IQ	not sure	.03	.05	.20	0
		no	.20	.36	1.00	0
	Validity	not sure	.03	.05	.15	0
		no	.08	.15	.50	0

Appendix B
Descriptive statistics: mean percentages, standard deviations and range of combined ratings

	Criteria	Mean	sd	high	low
Vocabulary	importance	.11	.14	.50	0
	taught	.26	.23	1.00	0
	SES	.38	.27	1.00	0
	IQ	.41	.24	1.00	0
	validity	.26	.20	.80	0
Reading comprehension	importance	.14	.16	.57	0
	taught	.26	.23	1.00	0
	SES	.38	.27	1.00	0
	IQ	.40	.30	.97	0
	validity	.26	.20	.80	0
Grammar and Language	importance	.08	.10	.33	0
	taught	.24	.25	1.00	0
	SES	.37	.27	1.00	0
	IQ	.38	.32	1.00	0
	validity	.21	.18	.77	0
Math problem-solving and reasoning	importance	.11	.12	.50	0
	taught	.24	.22	1.00	0
	SES	.19	.24	.97	0
	IQ	.33	.30	1.00	0
	validity	.21	.17	.80	0
Math Procedures	importance	.02	.05	.20	0
	taught	.07	.10	.40	0
	SES	.11	.27	1.00	0
	IQ	.22	.38	1.00	0
	validity	.10	.16	.60	0

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
New York University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
Ann Leavenworth Center for Accelerated Learning

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois—UC

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 1506 times since September 18, 2000

Education Policy Analysis Archives

Volume 8 Number 47 **September 18, 2000** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Teacher Supply and Demand: Surprises from Primary Research

Andrew J. Wayne
University of Maryland

Abstract
An investigation of primary research studies on public school teacher supply and demand revealed four surprises. Projections show that enrollments are leveling off. Relatedly, annual hiring increases should be only about two or three percent over the next few years. Results from studies of teacher attrition also yield unexpected results. Excluding retirements, only about one in 20 teachers leaves each year, and the novice teachers who quit mainly cite personal and family reasons, not job dissatisfaction. Each of these findings broadens policy makers' options for teacher supply.

With teacher quality atop local, state, and federal agendas, the body of policy research that addresses teacher quality is very much in the spotlight. Hopefully some of the knowledge generated by researchers can prove helpful to policy makers.

But to a surprising extent, the research community is not offering policy makers much that they can use. The policy researchers who shape public understanding of the teacher quality issue are now making considerable efforts to challenge each other's work (e.g., Ballou and Podgursky, 1999, 2000; Darling-Hammond, 1999). Although that debate will have salutary effects over the long-term, the short-term outlook for lay audiences is confusion over whom to trust.

This article attempts to make progress by focusing on questions whose answers depend on more broadly understood analytic tools. It focuses on teacher supply and demand—only one part of the teacher quality story. But knowledge about supply and demand can help policy makers, and the requisite analytic tools are so simple that disagreement is unlikely.

My examination of the knowledge base on the supply and demand of public school teachers led to several surprises. Rather than summarize all that is known, what follows focuses on those points where the common wisdom is wrong. Each of the four sections below contrasts what primary research studies say with what policy makers hear about supply and demand.

The original studies come from long-term federal investments in survey research, overseen by the National Center for Education Statistics (NCES). The NCES is regarded as the most authoritative source of national evidence on teacher supply and demand. Its survey methods and analyses are thoroughly documented, and all of its documents are publicly available at www.nces.ed.gov.

Enrollments are Leveling Off

Close examination of NCES projections reveals that enrollments are leveling off. Mischaracterizations of these projections are very common. A recent RAND publication referred to "a dramatic increase in enrollments" over the next decade (Kirby, Naftel, and Berends, 1999, p. 1). Combined with teacher retirements, says a U.S. Department of Education document, these enrollment increases spell a "demographic double-whammy" for the schools (U.S. Department of Education, 1998, p. 2).

The NCES counts students every year. Actually school districts do the counting and report their findings to state governments who, in turn, report numbers to the NCES. The error-checking and compilation process is somewhat time-consuming, so the most recently reported count was for 1998.

Those counts show that from 1988 to 1998 enrollments rose 16 percent. Contrast that with what the future holds. According to NCES's analyses, from 2000 to 2005 enrollments should rise only one percent, and from 2005 to 2010 enrollment should decline, though perhaps negligibly. Census Bureau population projections undergird these

estimates (Gerald and Hussar, 2000, p. 12).

In other words, the best available projection is that a school with 1000 students today will have about 1010 students *five years from now*. The Census Bureau can botch immigration assumptions (Ahlburg, 1993), and, to be sure, national averages are no guide for state policy makers. From 1990 to 1996, for example, elementary enrollment dropped about six percent in West Virginia and North Dakota, while it increased about 15 percent in California and New Jersey (Gerald and Hussar, 1998, p. 109). But if policy makers expected a wave of children to deluge the nation's schools, they were misled.

Keeping enrollment increases in perspective this way helps policy makers understand their options. If the projections are roughly correct, the teaching force will hardly need to grow at all. The only growth will derive from declines in pupil-teacher ratios.

Hiring Will Increase, On Average, Two Percent Per Year Over the Next Decade

It is true that a wave of retirements is about to hit (Hussar, 1999, p. 10). Policy makers are hearing that these retirements—combined with already high attrition rates—will drive hiring needs through the roof. How big is the crunch?

For some reason journalists, academics, policy wonks, and interest groups offer only an ambiguous answer: the nation will need to hire 2.2 million public school teachers over the next decade. This ten year total—admittedly from NCES analyses—does nothing to help policy makers gauge the problem; they would need to know the number hired in the past decade for comparison. In most contexts the figure just imparts urgency or draws attention to someone's proposal. Ironically, a closer read of the NCES projections would permit an even more captivating ten year total—2.5 million—given predictable drops in the pupil-teacher ratio (Hussar, 1999, p. 35).

A much more helpful characterization of hiring needs is possible. The 2.5 million figure is actually the sum of all *annual hiring* for the next ten years. NCES projection models predict that annual hiring will rise from 218,000 in 1999-2000 to 261,000 in 2009-10. During that period, the early increases will somewhat outpace the later ones (Hussar, 1999, p. 35). Thoughtfully developed assumptions about enrollments, pupil-teacher ratio changes, and teacher attrition drive the projections, but no one would be surprised if the estimates proved wrong by 15,000 hires in either direction.

Because no one explains NCES projections in terms of annual hiring, policy makers' informants routinely slip up. A prominent foundation referred to "the projected shortage of 2.2 million teachers" (Milken Family Foundation, 1999). The more common misinterpretation is that the nation's teacher preparation institutions must *train* over two million teachers. Not so. At last count, experienced teachers constituted over one quarter of annual hiring (Hussar, 1999, p. 7).

What should the research community tell policy makers? Projections lose accuracy quickly with time, so our message ought to be that the next few years probably hold annual hiring increases of two to three percent. That is about all we can say, for our guesses about how hard the additional hiring will be are probably no better than policy makers'.

Excluding Retirements, About One in 20 Teachers Leaves Each Year

With all the hyperbole, a reasonable legislator might guess that one in four teachers drops out of the profession every year. The hallmark of the teaching profession, they are told, is the "revolving door." John Merrow—a prominent and respected education journalist—recently analogized it to "a swimming pool with a serious leak" and drew the conclusion for policy makers: "We're misdiagnosing the problem as 'recruitment' when its really 'retention'." (Merrow, 1999)

The actual data provide a different perspective. The NCES followed a national sample of over 4,500 teachers from the 1993-94 school year. Only about seven percent of them were not teachers in the 1994-95 school year, and two of the seven percent were retirees (Henke et al., 1997, p. A-248). That means that excluding retirements, only about one in 20 teachers leaves each year. And many of these people will return to teaching.

Where the same vivid metaphors are applied to beginning teachers, they still leave the wrong impression. Attrition among teachers with less than four years experience is about nine percent per year (Henke et al., 1997, p. A-248). Admittedly, this adds up. Multiply by four, and it appears that over one-third of a beginning cohort will not begin a fifth year. But does this distinguish teaching from other professions? A recent Public Agenda Survey found the opposite to be true. Only 19 percent of beginning teachers reported expecting to change careers, while fully half of college graduates under 30 years of age made the same claim (Farkas, Johnson, and Foleno, 2000, p. 11).

Even low-income schools within urban areas exhibit manageable overall attrition rates: 5.7 percent according to the best tabulation of NCES data (Ingersoll, 1999, p. 22). This figure raises serious questions about the assumptions that currently guide efforts to improve teacher quality for low-income students.

It helps to distinguish between teacher attrition and teacher mobility. The discussion above focused on the former, but just as many —if not more—teachers change schools every year as leave them. Add in teachers who change assignments, and over one in four teachers changes status somehow every year (Boe et al., 1998, p. 10). Needless to say, conflating these phenomena would not help decision-makers address supply and demand.

Novice Teachers Who Quit Rarely Cite Job Dissatisfaction

Evidence notwithstanding, many prefer to assume that novice teachers 'leave in droves' and offer explanations. The Director of the National Foundation for the Improvement of Education recently did so: "Why do they drop out? It's mainly because nobody's taking care of them" (Marklein, 1999, p. 6). Another explanation policy makers hear is that "substandard training fails to prepare teachers for the demands of the classroom" (Merrow, 1999).

Via confidential surveys, the NCES asked teachers who left what the main reason was for their departure. Among departing teachers with less than four years experience, 17 percent left involuntarily, mostly due to staffing actions. Another 12 percent left to take courses. Only eight percent marked "dissatisfied with teaching as a career," though another 17 percent left mainly "to pursue other work or better salary" (Boe et al., 1998, p. 32).

The missing group: 44 percent of the beginning teachers who left cited personal and family reasons (Boe et al., 1998, p. 32; see also Henke et al., 1997, p. A-255). It's possible that many enter teacher education programs precisely because the profession allows for commitment to family responsibilities. Summer work is definitely optional, and recruiters do not frown on long periods of unemployment.

So if the teaching profession "eats its young," it eats only a few. Doing the math above, dissatisfaction and competing careers explain on the order of only one quarter of novices' departures.

Figures like these give real perspective on the policy options for teacher supply. They debunk the exaggerations policy makers currently hear, that attrition among novices is and will remain unbearably high *until* (1) schools become more supportive working environments or (2) universities prepare teachers for real classrooms. No doubt those factors matter, but the real numbers show state and federal policy makers that substantial leverage is possible via the blunt instruments before them. Perhaps a twelve-month calendar—and concomitant salary increases—would draw the mainstream labor market into schools. Given good information, we know not to ignore such options.

Conclusion

My investigation of primary research studies on public school teacher supply and demand revealed four major surprises. Basic survey research and demography contradict what many say about enrollments, hiring needs, attrition, and the loss of novice teachers. If my interpretations are not correct, hopefully the research community will arrive at better answers reasonably quickly.

Readers should beware that although the discussion above employed the best available evidence, much of it relied on a national survey last conducted in the 1993-94 school year. State level investigations may turn up different results. Furthermore the 2000 Census and a new NCES survey of the nation's teachers are both underway and may yield important course corrections.

But the contrast—between what our primary research studies say and what policy makers hear—imparts a lasting message for the researchers and analysts concerned with teacher quality. What inhibits policy makers' utilization of the research base on teacher supply and demand is not lack of research, nor is it disagreements whose resolution requires more technical sophistication than policy makers have. Instead, the problem is neglect. When distortions arise, whether by mistake or because of interest group politics, it is the research community that is supposed to correct them.

Acknowledgements

I wish to thank Willis Hawley, William Hussar, Daniel Laitsch, and Linda Valli for their feedback on earlier drafts of this paper. The author accepts sole responsibility for any errors.

References

Ahlburg, D. (1993). The Census Bureau's new projections of the U.S. population. *Population and Development Review, 19*(1), 159-174.

Ballou, D., & Podgursky, M. (2000, May). Reforming teacher preparation and licensing: Continuing the debate. *Teachers College Record*. http://www.tcrecord.org ID: 10524.

Ballou, D., & Podgursky, M. (1999, October 13). Reforming teacher preparation and licensing: What is the evidence? *Teachers College Record*. http://www.tcrecord.org ID: 10418.

Boe, E. E., Bobbit, S. A., Cook, L. H., Barkanic, G., & Maisling, G. (1998). Teacher turnover in eight cognate areas: National trends and predictors. Philadelphia, PA: Center for Research and Evaluation in Social Policy, University of Pennsylvania.

Darling-Hammond, L. (1999, October 13). Reforming teacher preparation and licensing: Debating the evidence. *Teachers College Record*. http://www.tcrecord.org ID 10419.

Farkas, S., Johnson, J., & Foleno, T. (2000). *A sense of calling: Who teaches and why*. New York, NY: Public Agenda.

Gerald, D. E., & Hussar, W. J. (1998). Projections of education statistics to 2008. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Gerald, D. E., & Hussar, W. J. (2000). Projections of education statistics to 2010. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Henke, R. R., Choy, S. P., Chen, X., Geis, S., & Alt, M. N. (1997).

America's teachers: Profile of a profession, 1993-94. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Hussar, W. J. (1999). Predicting the need for newly hired teachers in the United States to 2008-09. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Ingersoll, R. M. (1999). Teacher turnover, teacher shortages, and the organization of schools. Seattle, WA: Center for the Study of Teaching and Policy, University of Washington.

Kirby, S. N., Naftel, S., & Berends, M. (1999). Staffing at-risk school districts in Texas: Problems and prospects. Santa Monica, CA: RAND.

Marklein, M. B. (1999, October 13). Teaching of teachers evolves to raise standards. *USA Today*, p. D6.

Merrow, J. (1999, October 6). The teacher shortage: Wrong diagnosis, phony cures. *Education Week*.

Milken Family Foundation. (1999, June 18). Poll reveals high school students saying "no" to teaching careers. Retrieved October 15, 1999 from the World Wide Web: http://www.mff.org/newsstory.taf?page=72.

U.S. Department of Education. (1998). *Promising practices: New ways to improve teacher quality*. Washington, DC: Author.

About the Author

Andrew Wayne

Andrew Wayne is a doctoral candidate in social policy at the University of Maryland, College Park. His forthcoming dissertation is entitled *Federal policies to improve teacher quality for low-income students*. He can be reached at (301) 657-3715 or by Email: aw102@umail.umd.edu.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Greg Camilli
Rutgers University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher
Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doc.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 1020 times since October 1, 2000

Education Policy Analysis Archives

Volume 8 Number 48 **October 1, 2000** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

"Put Teaching on the Same Footing as Research?" Teaching and Learning Policy Review in Hong Kong and the U.S.

Orlan Lee
Hong Kong University of Science & Technology

Abstract
The Research Assessment Exercises (RAEs) in hugely expanded universities in Britain and Hong Kong attempt mammoth scale ratings of "quality of research." If peer review on that scale is feasible for "quality of research," is it less so for "quality of teaching"? The lessons of the Hong Kong Teaching and Learning Quality Process Reviews (TLQPRs), of recent studies on the influence of grade expectation and workload on student ratings, of attempts to employ agency theory both to improve teaching quality and raise student ratings, and of institutional attempts to refine the peer review process, all suggest that we can "put teaching on the same footing as research" and include professional regard for teaching content and objectives, as well as student ratings of effectiveness and personality appeal, in the process.

> ...in the winter term of 1992, the Simon School faculty passed a resolution, that determined: "[T]o establish a faculty committee to evaluate teaching content and quality on an on-going basis. *The intent of the proposal is to put the evaluation of teaching on the same footing as the evaluation of research.* The committee will have the responsibility to evaluate both the content and presentation of each faculty member on a regular basis to be determined by the committee.... The output of this process should be reports designed to provide constructive feedback to faculty and evaluations to be considered in promotion, tenure, and compensation decisions." (Faculty Meeting Minutes, University of Rochester, William E. Simon Graduate School of Business Administration, February 26, 1992, cited: Brickley and Zimmerman, 1997, p. 5, emphasis added).

Introduction

"Put teaching on the same footing as research?" I can hear my scholarly colleagues ask, "You mean another attempt to credit those who do 'teaching' to the detriment of their 'research'?" No, my friends,

what I understand from the quote in the box is that administration would measure the quality of "teaching" on the same basis they demand from "research."

In 1997, in response to growing concern about maintaining quality of teaching and learning in expanding institutions of higher education—not only in Hong Kong, but worldwide—the University Grants Committee (UGC) (Hong Kong), undertook a study of the process by which teaching and learning quality was to be evaluated in Hong Kong institutions of higher education. This became known as the Teaching and Learning Quality Process Review (TLQPR) of 1997.

A series of institutional studies addressed critical problems bound to arise in an atmosphere of democratic interest in promoting expansion of economic opportunity and social mobility by means of wider access to higher education. It also revealed concerns within the institutions and the academic profession at large regarding free exercise of the functions of research and teaching, and their survival in light of calls for greater public accountability.

The UGC panel assigned to conduct the Teaching and Learning Quality Process Review (TLQPR) of the author's own University expressed its concern about the institution's reliance, almost exclusively, on mean quantified scores of student responses to course surveys to assess the quality of teaching and learning. This has also been a significant problem in teaching quality assessment in U.S. institutions since adoption of formalized "student evaluation" mechanisms as the result of student protest movements in the late 1960s and the 1970s.

No doubt, every teacher likes to be appreciated by his or her students. Similarly every student has an interest in minimizing risk in evaluation of his or her own course performance. But surely this situation describes a source of conflict of interest—likely on both sides—as much as a demonstration of the "validity" of "student evaluation" of teaching and learning on the theory that "the customer is always right."

A considerable volume of published research in this area attributes a "validity" to figures that are allegedly replicable because of their apparent "consistency and stability." Yet, we are also told that: "The literature on validity, though extensive, remains very fluid and not perfectly conclusive." Still other researchers find that teaching ratings and learning are only "weakly related."

Some authorities on the literature tell us that in part this predicament arises from research concentrating on "construction of instruments to yield items and subscales which [are] intended to measure student learning outcomes." Yet they also report that others have found "content validity," i.e., "positive relationships between student ratings and achievement."

Chief factors that would establish "validity," these experts tell us, are that evidence suggests that students and instructors seem to agree on what constitutes "effective teaching" and on the qualities of "an ideal professor." This conclusion must be flawed if, as the present author suspects, the literature of education theory, and practical

experience of student responses indicate that these two do not always share agreement on what "achievement" is, what "good teaching" is, and perhaps even on what "education" itself should aspire to.

This article compares the presumption of "validity" of "student evaluation" of teaching quality with the results of recent studies at the University of Washington on the influence of grade expectation and workload on student ratings, on the results of attempts, at the University of Rochester, to employ agency theory both to improve teaching quality and raise student ratings, and with the peer review model employed at the City University of Hong Kong.

I. Concerns about Quality of Teaching and Research in Expanding Institutions in Times of Contracting Budgets

In the Plenary Address of an International Conference on the Application of Psychology to the Quality of Learning and Teaching held in Hong Kong, Professor Robert J. Sternberg of Yale University (Sternberg, 1998) warned that universities that have used IQ tests, and other standardized measures of practical intelligence or practical experience as sole standards of university admissions, have created self confirming systems. "Only those with high IQs succeed, because only those with high IQs are admitted." The "tragedy" of this self selection as a "social goal," he said, is that "in our emphasis on skills that benefit the individual, we have created societies in which. . .the optimization of our individual outcomes at the expense of common well-being is becoming ever more pervasive."

The point of this paper is similar: if by "Quality of Teaching and Learning" we mean what style of Teaching and Learning is most popular with our students, or most satisfies the expectations they bring with them from their schools, or what they believe most readily facilitates their immediate needs in getting jobs or obtaining professional certification, that is what they will confirm to us in student ratings.

If, on the other hand, our goal is to contribute to modifying the tendency to the rote learning and recitation method, and to promoting critical thinking and general education—as the Vice Chancellors of both sponsoring institutions of the Hong Kong Conference, the University of Hong Kong, and the Hong Kong University of Science & Technology, urged in their opening addresses—then we better attempt to balance student input, with reasonable professional efforts to meet those expectations.

In response to numerous and growing concerns about maintaining quality of teaching and research in expanding institutions of higher education, not only in Hong Kong, but worldwide (see, e.g., Clark, 1995)(Note 1), the University Grants Committee (UGC) (Hong Kong),(Note 2) has undertaken studies that will affect the funding of both the research and teaching sides of university functions. Three Research Assessment Exercises (RAEs), studies of the research being

done in Hong Kong universities, were carried out in 1994, 1996, and 1999. A study, not of teaching and learning quality as such, but of the process for reviewing the quality of teaching and learning in Hong Kong institutions of higher education—the Teaching and Learning Quality Process Review (TLQPR) (see: Massy ; French)—followed in 1997, and a second is proposed for 2000-2001.

Both sets of studies addressed critical problems bound to arise in an atmosphere of democratic interest in promoting expansion of economic opportunity and social mobility by wider access to higher education. Both also reveal concerns within the institutions and the academic profession at large regarding free exercise of the functions of research and teaching, and their survival in light of calls for greater public accountability. The author has already described some of the professional concerns arising in the Research Assessment Exercises, the RAEs (see: Lee, 1998). The following discussion will address similar concerns with respect to the TLQPR. Whereas the author has expressed some reservation with respect to the former (the RAEs), he is generally in agreement with the latter (the TLQPR)—and especially as it affects his home university.

II. Measuring Teaching and Learning Quality

The announcement of an International Conference on the "Application of Psychology to the Quality of Learning and Teaching" (Hong Kong, June, 1998), indicated that it "strongly emphasize[d] cutting-edge research on the application of psychological principles to improving learning and teaching quality, with the aim of developing a global perspective on learning and achieving motivation" (HKU; HKUST, 1997).

With research on psychology of teaching and learning so highly specialized that a paper submitted to the Hong Kong conference required at least one of 27 keyword codes to classify it before it could be considered, it would appear that there are at least that many psychological perspectives alone from which to evaluate quality of teaching and learning. No wonder the TLQPR was troubled to find institutions with only student ratings in place.

II. A. Standardized Student Ratings Surveys

II. A. 1. Sole Use of "Student Evaluations"

It is understandable, in light of the multiplicity of just the psychological perspectives on teaching and learning, that the UGC (Hong Kong) panel assigned to conduct the 1997 Teaching and Learning Quality Process Review (TLQPR) of the author's own University expressed its concern about our University's reliance, almost solely, on mean quantified scores of students responding to semester surveys to assess the quality of teaching and learning in our various courses: "There appears to be little systematic monitoring of teaching and learning quality [at HKUST] other than through the

[student] teaching evaluation questionnaires..." (TLQPR, 1998, para. 16). This phenomenon is doubtless far more pervasive than only at HKUST, or only in Hong Kong. The problem surely reflects not only that universities do not know better ways to evaluate teaching, but probably also that they have no clear idea of what they want to accomplish in their courses either.

Despite the University response to the TLQPR, this imbalance was still reflected in the subsequent HKUST, *Faculty Handbook, 1997*, where, after indication that review of faculty performance for retention or promotion would involve consideration of "research, teaching, and service," it is made clear that unlike the case with "research" and "service": Reviews of teaching performance rely to a greater or lesser extent on *student evaluations* . . . (HKUST, 1997, p. 169, emphasis added).

The appearance of being responsive to student concerns is such a pre-occupation with university administrations that follow the American model, that finding a professionally acceptable method of evaluating what reasonable people recognize to be the essential characteristics of good teaching continues to elude them. One of the leading American authorities on "student evaluation"—who has great hopes of reforming the prevailing system—concedes privately:

> Most universities in the USA give lip service to using information other than student ratings for teaching evaluation. However, at most places the information obtained by other means (teaching portfolios, peer evaluation) is rarely put into a form that permits ready use for evaluation. Consequently most places end up relying primarily on student ratings.

That was precisely the HKUST administration's response to the TLQPR. Despite elaborate verbal acknowledgment of the existence of all other means of evaluating teaching in theory, the official "Progress Report to the University Grants Committee" (2 March, 1998), comes full circle to student ratings, and essentially concedes that at HKUST there *is* nothing else—students evaluate teaching. The university administration then lists "repeat offenders" and "monitors" faculty "accountability":

> A more formal use of the student evaluation results to monitor Department accountability for teaching performance was introduced in the past year. It involves the identification, by the Academic Affairs office, of a group of instructors with particularly poor records of performance in the previous year. Department Heads were provided with a list of any faculty members in their own Departments who have been so identified, and asked to take appropriate corrective actions to help these instructors improve. In subsequent years, Department Heads will have to provide, for any instructor who turns up on the list as a "repeat offender," details on

what actions, if any, were taken, and a statement of planned future actions to address the problems. (TLQPR Progress Report, 1998, p. 2).

Surely, every teacher likes to be appreciated by students. But, is that why our University relies almost exclusively on that one measure—what our students say about us—to assess our teaching competence? I doubt it seriously.

In Hong Kong, as elsewhere, institutional growth accompanied growth of student population. A subsequent dramatic change in the rate of student population growth, together with declining economic growth, means that there is, now, a heightened awareness of inter-institutional competition for student applicants (see: e.g., JUPAS, 1997), which leads inevitably to greater sensitivity to student tastes and student demands—doubtless one of the chief sources of the "student evaluation of teaching" movement in the first place (cf. Imrie, 1996).

Institutional growth, especially in Hong Kong, had been phenomenal in recent years (see: UGC, 1996). We are told, that full time equivalent enrollments (FTEs) in higher education increased from 42,000 in 1990-91 to 62,000 in 1995-96, or an increase of roughly 47% in only five years, giving rise to concerns about how institutions would be able to maintain the quality of teaching and learning (HKU, 1997, para. 3), but also about how new institutions would fare in regard to competition for student enrollments.

II. A. 2. Why Is There No Other Established Measure?

Over the years, there has been a great deal written about the overemphasis on, and inherent conflict of interest in, "student evaluation" of professional performance—for which there is no parallel in any other profession (see: Appendix: "Conflict of Interest," 1974-82, and "Formative"and "Summative" uses, 1970s). But how did it happen that there was no existing institutional system of measurement of teaching and learning effectiveness in the first place, that would have addressed quality of teaching and learning concerns suitably, prior to the massive expansion of the use of "student evaluation"? Ask any college or university teacher and you are bound to get a sense of why: "Academic freedom" (Note 3) (cf. Flexner, 1967)—i.e., from the perspective of what the Germans call, "*Lehrfreiheit*," the "freedom to teach without interference." None of us is particularly fond of having other colleagues, or administrators, poking their noses into how or what we teach.

As a consequence of our profession's concern with generations of political and ideological attempts to control what we can do or say in the classroom, we have been brought up with an academic legacy of resistance to thought control and, therefore, have developed no mechanism or standard, universally accepted, for assessing what we do, professionally, or how well we do what we do in the classroom. Consequently, the teaching profession was an easy target for institutions seeking to satisfy reformist demands in this area in the late

institutions seeking to satisfy reformist demands in this area in the late 1960s and early 70s. For this reason, and because of our even greater subservience to those in the education schools, in teaching technology, and in educational testing, we have allowed new professions to arise which specialize in telling those of us who teach "how to do it better." (cf. UGC, 1996, p.8) (Note 4)

All of us in the academic world know that our students will observe and react to our flaws and weaknesses as much as to our strengths. Yet, when it comes to assessment of our professional performance and abilities, most of us expect the same courtesy in evaluation as is accorded to other professionals (cf. Appendix, "Consumerism," 1976-91) (Note 5)—and to our students:

- evaluation by those who understand what we are attempting to do;
- evaluation by those who have a professional understanding of what we should do;
- evaluation without conflict of interest; —as well as, of course,
- evaluation for effectiveness.

II. A. 3. Need for Student Feedback

There is no need to convince the present author—at one time or another a candidate for five university degrees—that students often have valid opinions and cogent arguments. Which one of us, as a student or a faculty member, has not sat through lectures, and even whole courses, that we would be ashamed to have given ourselves. Simply being boring is a malady that even the best of us suffers from at times. These are concerns, which certainly should not be silenced, and perhaps also deserve some greater outlet for discussion on all campuses.

The Harvard Crimson Confi-Guide once served a function like this. At one time the independent Harvard University student newspaper gathered and published student comments on their Harvard courses—a short web search revealed that they still do. But that is all it purports to be. It makes no pretense of being a "survey," of being "scientific," or even of being "quantitative" in its results. It refers to itself as embodying: "Irreverent and honest appraisals of your favorite (and not so favorite) Harvard courses":

> Be very careful what you do with this guide. Read. Enjoy. Laugh out loud. The goal of the *Confidential Guide to Courses* is . . . to help students by giving them the lowdown on classes. Is it good? Is it a gut? Does the professor give interesting lectures? Are the exams difficult?
>
> This guide generally succeeds in providing that information, but that doesn't mean the articles have all the answers. They are meant to be helpful, but they can't necessarily be taken at face value.
>
> Each article is an opinion piece written by a student

> who took the class recently. The author can say whatever he or she wants, no matter how big the chip on his or her shoulder. It's important to remember that different people can come away from the same class with different impressions. . . .(*Confi-Guide*, 1998).

Instructors know, or ought to know, that they can get feedback from their students on how effective their teaching style is. Some do this by survey; some by private chat; some by instinct. But this does not mean that every student comment is good as gold or ought to be taken to heart. A professional person has to know for himself or herself what to make of such comments. That is not what standardized testing or survey research does, however. As we all know, you cannot argue with the question where you already know that the tested population is so large that the examiners—or the survey experts—are only looking for a positive or negative response pre-defined to carry specific conclusory meaning. That may sound like poor survey or test writing. Nevertheless, practically speaking, any teaching rating questionnaire will call for these same up or down responses. Professor Wilbert McKeachie, probably the most authoritative figure in the student ratings genre writes critically of this technique:

> . . . effective teachers come in many shapes and sizes. Scriven (1981) has long argued that no ratings of teaching style (e.g., enthusiasm, organization, warmth) should be used, because teaching effectiveness can be achieved in many ways. Using characteristics that generally have positive correlations with effectiveness penalizes the teacher who is effective despite less than top scores on one or more of the dimensions usually associated with effectiveness. Judging an individual on the basis of characteristics, Scriven says, is just as unethical as judging an individual on the basis of race or gender (McKeachie, 1997, p. 1218).

With all respect, there is something disingenuous about this admission. Those who have done most to promote the concept of "validity" of measures here admit they may be accurate *only* for what they measure literally. Then they argue that they do not measure what administrators are known to want to apply their quantifiable results for. They give teaching assessment committees a howitzer and tell them to use it like a smart bomb:

> Almost as bad as dismissal of student ratings, . . . is the opposite problem—attempting to compare teachers with one another by using numerical means or medians. Comparisons of ratings in different classes are dubious not only because of between-classes differences in the students but also because of differences in goal, teaching methods, content, and a myriad of other variables . (McKeachie, 1997, p. 1222).

In other words, (1) ratings are considered "valid," yet, (2) the quantified results relate only to individual performance. That is, they may presumably be used for "formative" and "summative" purposes—i.e., to advise that particular instructor how to *improve teaching*, and, ultimately, to advise the personnel committee how to *judge effectiveness* of that instructor. However, whereas results are expressed in quantified form, the scores for identical qualities are to be considered "not comparative."

It may be that schools with great sophistication in the use of student survey scores express such a qualification as to how student numerical ratings are to be applied— publicly. In practice, however, I do not see any hesitation in considering an 80% rating of one instructor equivalent to an 80% rating of another. At the author's University, for example, both get congratulatory letters from the Dean. Similarly, with a 40% rating for two years in a row, any instructor is bound to be considered a "repeat offender."

Accordingly, with regard to survey sophistication at HKUST, we are forewarned: "Note that the descriptions of the ratings should not be taken literally." (HKUST, 1998) Read further, however, and one is told that: "The average scores for all courses is in the range 60-70, so that the 'average' course has an 'above average' rating (HKUST, 1998)."

Does this mean that our administrators are so sophisticated about statistical and survey measures that they count these scores for no more than a simple exercise in measuring student opinion? Not on your life. We already know from Section II.A.1. above, that "Reviews of teaching performance rely to a greater or lesser extent on *student evaluations*," and "repeat offenders" will be dealt with.

Let me say first of all that the Hong Kong University of Science & Technology would rate itself as among the top universities in Asia—if not in the world. But "the average scores for all courses," judged by our students, we are told here, are rated between D+/C- and C+/B-. Heaven help the instructors whose average grades for their own students actually looked like that! But perhaps you may say that our students are more honest about us than we are about them.

What is the source of this disparity in ratings between faculty of students and students of faculty? Grade inflation can also have varying sources—since, according to this report, at least, it is not simply producing higher faculty ratings. Presumably the faculty believe that they are achieving better results with students than students give them credit for. Does it go too far to suggest that the two may have different concepts or goals of teaching and learning in mind, and that that is what their respective grades and ratings scores are measuring?

This disparity in concepts and goals of education will be dealt with further below (at Section II.A.6). In this connection, however, let us take a closer look at something else Wilbert McKeachie alludes to in passing in his paper in the "Current Issues" section of the *American Psychologist* (November, 1997) devoted to controversy over findings in the students' ratings research. McKeachie is willing to admit exactly the inherent contradiction of goals and objectives in student evaluation of teaching:

> There are . . . two problems that detract from the usefulness of ratings for improvement. . . . Many students prefer teaching that enables them to listen passively—teaching that organizes the subject matter for them and that prepares them well for tests. . . .
> Cognitive and motivational research, however, points to better retention, thinking, and motivational effects when students are more actively involved in talking, writing, and doing.

This inherent conflict of interest, notwithstanding, McKeachie justifies the continued reliance on the ratings survey system on the basis of what it is conceptually intended to achieve, i.e., "feedback":

> The second problem is the negative effect of low ratings on teacher motivation. . . .A solution for both of these problems is better feedback (McKeachie, 1997, p. 1219:1).

Only one set of convictions can conceivably attempt to justify knowingly relying on a system of assessment that you concede is based on conflict of interest: (1) the persuasion that an institutional system of measurement of teaching effectiveness is mandatory for personnel decisions; and (2) that no professional measurement compares in "validity" (as we shall see shortly, he says as much) with student ratings.

Here, I suspect we do have the root of the dichotomy in the grading and ratings problem: "Many students prefer teaching that enables them to listen passively. . .and that prepares them well for tests," and judge faculty on that basis. On the other hand, many faculty members are persuaded that "retention, thinking, and motivational effects" are greater "when students are more actively involved in talking, writing, and doing." I suspect that they also tend to grade on the belief that they are achieving results of this kind. While each scoring system may be perfectly honest as far as what it purports to measure is concerned, as McKeachie says, ". . .the two problems detract from the usefulness of ratings for improvement," i.e., for the much vaunted "formative" effect. McKeachie, further on, gingerly admits, the two systems simply do not relate to each other: "However, student ratings are not perfectly correlated with student learning. . . . " (McKeachie, 1997, p. 1219: 2)

The "solution for both of *these* problems [may be] better feedback." However, while educational technologists may believe that they are promoting feedback, there is in reality little communication about these matters in large public institutions, either between faculty and students, or between each among themselves. Student ratings are an *educational technology product* that, regardless of the mildly qualified claims of those who argue "validity," provide academic administrators with what purports to be quantitative measurements of teaching effectiveness—and that is precisely how the survey

technologists expect them to be used:

> But what about the use of student ratings for personnel decisions? Here again the authors of the articles in this *Current Issues* section [of *American Psychologist*, November, 1997] provide reassurance. All of the authors (and I join them) agree that student ratings are the single most valid source of data on teaching effectiveness. In fact, as Marsh and Roche (1997) point out, there is little evidence of the validity of any other sources of data. (McKeachie, 1997, p. 1219:2).

II. A. 4. Attractiveness of "Student Evaluation" Surveys

The beauty of student ranking surveys for a college or university administration is that they are cheap, and that they purport to offer exact *quantitative*, and, like it or not, *comparative* figures between faculty members. On their face, they appear to be the unqualified ranking by a representative sampling of students taking a course—without need for discursive explanations—moral, legal, or professional. The president of the author's university also reports that instructors have been fired because of low ranking in student evaluation surveys: ". . . In terms of system, all courses are evaluated by students and the results are disclosed on the World Wide Web; unsatisfactory teaching performance has resulted in many cases of contract non- renewal or salary bar. . . ."(Woo, 1997)

In a note in reaction to the foregoing observations, the President seems to take a more balanced view: "We certainly cannot just rely on student evaluation *scores*. Good teachers often get remembered only long after the students have graduated." This was despite subsequent publication of the "Report to the University Grants Committee" (2 March, 1998) cited above. Obviously the President has sensibilities as a teacher as well as an administrator.

II. A. 5. Crucial Variables and Consistency and Stability of Results

With the exception of some actually sometimes crucial variables(Note 6)—prior subject interest, class size, time of day a course is taught, rank of the instructor, grades expected, and course load which educational measurement investigators acknowledge affect student ratings of faculty in some way (cf. Appendix)—there have been a number of student ratings researchers who have argued that the student survey system is "consistent and stable." That is, they argue, similar ratings are seen to be attributable to the same faculty members, regardless of the subject matter they teach, and from year to year. Moreover, some investigators attribute close correlations to more professional appearing reviews by peers, administrators, and alumni (cf. Appendix).

Yet, while such correlations between results of different groups

of survey subjects may exist at times, other researchers tell us that, teaching ratings and learning are only "weakly related" (Gramlich; Greenlee, 1993). To the extent that this is true, it would tend to link the rating with the faculty member's teaching style or personality, and would tend to obviate one supposed major purpose of ratings, i.e., that they are "formative," that they can be used to assist the instructor to achieve improvement either in the teaching itself, or in its reception by students.

Nevertheless, some researchers in this area attribute a "validity" to figures that are supposedly replicable because of their apparent "consistency and stability." Yet, the same authority tells us: "The literature on validity, though extensive, remains very fluid and not perfectly conclusive" (Arubayi, 1987, p. 270).

In what A.G. Greenwald has called "the best of the largest group of construct-validity studies" (Greenwald, 1997, p. 1184) there seemed to be evidence to support correlational validity between student ratings in multisection courses. Here the results of student ratings were compared for different instructors giving different sections of the same course, where similar or identical examinations were given to different sections with students with similar ability (Abrami; Cohen; d'Apollonia, 1988).

The present author, who has, heretofore, limited himself to reviewing the literature on this subject, must interject at this point that he has observed completely unforeseen but sharply conflicting statistical results on this particular kind of experiment. The author gives a non-technical course, required for certification, by undergraduate engineering students. The enrollment of 350 students was divided into five sections of circa 70 students, and given in consecutive hours on the same days—all with the same instructor and identical workload and examinations. However, a student ratings curve emerged that dipped 1.5 deciles from the first to the third sections, then rose again, at the same rate, from the third to the fifth. When the instructor wondered aloud whether he actually wore thin in "quality" from noon to 3:00 o'clock, then reverted to form from 3:00 to 5:00, students objected: "Oh, it has nothing to do with the time of day. You know we do not come to the sections we are assigned to. We come whenever we feel like it. So it has nothing to do with your teaching, or the time of day, at all. *It is a matter of which Department is enrolled in which section.*"

As it turns out, the Admissions, Records, and Registration (ARR) office had assigned students, not to their section of choice, but rather as blocs of students by Departments or Programme—and ratings were tallied accordingly. The first and fifth sections were 100% Mechanical Engineering (MECH) and Computer Engineering (CPEG) students respectively, and the third 100% Electrical and Electronic Engineering (EEE) students. The second and fourth sections were 50% EEE, and 50% MECH, and 50% CPEG, students respectively.

"Why this difference?" I asked. "Because EEE has the heaviest workload." "And, they think they are the best." Or so I was told. In other words, student reaction was: the ratings curve had little to do

with the difficulty of the course—or with the ability of the students. Rather, student opinion had it, it reflected primarily the EEE students' image of themselves, at worst, in inverse proportion to their actual ability.

Results of "construct-validity" studies, notwithstanding, a veritable shock wave occurred in "validity" studies when a prominent psychologist encountered even sharper, and apparently random, variation in his own students' ratings:

> My interest in student ratings had a sudden onset. In 1989, I received the highest student rating evaluations I had ever received at University of Washington, for teaching an undergraduate honors seminar. The sudden interest came, not then, but a year later, when I received my lowest ever evaluations. The two ratings were separated by eight deciles according to the university's norms—about 2.5 standard deviations apart. But these two ratings were for the same course, taught in the same fashion, with a syllabus that was only slightly changed.
> The two juxtaposed ratings contained more than a mild hint that my students' responses were determined by something other than the (unchanged) course characteristics or the (presumably unchanged) instructor's teaching ability (Greenwald, 1997, p. 1184).

The experimental results of A.G. Greenwald and G.M. Gillmore have since persuaded many that "grading leniency" and "workload" are the two leading influences on student ratings results. The University of Washington has adopted their modified student ratings questionnaire in an effort to compensate for this bias.

Yet for those who believe that student rating is a "valid" measurement of quality of teaching and learning, the discussion goes on. In his conclusion to his article in the "Current Issues" Section of the *American Psychologist* that the work of Greenwald and Gillmore precipitated, Wilbert McKeachie addresses the puzzlement of Greenwald in discovering the aberrations in his teaching scores in the article that set off this symposium:

> Had I been consulting with him about the ratings, I would have said something like this: Tony, classes differ. Effective teaching is not just a matter of finding a method that works well and using it consistently. Rather, teaching is an interactive process between the students and the teacher. Good teaching involves building bridges between what is in your head and what is in the students' heads. What works for one student or for one class may not work for others. Next time, get some ratings early in the term, and if things are not going well, let's talk about varying your stategies. (McKeachie, 1997, p. 1224).

One can not quarrel with that statement. But Professor McKeachie and Professors Greenwald and Gillmore are all very experienced at teaching. And they teach a popular medium—the measurement of student opinion. I dare say that nothing any one of them is likely to do in the classroom is going to damage his long-standing reputation. It is all the rest of us—who also know and share this philosophy—and who also measure how and what we teach in the classroom according to who we teach, and how they receive it, who should be concerned. It is not the philosophy that is the threat in the annual reviews. It is rather that it is unlikely to come to a discussion of the substance of quality teaching and learning in a personnel evaluation committee that only compares quantified results.

Since writing the foregoing, the present author has found that exactly the same curious ratings curve—for the same classifications of students, and mixes of students occurred in his classes the following year as reported above. Is there some lesson to be learned here from Professor McKeachie's advice to Professors Greenwald and Gillmore? Is it a matter of, as Professor McKeachie says:

> "classes differ" "Effective teaching is not just a matter of finding a method that works well and using it consistently"?

And that that advice applies hour by hour as well as year by year? How would I respond?

> Wilbert, I think there's something else afoot here. Do you think these strange results support those survey experts who argue "validity" on the basis of the "consistency" of student ratings of the same instructor from year to year—and "regardless of the subject matter" the instructor teaches? I suspect that those survey experts have neglected to mention the possibility of some deeper form of personality variation between MECH students, EEE students, and CPEG students, and classes where they are equally mixed!
>
> Remember what the students themselves had to say about this?
>
> "You know we do not come to the sections we are assigned to.
> We come whenever we feel like it."
>
> That does not sound like a case remediable by different teaching *strategies* hour by hour to me. I might point out that CPEG is an elite Programme within an already elite EEE Department. An administration concerned as much about student welfare as comparability of faculty ratings might be wise to look into a source of student disaffection between Programmes.
>
> At any rate, I think we owe the rigidity of ARR (the

office that assigned these students class hours by Programme classification rather than personal choice) a vote of thanks for uncovering what otherwise would be concealed in a mass teaching ratings survey. In years when the whole class enrollment of 350 would have been counted as one whole, we would only get the overall impression that the median rating of the instructor was poor. Now we know that student personality varies by programmes in our Engineering Faculty—and that they vote as a bloc—though what the source of this aberration is remains uncertain.

However, the situation suggests to me that there may be more to be learned here from what W.O. Weyrauch has called "the law of a small group" (Weyrauch, 1971) than from student teaching ratings theory. Students tell me that collective solidarity is very important among engineering students:

"If the others don't talk in class—you don't talk in class. If the others don't talk in English—you don't talk in English."
(Cf. Wong, 1984, cited at Section II.A.6.below)

There are also noticeable personality differences between classes—which for the most part means between Departments. You hear the boisterous roar of the MECH students as you get within 50 yards of the classroom.

"The Mechanical Engineering students are *naughty*!"

... they tell me in English—quoting my colleagues who prefer class discipline to Socratic problem solving. CPEG students are an elite Programme in an already elite EEE Department. And they seem sedately well content with the attention lavished on them by their mentors since an internationally acclaimed computer scientist came here to set up that Programme. There is more diversity, and less mentoring, I hear, among EEE students. My guess is that more of them work individually—and skip classes where they think they can make up the work later on. For these students, a clear set of organized class notes from a friend is vital. "None of your Socratic problem solving for me," they seem to say,

"Just give me the Notes!"

... I have read remarks such as these in the write-in blanks of the ratings questionnaires. Of course, "varying your strategies" could work here. But who is giving this course? Do you give up on MECH and CPEG because half the EEE students would rather work on their senior projects? No,

> Wilbert, thank you very much. Teaching is also about professional integrity and professional satisfaction. "[T]here is little evidence of the validity of any other sources of data," you say? Since when has that ever set the standard for academic decision-making?

II. A. 6. Is There Validity If There Is No Agreement on Outcomes?

The same authority on the literature who argued "validity" because of apparent "consistency and stability" tells us that part of the predicament of "fluidity in research results" lies in the research concentrating on "construction of instruments to yield items and subscales which were intended to measure student learning outcomes" (Arubayi, 1987). He reports that others have found "content validity," i.e., "positive relationships between student ratings and achievement" (Arubayi, 1987).

Other factors that would establish "validity," this expert tells us, are that:

> Evidence suggests that *students and instructors seem to agree as to what leads to good teaching.* Similarly, . . . very close *similarity between the perceptions of students . . . on what constitutes a[n] "ideal professor." If students can agree with their instructors* as to *what constitutes effective teaching* and the qualities of an ideal professor then one might be sage to conclude that students are mature enough to rate or evaluate instructors and instruction (Arubayi, 1987, p. 270f. emphasis added).

Reliance on near-exclusive use of "student evaluation" of teaching is bound to arouse concern for those of us in Hong Kong—where there are also faculty members to be found, who, while deeply attached to the region, their students, and the subject matter of their fields, do not share agreement with their students on what "achievement" is, what "good teaching" is, and perhaps even on what "education" itself represents.

In no way does it dispose of the issue to say that those faculty members are themselves out of joint, and that the situation will be cured by localizing expatriates out and putting local people in their place. The definitions of "education" and "achievement" are not simply heritage and culture-bound. An institution like the Hong Kong University of Science & Technology is overwhelmingly staffed by PhDs from the world's leading universities. Are we to believe that they are prepared to abandon the educational values they hold for themselves—and upon which they want their own research and career accomplishments to be judged—when they instruct their students?

"We ought to teach every course the same way we would teach majors in the United States," our University President Woo Chia Wei is reported to have opined—somewhat at odds with what as an administrator he seems to be telling us. Are we to believe that there is

one set of values for the world, and another for our own students?

How would I teach in the U.S.? Like an Ivy League graduate would be expected to:

- Evaluating how we GATHER FACTS;
- Establishing how we DEFINE A PROBLEM;
- IDENTIFYING ISSUES and METHODS leading to various SOLUTIONS of a problem;
- STRESSING REASONING over factual information;
- STRESSING HOW WE REACH CONCLUSIONS—NOT OPINIONS (Lee, 1997).

Does this form of teaching offer an advantage to Hong Kong and to China? Many of us believe it does—not least of all the Vice Chancellors who keynoted the international conference in Hong Kong on Teaching and Learning Quality.

By no means do all Western educated scholars in Hong Kong pursue this method. But, those who do, know that this style of teaching is not the mainstream tradition of the region. The instructor dedicated to this approach is, therefore, faced with the deliberate choice—of attempting to bring his or her students out of their protection of silence and anonymity to develop discursive verbal abilities (Lee, 2000) or—of abandoning what he or she believes is both sound practice—and attainable with persistence—in order to pursue the more accepted purely didactic approach that will gain him better ratings.

Many of our students are afraid that departure from their accepted learning habits—and how such a change in them will be received by their peers—will create a disadvantage to them in competing: first with their own classmates for grades, then with their fellow graduates, for jobs. They are, therefore, more at home with the standardized testing and curved grading results aspect of the American heritage, believing that they must receive and repeat exact information to be "testable," and that it is, therefore, "unfair" to them to introduce new standards of teaching and learning that suddenly give away their "place on the curve."

These conclusions are not based upon a formal scientific survey, but do derive from years of listening to student comments, both personal and anonymous. However, more formal case studies in Hong Kong have produced similar results. In a case study on law student learning in English at the University of Hong Kong, for example, three language use researchers conclude: "...by the time students reach the end of their secondary education and probably well before that point, they have internalised a set of unstated survival strategies for choosing which language to use [Cantonese or English] or, indeed, whether to communicate at all in a given situation." (Corcos; Churchill; Lam, 1998).

They refer to a set of implicit socio-cultural rules derived by an earlier researcher in this area:

- If you want to talk to another student in a friendly way and

without seeming superior, you must not use English;
- Do not show off your language proficiency in front of your peers;
- You should deny such proficiency if anyone praises you;
- You must hesitate and show difficulty in arriving at an answer when called upon by the teacher;
- You must not answer the teacher voluntarily or enthusiastically in English;
- You must not speak in fluent English (Wong, 1984, as cited).

Similar defenses to class response techniques apply in other parts of the world (even in some parts of the U.S. where "class participation" is established doctrine), however, in Hong Kong, university instruction in English, a foreign language, though still the basis for official and business communication, serves as cover for non-participation. Actually response in Cantonese is no better—if students are not accustomed to verbal reasoning.

II. B. Measurement and Enhancement of Teaching by Peer Review

Of course you listen to your students—and you adjust to whoever comes. But is that all there is? If better teaching and enhanced learning are desired, experience tells us that they can be encouraged or cultivated—the elements are all well-known. (Note 7)

We may agree that there is a difference between encouraging enhanced quality of teaching and learning, and merely conducting a survey to see whether teaching conforms to students' established expectations. However, encouraging better teaching by whatever method may involve changing incentives and investing greater resources, and may, therefore, discourage administrators from pursuing such a course too vigorously in times of contracting budgets. But testing is cheap, and appears to satisfy the student constituency.

II. B. 1. Changing Incentives from Research to Teaching

The process by which incentive structure can be changed in a university environment has been described in the literature in the same terms as changes in incentive structure in business. This process was employed in efforts to reinforce the teaching and learning environment at the William E. Simon School of Business Administration at the University of Rochester, and apparently in other leading American business schools, when the administrations determined that environmental factors affecting them, leading to competition for public funding and for student applicants, were similar to those described at the outset of this paper as leading to the Research Assessment Exercises (RAEs) and Teaching and Learning Quality Process Review (TLQPR) in Hong Kong (see: Brickley; Zimmerman, 1997—the following relies on that report).

The birth rate has long been declining in the United States, leading, over the years, to declining numbers of children in schools, and, as a result, declining numbers of students in colleges and universities. In the late 1980s this reduction in numbers of applicants

was also felt in the graduate schools of business—combined with a lower demand for MBAs as a result of economic conditions.

Competition for applicants among American business schools first led to enhanced spending on public relations, then on scholarships, and, finally, on enhanced spending on incentives to improve the teaching environment. At about that time, *Business Week* began publishing a biannual list of top-20 business schools, and asked graduating students and recruiters to rate the schools according to opportunities 1) either in class or in extracurricular activities, and 2) to nurture and improve your skills in leading others (Byrne; Leonhardt, 1996).

Focus on Research emphasis, so important in the competitive standing of former years, received no special mention, and seemed to have fallen by the wayside in a competition fired expressly by students' interests.

Concern with media rankings seems to have been quite intense. The Simon School at Rochester, was for example, listed in the *Business Week* top-20 business schools in 1988, and 1990, but not in 1992. As a result, a number of business schools, including Rochester, were led to serious reconsideration of their academic programs—emphasizing enhanced incentives to improve teaching. A faculty report at Rochester called for efforts to:

> . . . increase teaching incentives, and make the change clearly visible to applicants, students, administrators and faculty
> ("MBA Program Status Report," University of Rochester, William E. Simon Graduate School of Business Administration [June 14, 1991] cited: Brickley and Zimmerman, 1997. Cf. also: "The Report of the Task Force on Improvement," M.I.T., Sloan School of Management [May 7, 1991]).

To meet the demands of that situation, the School of Business Administration at the University of Rochester determined to become more competitive in the market for business school applicants. In the process, they determined to enhance their standing as a top-20 business school by seeking to attract student applicants by an enhanced teaching and learning environment—a significant change from the emphasis on advanced Research in the 1980s, when the applicant level was strong and rising.

II. B. 2. Changing to a Peer Review Measurement System

It is interesting to observe that at about the same time as The Simon School at Rochester was engaged in the process of re-assessing its system for teaching evaluation, a similar process was underway at the City University of Hong Kong—for different reasons.

In 1993, the year before full university status was conferred on the then City Polytechnic, the Academic Board (now the Senate)

established a Quality Assurance Committee which laid down guidelines for, among other things, teaching evaluation (QAC, 1993). While emphasizing that teaching evaluation "*must* include student feedback as a substantial primary element in the process," the Guide makes clear that teaching evaluation must also be an institutional determination: conforming with stated "*policy*" and "*principles*," based on all available "*evidence*," fully "*documented*," and "*accessible*":

> Teaching evaluation must conform to the *Principles* stated. . . . Teaching evaluation schemes must be documented. . . .
>
> The *primary* purpose of any teaching evaluation scheme should be to improve teaching. Teaching evaluation schemes *must* include student feedback as a substantial primary element. . . . Where a scheme is designed to evaluate teaching for assessment purposes, evidence *must* be included from other appropriate sources such as peer review, individual reflection, expert observation, etc., *in addition* to student feedback. . . . Those entrusted with using the information from teaching evaluations for decision-making related to career progression should be skilled in interpreting and drawing together the different sources of information. . . . In all cases the staff member being evaluated must be fully consulted. . . . Provisions should exist for regular review of the . . . evaluation schemes and of the institution's evaluation procedures (QAC, 1993, p. 1f.)
>
> (The first paragraph is taken from "policy," the remainder from "principles." The Guide is undated, but acknowledges Hall; Cedric; Fizgerald, 1994, as the source from which its principles were developed.)

This policy has been applauded in the TLQPR at City University. Yet, both from the TLQPR, and from faculty comments, one gets the impression that this system has not been fully implemented at City University either.

In both cases cited above, recourse to a peer review measurement system was motivated by new roles of the institution—calling for greater attention to the teaching and learning mission. On the other hand, both institutions (or their faculties?) were remarkably sensitive to the implication that either matters of professional competency or career decisions might be driven purely by reaction to data arising solely from student inputs. Clearly, both institutions were acutely attentive to the importance of maintaining ultimate *institutional responsibility* for professional decision-making, and correspondingly, professional information gathering.

As a result of the situation described in the foregoing section, the Simon School made a significant decision to change from dependence solely on the student quantitative rating system for course and instructor, to a highly organized qualitative peer review system.

Based on the evidence of the cited study that teaching ratings

and Learning was only "weakly related" (Gramlich; Greenlee, 1993), and on the concern that "some instructors game student ratings by reducing course work loads and cutting analytic content," or "...hand out cookies, bagels, and wine and cheese the last day of class when student ratings are administered" (Brickley; Zimmerman, p. 5), in the winter term of 1992, the Simon School faculty passed a resolution, that determined:

> [T]o establish a faculty committee to evaluate teaching content and quality on an on-going basis. *The intent of the proposal is to put the evaluation of teaching on the same footing as the evaluation of research.* The committee will have the responsibility to evaluate both the content and presentation of each faculty member on a regular basis to be determined by the committee. . . . The output of this process should be reports designed to provide constructive feedback to faculty and evaluations to be considered in promotion, tenure, and compensation decisions.
> ("Faculty Meeting Minutes," University of Rochester, William E. Simon Graduate School of Business Administration [February 26, 1992], Brickley; Zimmerman, p. 5 emphasis added).

In the case of City University of Hong Kong, the faculty Quality Assurance Committee (QAC) took a more systematic approach, in a manner befitting its role in determining future guidelines for policy of a major university, it devoted its early efforts to outlining statements of principles on quality and quality assurance. While these principles clearly were to acknowledge the role of students and other "stakeholders," e.g., employers and professional bodies, they were not to be construed in such a way as would utterly disenfranchise the teaching faculty: "The systems of quality assurance must be capable of operating independently of the participation of particular individuals and have an integrity which enables judgements to be formed that are unaffected by other managerial imperatives." (QAC, 1993, p. 4)

What is recognizable from the City University statements and principles is that these derive from faculty deliberations and are not simply imposed from above. In this respect, they are unique in circumscribing the activities of the whole institution: "Quality assurance policies should embrace all activities of the institution (QAC, 1993, p. 4). These principles not only recognize the institution's public roles and obligations to student's and other "stakeholders," they declare that they will apply "in all aspects of the staff's role including teaching, research, and administration" (QAC, 1993, p. 4).

II. B. 3. Implementation of the Peer Review System

As long as an informal quantitative student rating of course and faculty member was the only goal, it could be accomplished with comparative ease by passing out and collecting questionnaires at the

end of the semester. If the evaluation of teaching were now to be put "on the same footing as evaluation of Research," then an objective means of qualitative measurement of the work of the course and the faculty member had to be found. For this purpose, the Rochester Business School faculty established a "Committee on Teaching Excellence" (CTE). The Committee developed a set of procedures, following the example of psychoanalysis, by first setting about evaluating six of the courses taught by members of the Committee itself:

> By the end of the 1993 academic year the CTE established a process, that except for minor changes, remains in effect through 1997. This process includes *benchmarking* the class with other top business schools: using a two-person evaluation team to *observe lectures*, *review material*, and *conduct student focus groups*; *video taping* several classes; *full committee discussion of the course*; and a *final written report* which goes to the instructor and the Dean's office and which is included in the faculty member's personnel file.
>
> . . . In addition to evaluating nine individual courses each year, the CTE held several seminars to discuss teaching. These forums allowed faculty to share their experience on various topics including: teaching cases, using computer- based presentation packages, and managing class discussion ("cold" calling). These seminars in the 1995 academic year were the first faculty seminars devoted to teaching (Brickley; Zimmerman, 1997, p. 5, emphasis added).

Evaluating the teaching process—involving analysis of quality of inputs or preparation and materials, form of classroom delivery, and measurement of effect upon students and their achievement—is necessarily a time intensive effort for all Committee members. The opportunity cost to evaluate one course was estimated at (US)$15,000.

In the case of the City University of Hong Kong, as well, the section of the *CityU Policy and Guide for Developing Teaching Evaluation Schemes* dealing with peer review specifically refers to evidence drawing on the following topics, and calls for citation of evidence in each case:

1. **subject expertise**: (up-to-dateness of content material);
2. **module design**: (relationship between content and objective, sequence, etc.);
3. **enhancing student learning**: (activities included, assessment requirements, etc.);
4. **module organisation**: (variety of experiences, reading lists, availability of materials, etc.);
5. **supporting departmental goals**: (from departmental objectives);
6. **research supervision** (QAC, 1993, sec. 2.2.2).

The guidelines conclude with the admonition that any peer review scheme must emphasize "expertise," "integrity," and "training" (QAC. 1993, sec. 2.2.2), both in the collection of data and its interpretation. No doubt this system, as well, must require a considerable "opportunity cost" that the institution considers is justified.

III. An Assessment System that Dwells on the Past? Or Education Policy with Increased Incentives for Teaching?

It should not be necessary here to enumerate the extent of the literature on opinion survey research. Neglect of comparative validation of an investigator's particular empirical method, or neglect of the potential impact of pre- existing biases—both among the research subjects, and among the investigators—would ordinarily arouse sufficient consternation among scholars of the field that such results would receive little credibility.

As the foregoing has suggested, however, there has been little attempt to obtain general agreement on the standards of psychometric validity of student ratings of teaching despite the fact that investigators are well aware that their findings are being put to practical use in so-called "formative" and "summative" evaluation of members of their own profession.

Very simply, there appear to be two camps: 1) Those who treat student ratings as a reasonable "input" to "formative" and/or "summative" teaching assessment—along with all other professionally accepted indices; and 2) Those who consider that student ratings are the "valid" and sufficient basis for "formative" and "summative" evaluation of teaching by themselves. Institutions that employ student ratings alone tend to be interested primarily in quantitative and comparative results—i.e., numerical values that can be employed across the board to gauge and reward faculty performance.

Within the context of the empirical research reports, however, little interest is shown in qualitative criticism of the formulation of survey questions in student opinion surveys—and little attention is given to the impact of value systems in interpretation of survey questions. The foregoing has shown that leading authorities in the area: e.g., Scriven and McKeachie recognize the danger of confusing "characteristics that generally have positive correlations with effectiveness" with either "effectiveness" per se, or as all there is to be said for good teaching, or, more important, what teaching policy should aspire to.

Recognizing the needs of students in acquiring the skills to comprehend and master the subject matter of their field, and response of the instructor to the needs of a particular body of students is certainly one aspect of good teaching. But formation of forward looking education policy, cannot endlessly avoid the necessity of considering the obligation of the instructor—and of the institution— to the public and to the profession of teaching, to pursue clear

educational goals which reflect the ambitions of our civilization and not simply those of any one generation of students whose priority is solely admission to professional qualification.

III. A. Haskell's Survey of the Literature of Pschometric Validity of Student Ratings and of Whether There is a Cause of Action for Violation of Academic Freedom for Reliance on Student Ratings in Personnel Decisions to the Exclusion of Everything Else

The serious omission of a qualitative discussion of psychometric validity of student ratings has been addressed in a comprehensive, at times rambling, series of four articles, a study of the literature of student ratings theory by Robert E. Haskell, Professor of Psychology at the University of New England in the United States (Haskell, 1997a,b,c,d).

Haskell is clear about his own personal position, "SEF [student evaluation of faculty] is deceptive regarding its negative implications for higher education" (1997b, p.3), and that the present system ". . .sets up a conflict of interest between the instructor and quality of education. . .[the] opposite of the original intent of SEF which was the improvement of instruction" (1997a, p. 16). It is inescapable that these considerations must return to the forefront of academic discussion at the turn of the century as democratization of access to higher education, now combined with increasing budgetary constraint, forces institutions to concentrate on issues of "*quality*" and "*accountability*."

Haskell's contribution lies in providing a kind of qualitative comparative survey of the ratings literature. He also recognizes that improper use of student ratings can result, and has resulted, in litigation over abuse of process in renewal, salary, and tenure decisions. He has attempted to study the possible remedy of use of the issue of violation of "academic freedom" in such litigation where litigants have attempted to identify academic freedom with freedom of speech, which enjoys unqualified protection under the American Constitution.

Haskell points out the conspicuous disregard of faculty rights throughout the period in which reliance on student ratings of faculty has been associated with student and minority rights causes: "A recent booklet on '*The Law of Teacher Evaluation*' (Zirkel, 1996) contains no mention of SEF cases. Nor does a recent comprehensive legal guide for educational administrators (Kaplin and Lee, 1995), nor do other reports (Poch, 1993) on the legalities of academic freedom, tenure and promotion" (Haskell, 1997b, p. 2).

Haskell's insight into the value of considering how the courts have reacted to cases based on student ratings could have led to a more significant contribution if his results had been more systematic and analytical. The second article, particularly, would have benefited from closer collaboration with a person trained in handling this kind of material. The colossal labor represented by this vast qualitative review of the literature of the field, notwithstanding, the value of the author's discussion of judicial opinion, is practically limited to the enumeration

of 78 cases where the issue of over reliance on, or neglect of, student ratings has been raised. Some of the cases are properly cited, others are not. High level court reports are listed side by side with low level. There is no attempt to distinguish between where reference to ratings would support the faculty member's case but are ignored, and cases where negative results are relied on to make decisions that should have been supported by professional opinion. There is little analysis of whether arguments for use of ratings on either side were well-taken.

There is, furthermore, no distinction made between *decisions* based upon use of ratings, and mere *obiter dicta*, or comments in passing mentioning ratings. Nevertheless, from Haskell's investigation of this problem we can begin to recognize that the concept of "academic freedom" does not seem to have been developed very far by the American courts themselves as a First Amendment (i.e., freedom of speech) category in connection with student ratings. (Note 8) On the other hand, there appear to be a number of efforts to combine complaints supported by reliance on student ratings with a theory of discrimination on the basis of sex or race—which is statutorily based and has a more consistent jurisprudence. Courts have developed measures such as "disparate impact" of policies on protected groups to support claims of illegal discrimination.

Haskell makes the valid point that whereas some lower courts have, in the past, distinguished between "freedom of speech," that was protected, and "action" in connection with expression of opinion, that was not protected (notably in *Lovelace v. S.E. Mass. Univ.*, 793F.2d 419 [1st Cir.1986]), the U.S. Supreme Court has overtaken them (Haskell, 1997d, p. 5). In 1989, the U.S. Supreme Court ruled that flag burning could be seen as political expression, and would, in that sense, be protected under the First Amendment (*Texas v. Johnson*, 491 U.S. 397 [1989]; see also: *United States v. Eichman*, 496 U.S. 310 [1990]).

On the other hand, there appears to be no American case law expressly protecting what the Germans call "*Lehrfreiheit*," i.e., freedom to teach with respect to methodology, coverage or organization of material, and grading. Indeed the cases cited suggest that some courts would allow interference in this area on the basis of institutional or public policy.

A teacher's right to say, or teach, what he or she believed to be professionally defensible would be protected. Of course, the requirement that a faculty member's expression of opinion be professionally defensible is clearly a limitation that would not apply to others—students, for example, or student ratings. Students, and other interested members of the public, can say whatever comes into their heads—providing that it is not outright defamation.

Perhaps because of lack of a sufficient number of appeals one does not learn whether any of these cases has led to a rule adopted either in the American state or federal courts. However, we do learn that numerous judicial reservations can be cited against relying on student ratings alone—to the exclusion of professional opinion—in faculty personnel decisions (Haskell, 1997b, *passim*). Impressively, the Canadian examples cited seem to stress the need for balance between

student ratings and professional assessment more than the American cases.

At the same time, we see the courts' hesitation to interject themselves into institutional decision-making. Haskell quite accurately characterizes the courts' unwillingness (unlike juries) to inquire into substantive criteria an institution applies for personnel evaluation as long as the procedural safeguards appear adequate—i.e., that the standard is applied generally to all faculty members (Haskell, 1997c, p. 4)—even though such criteria may appear to be incompetent when applied for the purpose. That was the case for a schoolteacher previously renewed over a 10 year period but terminated because her pupils ranked too low on the Iowa Test of Basic Skills (ITBS) and Iowa Test of Educational Development (ITBD). If measuring teaching effectiveness of the teacher on the basis of the performance of her pupils in standardized testing could be shown to be totally absurd or incompetent, the teacher might have been successful in thwarting dismissal. On the other hand, if a political decision, or public policy, calls for such a measure of teaching effectiveness, courts tend to leave judgment to the political arm, public policy, or simply institutional practice.

Yet, we must take care in characterizing judicial perspective. For, whereas course content and grading standards *may* be treated as a matter of institutional policy (Haskell, 1997d, p. 7), we also hear: "assignment of a letter grade is protected speech" (Haskell, 1997d., p. 6):

> [B]ecause the assignment of a letter grade is symbolic communication intended to send a specific message to the student, the individual professor's communicative act is entitled to some measure of First Amendment protection. (*Parate v. Isibor*, 868 F.2d 821, at 828 [6th Cir. 1989])(Note 9)

More disturbing is an allegation of professional incompetence in use of ratings by institutions which should know better, such as:

> According to Thompson (1988, p. 217), "Bayes Theorem shows that anything close to an accurate interpretation of the results of imperfect predictors is very elusive at the intuitive level. Indeed, empirical studies have shown that persons unfamilliar with conditional probability are quite poor at doing so (that is interpreting ratings results) unless the situation is quite simple." It seems likely that the combination of less than perfect data with less than perfect users could quickly yield completely unacceptable practices, unless safeguards were in place to insure that users knew how to recognize problems of validity and reliability, understood the inherent limitations of ratings data and knew valid procedures for using ratings data in the context of summative and formative evaluation (Franklin & Theall,

1990, pp. 79f.) (Haskell, 1997c, p. 6).

It asks a great deal of a court to assess an argument of this kind. Yet, there appears to be accumulating evidence that educational institutions, which are capable of evaluating psychometric standards, choose to ignore such weaknesses in favor of the efficiency of the continued unquestioned reliance on student polling results. All-in-all, we see diversity of judicial opinion may be comparable to the diversity of opinion in the psychometric survey discipline. Yet, what does appear from these citations is that while courts have not equated freedom of speech with academic freedom in all its manifestations, nor created a protected zone around assessment of teaching effectiveness, they have, from time to time, expressed clear reservations about reliance on student ratings in personnel decisions to the exclusion of everything else.

III. B. Should Forward Looking Education Policy Concentrate on Goals and Incentives to Improve Teaching?

The two authors of the study of the report on the shift to peer review of teaching at the Simon School of Business at Rochester tell us that there was a very rapid adjustment to changes in incentives—that was reflected by a corresponding rapid rise in student teaching evaluations:

> During the 1990s, there was a substantial environmental shift that increased the importance of teaching relative to academic research at top business schools. The Simon School, like other business schools, changed its performance evaluation and reward systems to increase the emphasis on teaching. One might have expected the effects of these changes to be gradual, given the human capital constraints implied by the composition of existing faculty.
>
> Our results, however, suggest a very rapid adjustment to the changes in incentives. Average teaching ratings increased from about 3.8 to over 4.0 (scale of 5) almost immediately. Teaching ratings continue to rise after the changes in incentives, suggesting additional learning and turnover effects (Brickley; Zimmerman, 1997, p. 21).

They believe this dramatic effect was owed to incentives rather than peer review. Whereas they had found that: "Some evidence suggests that research output fell" (Brickley; Zimmerman, 1992, abstr.) they continue that, thereafter: ". . .we find some evidence that faculty substituted research for teaching following the incentive changes" (Brickely; Zimmerman, 1997, abstr.).

On the other hand, these authors find that, in the long run, peer review may support "quality"— the declared objective of efforts in Hong Kong associated with the TLQPR, and with the City University QAC. But they are forced to recognize an inherent conflict of interest

when it comes to recognition of these efforts in student ratings:

> . . . Intense peer review of classes had no obvious effect on either teaching ratings for the evaluated classes or subsequent classes.
>
> One possible *reason peer review is not associated with higher student evaluations in the reviewed or subsequent courses might be due to the complementary nature of performance evaluation and compensation* [citing: Milgrom; Roberts, 1995]. The Deans' office did not formally announce that CTE reviews would explicitly enter the compensation policy of the School. An alternative explanation of the lack of statistical association is that "good" teaching as perceived by faculty evaluators and by students are orthogonal. For example, *faculty evaluations value courses with more intellectual rigor and greater work loads, whereas students value courses with more current business content, more entertaining lectures, and lower work loads*. (Brickley; Zimmerman, 1997, p. 22, emphasis added).

The turnaround process is described for us in terms of agency theory by the two faculty members of the Simon School:

> Agency theory suggests that the principal is interested in both the amount of effort exerted by the agent, as well as the agent's allocation of effort across tasks. As environments change, firms are expected to adjust incentive contracts on both dimensions. For example, the 1990s witnessed significant developments in information technology, which lowered the costs of measuring performance. These changes potentially help to explain why many firms increased their use of incentive compensation over this period. Similarly, changes in competition and technology motivated numerous firms to *increase their focus on quality over quantity*, for example, through the adoption of TQM programs (Brickley; Zimmerman, 1997, p. 22, emphasis added).(Note 10)

Changing incentives and "focus on quality over quantity" to concentrate more on teaching and learning—particularly in an environment which esteems research and/or technological development higher—is, perhaps, just as likely to involve more than merely issuing letters of congratulation to those who score high on student ratings polls.

IV. Open Decisions Openly Arrived At

Teachers may be stung by what students say if they ask for their students' opinions and find that they are significantly out of keeping with their own expectations. Of course, students have a right to their

own opinions. But teachers would be foolish to let themselves become ruled by *everything* students have to say—especially on those occasions when what they have to say derives from wholly different concepts of educational goals and/or is based on teaching practices contrary to wise learning patterns. They are students, and students test what they are thinking by saying it aloud.

If there are legitimate differences about teaching and learning, they must be addressed by the institution as well as the individual instructor. On the other hand, if low "student evaluation" figures reflect that an instructor comes into a class drunk, or is on drugs, perhaps does not come at all, or does not prepare, or preys upon those in his or her charge, then that instructor ought to be fired—you do not put his or her name up on the world wide web!.

But it is not students who post their opinions on the web. It is a university administration, which does this in place of deeper thought or due diligence. If a student calls me a fool, it may be an inept way to open a conversation—about what fools are. If a university administrator calls me a fool—he robs me of my right to teach.

Is there an inherent problem in recognizing a qualitative measurement for rating of teaching? For putting teaching evaluation "on the same footing as evaluation of research"? Isn't that what Universities do? In the 1996 Research Assessment Exercise (RAE) in Hong Kong, we are told, the research "output" of all research academics in the territory's then seven traditional "tertiary" institutions—covering 14,000 publications of 3,300 academic personnel—was assessed by 110 experts, many chosen worldwide, and all in less than nine months. If there is a way of obtaining assent of universities to standards for a monumental task of that kind, there must surely be an acceptable means of, at least, setting the standards for a professional teaching and learning quality review.

There is a reason, however, why the *CityU Policy and Guide for Developing Teaching Evaluation Schemes* takes such a judicious stand on the collecting of concrete evidence for teaching evaluation—this is a step that cannot be undone. And there is a reason why it calls for "expertise," "integrity," and "training," and applies the "quality" standards to the administration as well as the faculty. Too often these decisions are made behind closed doors not simply to protect confidentiality, but because ill-defined standards applied in secret leave no trace.

There may be a right of appeal. But no appeal ever corrected injustice that should not have been done in the first place. If we know the standards of "quality," and they are as clear as, for example, those in the *CityU Policy and Guide*, or those pursued by the Committee on Teaching Excellence at the Simon School, then let the sun shine in.

Notes

1. These concerns are well illustrated and documented by Clark. He considers the difficulties facing universities around the world from loss of funding for research and emphasis on mass

education. He describes the situation in universities in the United States, Britain, France, Germany and Japan, also as they form a model for their areas of cultural influence.

2. The UGC is an advisory committee appointed by the Chief Executive of the Hong Kong Special Administrative Region (SAR). Although the UGC has neither statutory nor executive powers, it administers public funds to the eight leading institutions of higher education in Hong Kong through its Secretariat, which is "staffed by civil servants."
3. The ideals of "academic freedom" derive from many sources: They were formalized as a pre-requisite of the research and teaching functions of the modern university by Wilhelm von Humboldt in the establishment of the University of Berlin in 1810. These ideals of "Lernfreiheit," the "freedom of inquiry, or advanced study," and "Lehrfreiheit," "the freedom to teach what one perceives to be the principles of one's special field," became institutional ideals not only of the German universities (until 1933, and again in the Federal Republic), but also, in a way, of the American graduate schools created on the German model. Intellectually, they derive from the same background of the European philosophers of the Renaissance and the Enlightenment that led to the creation of political institutions in the United States of America. (Cf. Flexner, 1967).
4. Importance of Educational Technology: All technology has to recommend itself to users to be adopted. There have been enormous changes in business and the professions, including education, as the result of improvements in technology in the last generation. Angela Castro of the Social Sciences Research Centre, of the University of Hong Kong writes on adoption of new technology: I do not believe professional development can be externally imposed on an individual, it must come from a personal prioritising of needs and values. If that passionate conviction is there, then the individual will seek ways to improve him/herself. (Castro, 1996) Even the authors of the "TLQPR Review" cannot resist referring to the fear of "Educational development units" being "cast in the role of 'teach police' " (TLQPR Review, 1996, p. 8).
5. Of course there are some who believe that, even in education, "the customer is always right." See: "Consumerism" in Appendix.
6. Other variables: sex of the student, sex of the instructor, personality of the student, and mood of the student, have also been studied in this context. More will be said about "personality" and "mood" of the student as they appear in Hong Kong student culture below.
7. Elements of Better Teaching Defined: e.g., breadth and depth of subject matter covered, development of understanding by students, amount and quality of such understanding retained, development of case material and textbooks, etc., and cooperation and collegiality between teachers and teachers and students.
8. The authors of the Basic Law (i.e., the mini-Constitution) of

Hong Kong, had the foresight to include reference to the concept of "academic freedom," which "institutions" may retain and enjoy:

> Art. 137: Educational institutions of all kinds may retain their autonomy and enjoy academic freedom. They may continue to recruit staff and use teaching materials from outside the Hong Kong Special Administrative Region. Schools run by religious organizations may continue to provide religious education, including courses in religion. Students shall enjoy freedom of choice of educational institutions and freedom to pursue their education outside the Hong Kong Special Administrative Region.

As is apparent, however, even with statutory protection of a specific right, it can not be foreseen how a court might interpret that right—or indeed whether a court might limit that right to what is immediately ascertainable within the four corners of Art. 137 itself.

9. With respect, this decision should not be written in stone either. On the one hand, what a faculty member ought to be able to bring to an institution is professional perspective on course design and grading standards. Yet, whereas a professional person should certainly enjoy a right to expression of professional opinion with respect to a grade, he or she cannot be said to have a right to create or destroy a career with that opinion. Even judicial decisions are subject to appeal.
10. On the application of agency theory, they refer to: Holmstron. B., and Milgrom, P. (1991); and Feltham, G., and Xie, J. (1994). For focus on quality over quantity, see also: Wruck, K., and Jensen, M. (1994); and Brickley, J., Smith, C.; Zimmerman, J.(1997).

References

Abrami, P.C.; Cohen, P.A.; d'Apollonia, S. (1988). Implementation Problems in Meta-Analysis, *Review of Educational Research* , 58: 151-79.

Aleamoni, L.M. and Graham, M.H. (1978). The Relationship between CEQ Ratings and Instructor's Rank, Class Size, and Course Level, *Journal of Educational Measurement* 34: 189-202.

Archibold, R.C. (1998). Payback Time: Give Me an 'A' or Else, *The New York Times*, Week in Review, May 24, 1998, and http:// www. nytimes. com/ library/ review/ 052498 students-evaluate-review.html, a review of the situation on American university campuses

Arubayi, Eric A. (1987). Improvement of Instruction and Teacher Effectiveness: Are Student Ratings Reliable and Valid? *Higher Education* 16: 267-78, at 270.

Arubayi, Eric (1985). Subject Disciplines and Student Evaluation of a Degree Programme in Education, *Higher Education* , 114: 683-91.

Avi-Itzhak, T. (1982). Teaching Effectiveness as Measured by Student Ratings and Instructor Self-Evaluation, *Higher Education* 11: 629-37.

Barnoski, R.P. and Sockloff, A.L. (1976). A Validation Study of the Faculty and Course Evaluation (FACE) Instrument, *Educational and Psychological Measurement* 36: 391- 400.

Brickley, James A. and Zimmerman, Jerold L. (1997). Changing Incentives in a Multitask Environment: The Case of Teaching at a Top-25 Business School. Working Paper, William E. Simon Graduate School of Business, University of Rochester.

Brickley, J.; Smith, C.; Zimmerman, J. (1997). *Managerial Economics and Organizational Architecture* , Irwin.

Byrne, J. and Leonhardt, D. (1996). The Best B-Schools, *Business Week* (October 21, 1996).

Castro, Angela (1996). Professional Development in IT for Tertiary Academics. Available: http://www.ugc.edu.hk/UGCweb/inte.

Clark, Burton R. (1995). *Places of Inquiry: Research and Advanced Education in Modern Universities,* Berkeley: University of California Press.

Corcos, R. Churchill, D. and Lam, A., (1998). Enhancing the Participation of Law Students in Academic Tutorials, in Kember, D.; Lam, D.-H.; Yan, L.; Yum, J.C.-K.; Liu, S.B., *Case Studies of Improving Teaching and Learning from the Action Learning Project,* Hong Kong: Hong Kong Polytechnic University, p. 358.

Crittenden, K.S.; Norr, J.L.; Lebailly, R.K. (1975). Size of University Classes and Student Evaluations of Teaching, *Journal of Higher Education* 46: 461-70.

Danielson, A.L. and White, R.A. (1976). Some Evidence on the Variables Associated with Student Evaluations of Teaching, *Journal of Higher Education* 7: 117-19.

Downie, N.W. (1952). Student Evaluation of Faculty, *Journal of Higher Education* 23: 495-96, 503.

Doyle, K and Whitely, S. (1974). Student Ratings as Criteria of Effective Teaching, *American Educational Research Journal* 11: 259-74.

Feltham. G., and Xie, J. (1994). Performance, Measure, Congruity, and

Diversity in Multitask Principal/Agent Relations, *Accounting Review* , 69: 429-53.

Franklin, J, and Theall, M. (1990). Communicating Student Ratings to Decision Makers: Design for Good Practice, in Theall, M. and Franklin, J., eds., *Student Ratings of Instruction: Issues for Improving Practice*, San Francisco: Jossey-Bass.

Flexner, Abraham (1967). *Universities: American, English, German*, with an introduction by Robert Ulich (New York: Teachers College Press, 1967).

Gage, N. L. (1974). Students' Ratings of College Teaching: Their Justification and Proper Use, in N.S. Glasman and B.R. Killait, eds., *Second UCSB Conference on Effective Teaching*, University of California at Santa Barbara, pp. 72-86.

Gage, N.L. (1961). The Appraisal of College Teaching: An Analysis of Ends and Means, *Journal of Higher Education* 32: 17-22.

Gayles, A.R. (1980). Student Evaluation in a Teacher Education Course, *Improving College and University Teaching* , 28: 128-31.

Gillmore, G.M. (1973). Estimates of Reliability Coefficients for Items and Subscales of the Illinois Course Evaluation Questionnaire, *Research Report, No. 341,* (Urbana, Ill.: Measurement and Research Division, Office of Instructional Resources, University of Illinois).

Gramlich, E. and Greenlee, G. (1993). Measuring Teaching Performance, *Journal of Economic Education* , Winter: 3-13 (based on a study of over 15,000 economics students who study the same subject matter, are graded in a common examination, and have instructors who each receive a similar form of student evaluation).

Greenwald, A.G. (1997). Validity Concerns and Usefulness of Student Ratings of Instruction, *American Psychologist,* 52:11: 1182-86, at p. 1184.

Greenwald, A.G. and Gillmore, G.M. (1997a). Grading Leniency is a Removable Contaminant of Student Ratings, *American Psychologist* 52:11: 1209-17.

Greenwald, A.G. and Gillmore, G.M.. (1997b). No Pain, No Gain? The Importance of Measuring Course Workload in Student Ratings of Instruction, *Journal of Educational Psychology* 89:4: 743-51.

Greenwald, A.G. and Gillmore, G.M.. (1997c). Supplement to UW's December 4 Press Release, http:// weber. u. washington. edu/ ~agg/ paingain/ supplement. Html.

Guthrie, E.R. (1954). *The Evaluation of Teaching: A Progress Report* (Seattle: University of Washington).

Hall, Cedric and Fizgerald, C. (1994). *Student Summative Evaluation of Teaching: Code of Practice* , Wellington, N.Z.: Association of University Staff of New Zealand.

Harris, E.L. (1982). Student Ratings of Faculty Performance: Should Departmental Committees Construct the Instruments, *Journal of Educational Research* 76: 100-106.

Harvard Crimson (1998). *Confi-Guide* 1/25/98, at http://www.thecrimson.harvard.edu/cgi-bin/ . . . .

Haskell, R.E. (1997a,b,c,d). Academic Freedom, Tenure and Student Evaluation of Faculty, *Education Policy Analysis Archives* 5: 6,17,18, 21; a: Galloping Polls in the 21st Century; b: Views from the Court; c: Accuracy and Psychometric Validity; d: Analysis and Implications of Views from the Courts in Relation to Academic Freedom, Standards, and Quality Instruction. Available: http://epaa.asu.edu/epaa/v5n6.html and v5n17.html, v5n18.html, v5n21.html.

Hillery, J.M. and Yuk, G.A. (1974). Convergent and Discriminant Validation of Student Ratings of College Instructors, *JSAS Catalog of Selected Documents in Psychology* 4: 26.

Hogan, T.P. (1973). Similarity of Student Ratings across Instructors, Course, and Time, *Research in Higher Education* 1: 149-54.

Holmstron, B., and Milgrom, P. (1991). Multitask Principal- Agent Analysis: Incentive Contracts, Asset Ownership and Job Design, *Journal of Law, Economics & Organization* (special issue) 7: 24-52.

Hong Kong University of Science & Technology (HKUST) (1998). Progress Report to the University Grants Committee (2 March, 1998). Available http://www.ust.hk/%7Ewebaa /TLQPR/account.htm, p.2.

Hong Kong University of Science & Technology (HKUST) (1998). Course Evaluations, http://www.ust.hk/~webaa/courseeval/#SOURCE.

Hong Kong University of Science & Technology (HKUST) (1997). *Faculty Handbook, 1997.*

Imrie, Bradford W. (1993). Professional Development as Quality Assurance, a summary of official thinking of the last 30 years leading up to the TLQPR. Available: http://www.ugc.edu.hk/documents/tlqpr/INQAAHE.html.

Joint University Programmes Admissions System (JUPAS) (1997). 1997 Admissions Grades Achieved by the 'Median' & 'Lower Quartile'

Applicants in Programmes Offered by the 7 Institutions.

Kaplin, W.A. and Lee, B. (1995). *The Law of Higher Education: A Comprehensive Guide to Legal Implications of Administrative Decision Making*, 3rd ed., San Francisco: Jossey Bass.

Kennedy, R.W. (1975). Grades Expected and Grades Received—their Relationship to Students' Evaluations of Faculty Performance, *Journal of Educational Psychology* 57: 109-15.

Kohlan, R.G. (1973). A Comparison of Faculty Evaluations Early and Late in the Course, *Journal of Higher Education* 44: 587-95.

Lee, O. (1998). 'Accountability' and 'Quality' in the Research Assessment Policy of the University Grants Committee (Hong Kong), which appeared as: Scrutineers perpetuate reign of error, in *The Times Higher Education Supplement*, August 14, 1998, Commonwealth Section, p. xiii.

Lee, O. (1997). *Hong Kong Business Law in a Nutshell*, New York: Juris Publishing, pp. xiii ff.

Lee, O. with the multimedia assistance of She, James (1999). 'I Want To See—Not To Be Seen!' Teaching 'Moot Court' Debating Skills through Interactive Multimedia, in James, Jeff, ed., *Quality in Teaching & Learning in Higher Education,* Hong Kong: University Grants Committee, 2000.

Marsh, H.W. (1983). Students as Evaluators of Teaching, *International Encyclopaedia of Education: Research and Studies* (New York: Pergamon Press).

Marsh, H.W. and Overall, J.U. (1981). The Relative Influence of Course Level, Course Type, and Instructor, on Students' Evaluations of College Teaching, *American Educational Research Journal* 18: 103-11.

Marsh, H.W. and Roche, L.A. (1997). Making Students' Evaluations of Teaching Effectiveness Effective: The Critical Issues of Validity, Bias, and Utility, *American Psychologist*, 52: 1187-97.

Massy, William F. and French, Nigel J. (1993). Teaching and Learning Quality Process Review: A Review of the Hong Kong Programme (hereafter, TLQPR Review), cited from the UGC internet home page, http://www.ugc.edu.hk/documents/papers/wfm_njf5.html.

McKeachie, Wilbert J. (1997a). Student Ratings: The Validity of Use, *American Psychologist* 52: 1218-25.

McKeachie, W.J. (1997b). Student Ratings of Faculty: A Reprise,

Academe, 65: 384-97.

Milgrom, P. and Roberts, J., (1995). Complementarities and Fit: Strategy, Structure, and Organizational Change in Manufacturing, *Journal of Accounting and Economics* , 19:179-208.

Murray, H.G. (1980). *Evaluating University Teaching: A Review of Research* , (Toronto: Confederation of University Faculty Associations).

Nichols, A, and Soper, J.C. (1972). Economic Man in the Classroom, *Journal of Political Economy*, 80: 1069- 73.

Payne, D.A. and Hobbs, A.M. (1979). The Effect of College Course Evaluation Feedback on Instructor and Student Perceptions of Instructional Climate and Effectiveness, *Higher Education* 8: 525-33.

Perry, R.R. and Baumann, R.R. (1973). Criteria for Evaluation of College Teaching: Their Reliability and Validity at the University of Toledo, in A.I Sockloff, ed., *Proceedings: Faculty Effectiveness as Evaluated by Students* (Philadelphia: Temple University).

Poch, R.K. (1993). Academic Freedom in American Higher Education: Rights, Responsibilities, and Limitations, *ASHE-ERIC Higher Education Report No. 4.*

Pohlmann, J.T. (1975). A Description of Teaching Effectiveness as Measured by Student Ratings, *Journal of Educational Measurement* 12: 49-54.

Quality Assurance Committee (QAC), City University of Hong Kong, (1996). CityU Policy and Guide for Developing Teaching Evaluation Schemes, http:// www. cityu. edu.hk/ QAC/ scheme. htm.

Quality Assurance Committee (QAC), City University of Hong Kong (1996). Statement of Principles on Quality and Quality Assurance. Available http://www.edu.hk/QAC/aboutQAC.htm

Riggs, R.O. (1975). The Prevalence and Purposes of Student and Subordinate Evaluations among AACTE Member Institutions, *Journal of Teacher Education* , 26: 218-21.

Rosenshine, B.; Cohen, A.; Furst, N. (1974). Correlates of Student Preference Ratings, *Journal of Economic Education* 4: 90-99.

Seldin, F. (1976). New Ratings Names for Professors, *The Peabody Journal of Education,* 53: 254-59.

Scott, C.A. (1977). Student Ratings and Instructor-Defined Extenuating Circumstances, *Journal of Educational Psychology* 69:

744-47.

Schwab, D.P. (1975). Course and Student Characteristic Correlates of the Course Evaluation Instrument, *Journal of Applied Psychology* 60: 742-47.

Scriven, M. (1981). Summative Teacher Evaluation, in J. Millman, ed., *Handbook of Teacher Evaluation*, Beverley Hills, CA: SAGE, pp. 244-71.

Sternberg, Robert J. (1998). Plenary Address, Practical Intelligence: Wisdom, Schooling, and Society, International Conference on the Application of Psychology to the Quality of Learning & Teaching, Hong Kong, June 13-18, 1998.

Sullivan, A. and Skanes, G. (1974). Validity of Student Evaluation of Teaching and the Characteristics of Successful Instructors, *Journal of Educational Psychology* 66: 584-90.

TLQPR of Hong Kong University of Science & Technology (1996).

University Grants Committee (UGC) (1996). Higher Education in Hong Kong—A Report by the University Grants Committee, Hong Kong, November, 1996.

University of Hong Kong (HKU), Department of Psychology, and Hong Kong University of Science & Technology (HKUST), Division of Social Science (1997). Call for Submissions, *International Conference on the Application of Psychology to the Quality of Learning and Teaching*, Hong Kong, June 13-18, 1998.

University of Hong Kong (HKU) (1997). Response to the TLQPR Report, http://www.hku.hk/acad/hku-tlqpr/response.htm , 11/23/97,

University of Washington (1997). Press Release: Student Evaluations Don't Get a Passing Grade: Easy-Grading Professors Get Too-High Marks, new UW Study Shcws, http://www. washington. edu/ newsroom/news/k120497.html .

Walker, B.D. (1968). An Investigation of Selected Variables Relative to the Manner in which a Population of Junior College Students Evaluate their Teachers, Diss., University of Houston, *Dissertation Abstracts*, 29: (1969) 3474B.

Weyrauch, W.O. (1971). The 'Basic Law' or 'Constitution' of a Small Group, *Journal of Social Issues* 27: 49, examining the emergence of community standards of behavior in a group isolated for a nutritional experiment, although many members saw themselves beyond culturally imposed rules.

Weyrauch, W.O. (1969). Governance Within Institutions, *Stanford Law Review* 22:141 (1969), reviewing Rubenstein and Lasswell, *The Sharing of Power in a Psychiatric Hospital* (1966).

Wong, C., (1984). Sociocultural Factors Counteract the Instructional Efforts of Teaching through English in Hong Kong, Seattle: University of Washington.

Woo Chia Wei (1997). The President's Progress Report, *HKUST Newsletter*, Fall, 1997, and email 9/10/97.

Wruck, K., and Jensen, M. (1994). "Science, Specific Knowledge, and Total Quality Management, *Journal of Accounting and Economics*, 18: 247-87.

Zirkel, P.A. (1996). *The Law of Teacher Evaluation*, Bloomington, IN: Phi Delta Kappa Educational foundation.

About the Author

Orlan Lee
Hong Kong University of Science & Technology
School of Business & Management; and
Visiting Fellow, Clare Hall, University of Cambridge

A.B. (hons.), Harvard; M.A., Yale; Ph.D., Freiburg (Germany); JurisDr., Pennsylvania; LL.M., Virginia. Dr. Lee teaches business law and cyberlaw in the School of Business and Management of the Hong Kong University of Science & Technology, and is also Visiting Fellow at Clare Hall, the college of advanced studies at the University of Cambridge. He is trained in both the civil law and common law systems, and as a social scientist, and has had extensive practical field experience. He has published widely on emergence of law and issues of public policy.

Acknowledgments

This article was prepared in part with support of a Direct Allocation Grant of the University Grants Committee of Hong Kong, and with the rare opportunity for research and reflection provided by Clare Hall, of the University of Cambridge. The author would like to thank Mr. Chan Tai Yat, and Mr. Ng Yiu Fai for their assistance in research and production of the manuscript.

Appendix
Divergent Findings

- **Those Duscussing the Conflict of Interest in Student Evaluation:**
  Gage, N. L. (1974);

Harris, E.L. (1982).

- **Those Studying the Widespread use of Student Evaluation for Formative and Summative Purposes:**
  In the 1970s, the American Council on Education surveyed 669 American colleges and universities and found 65% using such student ratings; 35% used these for so-called "summative" purposes, i.e., for faculty hiring, tenure, termination or promotion. See: Payne, D.A. and Hobbs, A.M. (1979).
  Obviously this form of questionnaire was even more at home in schools of teacher education, where 86% of the American Association of Colleges for Teacher Education (AACTE) reported using these measures. See: Riggs, R.O. (1975).

- **Those Advocating "Consumerism" in Education:**
  Seldin, F. (1976);
  Gayles, A.R. (1980);
  Arubayi, Eric (1985).

- **Those Attributing High Rating to Impact of Prior Interest in Subject:**
  Marsh, H.W. (1980);
  Greenwald, A.G. (1997).

- **Those Believing that Ratings are Consistent for the Same Faculty Members from Year-to-Year:**
  Guthrie, E.R. (1954).

- **Those Finding that Smaller Class Size Produced Higher Ratings:**
  Danielson, A.L. and White, R.A. (1976);
  Crittenden, K.S.; Norr, J.L.; Lebailly, R.K. (1975);
  Scott, C.A. (1977);
  Perry, R.R. and Baumann, R.R. (1973);
  Avi-Itzhak, T. (1982).

- **Those Still Arguing that Class Size Has NO Effect:**
  Aleamoni, L.M. and Graham, M.H. (1978).

- **Those Finding Student Ratings Correlate with Professional and Alumni Evaluation:**
  Marsh, H.W. (1983);
  Murray, H.G. (1980).

- **Those Finding that Time of Day Affects the Survey (Afternoon Ratings Lower than Morning):**
  Nichols, A, and Soper, J.C. (1972).

- **Those Finding that Lecturers are Rated Lower than Professors:**

Downie, N.W. (1952);
Gage, N.L. (1961);
Walker, B.D. (1968).

- **Those Finding that Students at Lower Levels Tend to Rank Lecturers Less Favorably than Professors:**
  Downie, N.W. (1952);
  Gage, N.L. (1961);
  Pohlmann, J.T. (1975);
  Kohlan, R.G. (1973).

- **Those Finding that Students at Lower Levels Do NOT Tend to Rank Lecturers Less Favorably than Professors:**
  Hillery, J.M. and Yuk, G.A. (1974).

- **Those Finding that "Grades Expected" Affect Ratings:**
  Barnoski, R.P. and Sockloff, A.L. (1976);
  Kennedy, R.W. (1975);
  Schwab, D.P. (1975);
  Sullivan, A. and Skanes, G. (1974);
  Hillery, J.M. and Yuk, G.A. (1974);
  Perry, R.R. and Baumann.R.R. (1973);
  Rosenshine, B.; Cohen, A.; Furst, N. (1974).

- **Those Finding that "Grades Expected" Do NOT Affect Ratings:**
  Doyle, K and Whitely, S. (1974).

- **Those Finding that Ratings Are Consistent for the Same Faculty Members Regardless of Subject Matter Taught:**
  Marsh, H.W. and Overall, J.U. (1981);
  Gillmore, G.M. (1973);
  Hogan, T.P. (1973).

- **Those Finding that Teaching Ratings and Learning are Only "Weakly Related":**
  Gramlich, E. and Greenlee, G. (1993).

- **Those Who Surveyed the Literature on Validity:**
  Arubayi, Eric A. (1987);
  McKeachie, W.J. (1997b).
  Haskell, R.E. (1997a, b, c, d).

- **Current Research Returning to the Conclusion that Grades Expected and Course Workload are Dominant Factors:**
  Greenwald, A.G. (1997);
  Greenwald, A.G. and Gillmore, G.M. (1997a);
  Greenwald, A.G. and Gillmore, G.M.. (1997b);
  University of Washington (1997);
  Greenwald, A.G. and Gillmore, G.M.. (1997c);

Archibold, R.C. (1998).

- **Those Discussing the Disparity in the Concepts of Teaching and Learning:**
  Lee, O. with She, James, (2000);
  Haskell, R.E. (1997a,b,c,d).

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
New York University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers

Jay D. Scribner

Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

University of Texas at Austin

Robert E. Stake
University of Illinois—UC

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografía e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 1900 times since October 26, 2000

Education Policy Analysis Archives

Volume 8 Number 49 October 26, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

What Do Test Scores in Texas Tell Us?

Stephen P. Klein
Laura S. Hamilton
Daniel F. McCaffrey
Brian M. Stecher

RAND

Related articles:
Haney Vol 8 No 41
Camilli Vol 8 No 42

Abstract
We examine the results on the Texas Assessment of Academic Skills (TAAS), the highest-profile state testing program and one that has recorded extraordinary recent gains in math and reading scores. To investigate whether the dramatic math and reading gains on the TAAS represent actual academic progress, we have compared these gains to score changes in Texas on another test, the National Assessment of Educational Progress (NAEP). Texas students did improve significantly more on a fourth-grade NAEP math test than their counterparts nationally. But, the size of this gain was smaller than their gains on TAAS and was not present on the eighth-grade math test. The stark differences between the stories told by NAEP and TAAS are especially striking when it comes to the gap in average scores between whites and students of color. According to the NAEP results, that gap in Texas is not only very large but increasing slightly. According to TAAS scores, the gap is much smaller and decreasing greatly. Many schools are devoting a great deal of class time to highly specific TAAS preparation. While this preparation may improve TAAS scores, it may not help students develop necessary reading and math skills. Schools with relatively large percentages of minority and poor students may be doing this more than other schools. We raise serious questions about the validity of those gains, and caution against the danger of making decisions to sanction or reward students, teachers and schools on the basis of test scores that may be inflated or misleading. Finally, we suggest some steps that states can take to increase the likelihood that their test results merit public confidence and provide a sound basis for educational policy.

Introduction

BEST COPY AVAILABLE

During the past decade, several states have begun using the results on statewide tests as the basis for rewarding and sanctioning individual students, teachers, and schools. Although testing and accountability are intended to improve achievement and motivate staff and students, concerns have been raised in both the media and the professional literature (e.g., Heubert & Hauser, 1999; Linn, 2000) about possible unintended consequences of these programs.

The high-stakes testing program in Texas has received much of this attention in part because of the extraordinarily large gains the students in this state have made on its statewide achievement tests, the Texas Assessment of Academic Skills (TAAS). In fact, the gains in TAAS reading and math scores for both majority and minority students have been so dramatic that they have been dubbed the "Texas miracle." However, there are concerns that these gains were inflated or biased as an indirect consequence of the rewards and sanctions that are attached to the results. Thus, although there is general agreement that the gains on the TAAS are attributable to Texas' high-stakes accountability system, there is some question about what these gains mean. Specifically, do they reflect a real improvement in student achievement or something else?

We conducted several analyses to examine the issue of whether TAAS scores can be trusted to provide an accurate index of student skills and abilities. First, we used scores on the reading and math tests that are administered as part of the National Assessment of Educational Progress (NAEP) to investigate how much students in Texas have improved and whether this improvement is consistent with what has occurred nationwide. NAEP scores are a good benchmark for this purpose because they reflect national content standards and they are not subject to the same external pressures to boost scores as there are on the TAAS.

Next, we assessed whether the gains in TAAS scores between 1994 and 1998 were comparable to those on NAEP. We did this to examine how much confidence can be placed in the TAAS score gains. Similarly, we measured whether the differences in scores between whites and students of color on the TAAS were consistent with the differences between these groups on NAEP. Specifically, is the gap on TAAS credible given the gap on NAEP? And finally, we investigated whether TAAS scores are related to the scores on a set of three other tests that we administered to students in 20 Texas elementary schools.

Our findings from this research raise serious questions about the validity of the gains in TAAS scores. More generally, our results illustrate the danger of relying on statewide test scores as the sole measure of student achievement when these scores are used to make high-stakes decisions about teachers and schools as well as students. We anticipate that our findings will be of interest to local, state, and national educational policymakers, legislators, educators, and fellow researchers and measurement specialists.

Readers also may be interested in a RAND study by Grissmer et al. (2000) that compared the NAEP scores of different states across the country. Grissmer and his colleagues found that after controlling for

various student demographic characteristics and other factors, Texas tended to have higher NAEP scores than other states and there was some speculation as to whether this was due to the accountability system in Texas. Thus, while the Grissmer et al. (2000) report and the research presented in this issue paper both used NAEP scores, these studies differed in the questions they investigated, the data they analyzed, and the methodologies they employed. A forthcoming RAND issue paper will discuss some of the broader policy questions about high-stakes testing in schools.

Background

Scores on achievement tests are increasingly being used to make decisions that have important consequences for examinees and others. Some of these "high-stakes" decisions are for individual students--such as for tracking, promotion, and graduation (Heubert & Hauser, 1999). Some states and school districts also are using test scores to make performance appraisal decisions for teachers and principals (e.g., merit pay and bonuses) and to hold schools and educational programs accountable for the success of their students (Linn, 2000). Although the policymakers who design and implement such systems often believe they lead to improved instruction, there is a growing body of evidence which indicates that high-stakes testing programs can also result in narrowing the curriculum and distorting scores (Koretz & Barron, 1998; Koretz et al., 1991; Linn, 2000; Linn, Graue, & Sanders, 1990; Stecher, Barron, Kaganoff, & Goodwin, 1998). Consequently, questions are being raised about the appropriateness of using test scores alone for making high-stakes decisions (Heubert & Hauser, 1999).

In this issue paper, we examine score gains on one statewide test in an effort to assess the degree to which they provide valid information about student achievement in that state and about improvements in achievement over time. This investigation is the latest in a decade-long series of RAND studies of high-stakes testing (e.g., Koretz & Barron, 1998). We believe that this work will provide lessons to help policymakers understand some of the challenges that arise in the context of high-stakes accountability systems.

Our interest in Texas was prompted by an unusual empirical relationship we observed between scores on TAAS and tests we administered to students in a small sample of schools as part of a larger study on teaching practices and student achievement. Because our set of schools was small and not representative of the state, we decided to explore statewide patterns of achievement on TAAS and on NAEP. In addition, Texas provides an ideal context in which to study high-stakes testing because its accountability system has received attention from the media and from the policy community, and it has been cited as possibly contributing to improved student achievement (e.g., Grissmer & Flanagan, 1998; Grissmer et al., 2000). TAAS scores are a central component of the accountability system. For example, students must pass the TAAS to graduate from high school, and TAAS

scores affect performance evaluations (and, in some cases, compensation) for teachers and principals.

The TAAS program has been credited not only with improving student performance, but also with reducing differences in average scores among racial and ethnic groups. For example, a recent press release announced a record high passing rate on the TAAS. According to Commissioner of Education Jim Nelson, "Texas has justifiably gained national recognition for the performance gains being made by our students." Nelson also stated that Texas has "been able to close the gap in achievement between our minority youngsters and our majority youngsters, and we've again seen how we're progressing in that regard" (Jim Nelson as quoted by Mabin, 2000).

The unprecedented score gains on the TAAS have been referred to as the "Texas miracle." However, some educators and analysts (e.g., Haney, 2000) have raised questions about the validity of these gains and the possible negative consequences of high-stakes accountability systems, particularly for low-income and minority students. For example, the media have reported concerns about excessive teaching to the test, and there is some empirical support for these criticisms (Carnoy, Loeb, & Smith, 2000; McNeil & Valenzuela, 2000; Hoffman et al., in press). For instance, teachers in Texas say they are spending especially large amounts of class time on test preparation activities. Because the length of the school day is fixed, the more time that is spent on preparing students to do well on the TAAS often means there is less time to devote to other subjects.

There are also concerns that score trends may be biased by a variety of formal and informal policies and practices. For example, policies about student retention in grade may affect score trends (McLaughlin, 2000). States may vary in the extent to which their schools promote students who fail to earn acceptable grades and/or statewide test scores. Eliminating these so-called "social promotions" would most likely raise the average scores at each grade level in subsequent years while lowering it at each age level. This is likely to occur because although the students who are held back may continue to improve, they are likely to do so at a slower rate than comparable students who graduate with their classmates (Heubert & Hauser, 1999). Another concern is inappropriate test preparation practices, including outright cheating. There have been documented cases of cheating across the nation, including in Texas. If widespread, these behaviors could substantially distort inferences from test score gains (Hoff, 2000; Johnston, 1999).

The pressure to raise scores may be felt most intensely in the lowest-scoring schools, which typically have large populations of low-income and minority students. Students at these schools may be particularly likely to suffer from overzealous efforts to raise scores. For example, Hoffman et al. (in press) found that teachers in low-performing schools reported greater frequency of test preparation than did teachers in higher-performing schools. This could lead to a superficial appearance that the gap between minority and majority students is narrowing when no change has actually occurred.

Evidence regarding the validity of score gains on the TAAS can be obtained by investigating the degree to which these gains are also present on other measures of these same general skills. Specifically, do the score trends on the TAAS correspond to those on the highly regarded NAEP? The NAEP tests are generally recognized as the "gold standard" for such comparisons because of the technical quality of the procedures that are used to develop, administer, and score these exams. Of course, NAEP is not a perfect measure. For example, there are no stakes attached to NAEP scores, and therefore student motivation may differ on NAEP and state tests, such as TAAS. However, it is currently the best indicator available.

There are several other reasons why score gains on the TAAS are not likely to have a one-to-one match with those on NAEP if these tests assess different skills and knowledge. However, the specifications for the NAEP exams are based on a consensus of a national panel of experts, including educators, about what students should know and be able to do. Hence, NAEP provides an appropriate benchmark for measuring improvement. As Linn (2000) notes, "Divergence of trends does not prove that NAEP is right and the state assessment is misleading, but it does raise important questions about the generalizability of gains reported on a state's own assessment, and hence about the validity of claims regarding student achievement" (p. 14).

Questions for Our Research

Understanding the source and consequences of the impressive score gains on the TAAS would require an extensive independent study. We have not done that. Instead, the analyses described below address the following questions about student achievement in Texas:

1. Have the reading and math skills of Texas students improved since the full statewide implementation of the TAAS program in 1994 (e.g., are fourth graders reading better today than fourth graders a few years ago); and, if their skills did improve: (a) how much improvement occurred and (b) was the amount of improvement in reading the same as it was in math?

2. Are the gains in reading and math on the TAAS consistent with what would be expected given NAEP scores in Texas and the rest of the country?

3. Has Texas narrowed the gap in average reading and math skills between whites and students of color?

4. Do other tests given in Texas at a sample of 20 schools produce results that are consistent with those obtained with the TAAS?

We begin by describing certain important features of the TAAS and NAEP exams. We then answer the first three questions through

analyses of publicly available TAAS and NAEP data and discuss the findings. Next, we answer the fourth question by reporting the results from a study that administered other tests to about 2,000 Texas students. Finally, we present our conclusions.

Description of the TAAS

TAAS was initiated in 1990 to serve as a criterion-referenced measure of the state's mandated curriculum. It is intended to be comprehensive and to measure higher-order thinking skills and problem-solving ability (Texas Education Agency, 1999). Since the full implementation of the TAAS program in 1994, it has been administered in reading and mathematics in grades 3, 4, 5, 6, 7, 8, and 10. Other subjects are also tested at selected grade levels. Last year, for example, a writing test was given at grades 4, 8, and 10. Science and social studies were tested at grade 8. The TAAS tests consist primarily of multiple-choice items, but the writing test includes questions that require written answers.

Teachers administer the TAAS tests to their own students. Answers are scored by the state. The questions are released to the public after each administration of the exam, and a new set of TAAS tests is administered each year. However, the format and content of the questions in one year are very similar to those used the next year. Each form of the TAAS contains items that are being field-tested for inclusion in the forms to be used in subsequent years. These items are also used to link test scores from one year to the next to help ensure consistent difficulty over time. These experimental items are not used to compute student scores nor are they released to the public. This practice is consistent with that employed in many other large-scale testing programs.

The TAAS is administered only in Texas. Thus, there are no national norms or benchmarks against which to compare the performance of Texas students on this test. However, the Texas Education Agency administered the Metropolitan Achievement Tests to a sample of Texas students to determine how well these students performed relative to a national norm group. We discuss this study in a later section of this issue paper.

Description of NAEP

The national portion of NAEP is mandated by Congress and is administered through the National Center for Education Statistics. It is currently the only assessment that provides information on the knowledge and skills of a representative sample of the nation's students. The content of NAEP tests is based on test specifications that were developed by educators and others, and is intended to reflect a consensus about what students should be learning at a given grade level. Hence, the questions are not tied to standards of a single state or district. (Note 1) Like TAAS, NAEP is designed to assess

problem-solving skills in addition to content knowledge. A national probability sample of schools is invited to participate in NAEP. Schools that decline are replaced with schools where the student characteristics are similar to those at the schools that refused to participate.

Most states, including Texas, also arrange to have the NAEP exams administered to another (and larger) group of their schools to allow for the generation of reliable state-level results. This state-level testing utilizes the same general procedures as the national NAEP program does; e.g., third-party selection of the participating schools and having a cadre of trained consultants (rather than classroom teachers) administer the tests. However, unlike the national program, these consultants may be local district personnel.

In both the national and state-level programs, a given student is asked a sample of all the questions that are used at that student's grade level. This permits a much larger sampling of the content domain in the available testing time than would be feasible if every student had to answer every item. Different item formats (including multiple-choice, short-answer, and essay) are used in most subjects. The breadth of content and item types, as well as the consensus of a national panel of experts that is reflected in NAEP frameworks, makes NAEP a useful indicator of achievement trends across the country.

The validity of NAEP scores is enhanced by the procedures that are used to give the exams and ensure test security (e.g., test administrators do not have a stake in the outcomes). However, the utility of NAEP scores is limited by some of the other features of this testing program. For instance, NAEP is not administered every year, and when it is administered, not every subject is included, only a few grade levels are tested, and individual student, school, and district scores are not available. These features preclude examining year-to-year trends in a particular subject or tracking individual student progress over time. The motivation to do well on the NAEP tests is intrinsic rather than driven by external stakes. However, any reduction in student effort or performance that may stem from NAEP being a relatively low-stakes test should be fairly consistent over time and therefore not bias our measurement of score improvements across years.

How We Report Results

NAEP and TAAS results are typically reported to the public in terms of the percentage of students passing or meeting certain performance levels (or "cut" scores). Although this type of reporting seems easier to understand, it can lead to erroneous conclusions. For example, the difficulty of achieving a passing status or a certain level of performance (such as "proficient") may vary between tests as well as within a testing program over time. Making comparisons based on percentages reaching certain levels also does not account for score changes among students who perform well above or below the cut score.

To avoid these and other problems with percentages, we adopted the research community's convention of reporting results in terms of "effect" sizes. The effect size is the difference in mean scores (between years or groups) divided by the standard deviation of those scores. In other words, it is the standardized mean difference. The major advantage of using effect sizes is that they provide a common metric across tests.

As a frame of reference for readers who are not familiar with this metric, the effect size for the difference in achievement between white and black students has ranged from 0.8 to 1.2 across a variety of large-scale tests (Hedges & Nowell, 1998). The effect size for the difference in third grade student reading scores between large and small classes in Tennessee was approximately 0.25 (Finn & Achilles, 1999). (Note 2)

Have Reading and Math Skills Improved in Texas?

NAEP data have been cited as evidence of the effectiveness of educational programs in Texas (e.g., Grissmer & Flanagan, 1998). For instance, within a racial or ethnic group, the average performance of the Texas students tends to be about six percentile-points higher than the national average for that group (Grissmer et al., 2000; Reese et al., 1997).

These results are consistent with the findings obtained by the Texas Education Agency in its 1999 Texas National Comparative Data Study, in which a sample of Texas students took the Metropolitan Achievement Tests, Seventh Edition (MAT-7). Texas students at every grade level scored slightly higher than the national norming sample in most subjects (Texas Education Agency, 1999). However, it is difficult to draw conclusions from this study because, according to the sampling plan for this research, each participating school selected the classrooms and students that would take the MAT. Moreover, Texas did not report the mean TAAS scores of the students who took the MAT. Under the circumstances, the TAAS data are vital for determining whether those who took the MAT were truly representative of their school or the state. For example, the interpretation of the MAT findings would no doubt change if it was discovered that the mean TAAS scores of the students who took the MAT were higher than the corresponding state mean TAAS scores.

Data from a single year cannot tell us whether achievement has improved over time or whether trends in TAAS scores are reflected in other tests. To answer the question of whether performance improved, we compared the scores of Texas fourth graders in one year with the scores of Texas fourth graders four years later. We did this in both reading and mathematics. We also did this for eighth graders in mathematics (NAEP's testing schedule precluded conducting a similar analysis for eighth graders in reading). We then contrasted these results with national trends to assess whether the gains in Texas after the full statewide implementation of the TAAS differed from those in other states.

Figures 1 through 3 present the results of these analyses. The main finding is that over a four-year period, the average test score gains on the NAEP in Texas exceeded those of the nation in only one of the three comparisons, namely: fourth grade math.

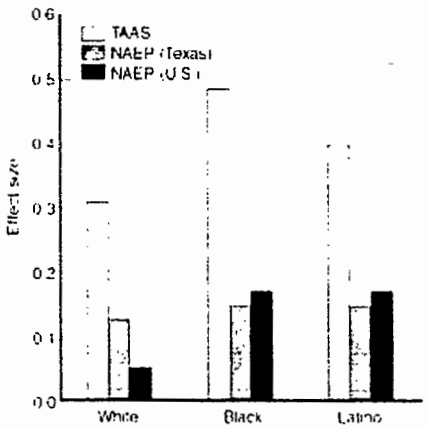


Figure 1
Reading Effect Sizes for 4th Graders on NAEP

Figure 1 shows that the Texas fourth graders in 1998 had higher NAEP reading scores than did Texas fourth graders in 1994. The size of the increase was .13 standard deviation units for white students and .15 units for students of color. However, these increases were not unique to Texas. The national trend was for all students to improve. In fact, only among white fourth graders was the improvement in Texas greater than improvement nationally, and then only slightly (the difference in the effect sizes between Texas and the United States was .08). We discuss the implications of this difference in score gains between groups when we discuss the question of whether Texas has narrowed the gap in performance among racial and ethnic groups.

The TAAS data tell a radically different story (see Figure 1). They indicate there was a very large improvement in TAAS reading scores for all groups (effect sizes ranged from .31 to .49). Figure 1 also shows that on the TAAS, black and Hispanic students improved more than whites. The gains on TAAS were therefore several times larger than they were on NAEP. And, contrary to the NAEP findings, the gains on TAAS were greater for students of color than they were for whites.

Figure 2 shows that fourth graders in Texas in 1996 had substantially higher NAEP math scores than did fourth graders in 1992 (effect sizes ranged from .25 to .43). Moreover, this improvement was substantially greater than the increase nationwide. This was especially true for white students. Nevertheless, the gains on TAAS were much larger than they were on NAEP, especially for students of color. (Note 3)

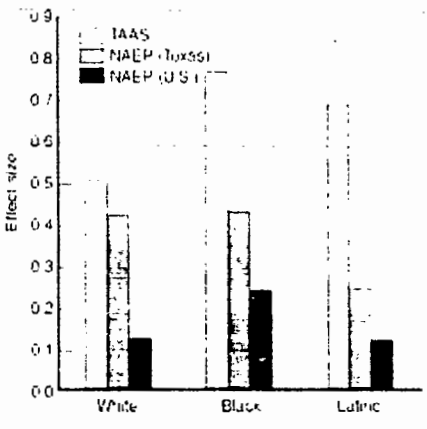


Figure 2
Math Effect Sizes for 4th Graders on NAEP Between 1992 and 1996, and on TAAS Between

Figure 3 shows that Texas eighth graders in 1996 had higher NAEP scores than did Texas eighth graders in 1992, but these differences were only slightly larger than those observed nationally.

BEST COPY AVAILABLE

Thus, as with fourth grade reading, there was nothing remarkable about the NAEP scores in Texas, and students of color did not gain more than whites. In contrast, there were huge improvements in eighth grade math scores on the TAAS during a similar four-year period, and these increases were much larger for students of color than they were for whites. The same was true for eighth grade TAAS reading scores during this period (effect sizes for whites, blacks, and Hispanics were .28, .45, and .37, respectively).

To further examine the question of whether there has been an improvement in reading and math skills of Texas students, we compared the NAEP scores of fourth graders in one year with the NAEP scores of eighth graders four years later. Because of the way NAEP samples students for testing, this is analogous (but not equivalent) to following the same cohort of students over time. In fact, the redesign of NAEP in 1984, which established a practice of testing grade levels four years apart and conducting the assessment in the core subjects every four years, was intended in part to support this type of analysis (Barton & Coley, 1998). We present results for Texas and the nation so readers can see the extent to which Texas students are progressing relative to students in other states.

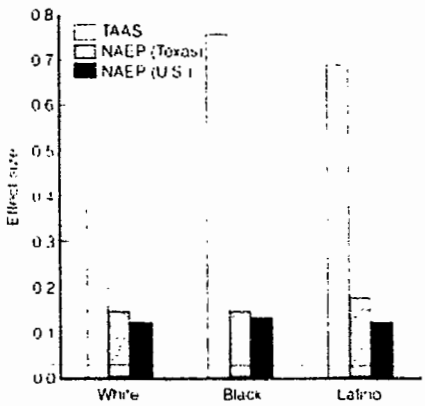


Figure 3
Math Effect Sizes for 8th Graders on NAEP Between 1992 and 1996, and on TAAS Between [illegible]

Table 1 shows that the average NAEP math scale score for white Texas fourth graders in 1992 was 229. Four years later, the mean score for white eighth graders was 285, i.e., a 56-point improvement. However, there was a 54-point improvement nationally for whites during this same period. There was a similar pattern for minority students, and these trends held for both math and reading (Table 2). In short, the score increases in Texas were almost identical to those nationwide (we could not conduct the corresponding analysis with TAAS data because TAAS does not convert scores to a common scale across grade levels).

BEST COPY AVAILABLE

Table 1
Mean NAEP Math Scores for

4th Graders in 1992 and 8th Graders in 1996

Group	Texas			United States			Texas – U.S.
	4th	8th	Gain	4th	8th	Gain	
White	229	285	56	227	281	54	2
Black	199	249	50	192	242	50	0
Hispanic	209	256	47	201	250	49	-2

Table 2
Mean NAEP Reading Scores for
4th Graders in 1994 and 8th Graders in 1998

Group	Texas			United States			Texas – U.S.
	4th	8th	Gain	4th	8th	Gain	
White	227	273	46	223	270	47	-1
Black	191	245	54	186	241	55	-1
Hispanic	198	252	54	188	243	55	-1

Is Texas Closing the Gap Between Whites and Students of Color?

In 1998, the mean fourth grade NAEP reading score for whites in Texas was one full standard deviation higher than the mean for blacks. To put this in perspective, the average black student was at roughly the 38th percentile among all Texas test takers whereas the average white student was at about the 67th percentile. This gap was slightly larger than the difference between these groups in 1994. In other words, the black-white reading gap actually increased during this four-year period. The same pattern was present in fourth and eighth grade math scores (see Figure 4a).

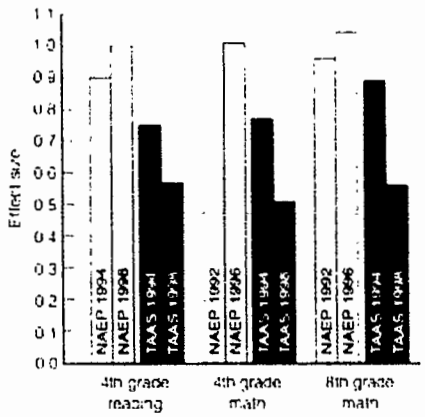


Figure 4a
Difference Between White and Black Mean Score in Texas on NAEP and TAAS (in Standard

In contrast, the difference in mean TAAS scores between whites and blacks was initially smaller than it was on NAEP, and it decreased substantially over a comparable four-year period. Consequently, by 1998, the black-white gap on TAAS was about half what it was on NAEP. In other words, whereas the gap on NAEP was large to begin with and got slightly wider over time, the gap on TAAS started off somewhat smaller than it was on NAEP and then got substantially smaller.

The same radically disparate NAEP and TAAS trends were also present for the Hispanic-white gap; i.e., the gap got slightly wider on NAEP but substantially smaller on TAAS over comparable four-year periods (see Figure 4b). In addition, although fourth grade math was the subject on which Texas showed the largest gains over time relative to the nation, the white-Hispanic NAEP gap grew in Texas but not nationally, and the white-black gap remained constant in Texas but actually shrank nationally. In short, gap sizes on NAEP were moving in the opposite direction than they were on TAAS.

BEST COPY AVAILABLE

It is worth noting that even the relatively small NAEP gains we observed might be somewhat inflated by changes in who takes the test. As mentioned earlier, Haney (2000) provides evidence that exclusion of students with disabilities increased in Texas while decreasing in the nation, and Texas also showed an increase over time in the percentage of students dropping out of school and being held back. All of these factors would have the effect of producing a gain in average test scores that overestimates actual changes in student performance.

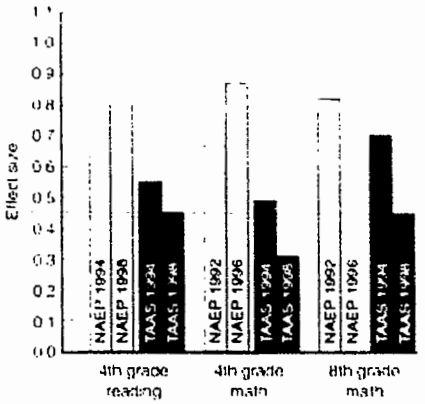


Figure 4b
Difference Between White and Hispanic Mean Scores in Texas on NAEP and TAAS in Standard

Why Do TAAS and NAEP Scores Behave So Differently?

The large discrepancies between TAAS and NAEP results raise serious questions about the validity of the TAAS scores. We do not know the sources of these differences. However, one plausible explanation, and one that is consistent with some of the survey and observation results cited earlier, is that many schools are devoting a great deal of class time to highly specific TAAS preparation. It is also plausible that the schools with relatively large percentages of minority and poor students may be doing this more than other schools.

TAAS questions are released after each administration. Although there is a new version of the exam each year, one version looks a lot like another in terms of the types of questions asked, terminology and graphics used, content areas covered, etc. Thus, giving students instruction and practice on how to answer the specific types of questions that appear on the TAAS could very well improve their scores on this exam. For example, in an effort to improve their TAAS scores, some schools have retained outside contractors to work with teachers, students, or both.

If the discrepancies we observed between NAEP and TAAS were due to some type of focused test preparation for the TAAS, then this instruction must have had a fairly narrow scope. With the possible exception of fourth grade math, it certainly did not appear to influence NAEP scores. In short, if TAAS scores were affected by test preparation for the TAAS, then the effects of this preparation did not appear to generalize to the NAEP exams. This explanation also raises questions about the appropriateness of what is being taught to prepare students to take the TAAS.

A small but significant percentage of students may have "topped out" on the TAAS. In other words, their TAAS scores may not reflect just how much more proficient they are in reading and math than are other students. If that happened, it would artificially narrow the gap on the TAAS between whites and students of color (because majority

BEST COPY AVAILABLE

students tend to earn higher scores than minority students). Thus, the reduced gap on the TAAS relative to NAEP may be an artifact of the TAAS being too easy for some students. (Note 4) If so, it also would deflate the gains in TAAS scores over time. In short, were it not for any topping-out, the TAAS gain scores in Figures 1 through 3 would have been even larger, which in turn would further increase the disparity between TAAS and NAEP results.

What Happens on Other Tests?

We collected data on about 2,000 fifth graders from a mix of 20 urban and suburban schools in Texas. This study was part of a much larger project that included administering different types of science and math tests to students who also took their state's exams. The 20 schools were from one part of Texas. They were not selected to be representative of this region let alone of Texas as a whole. Nevertheless, some of the results at these schools also raised questions about the validity of the TAAS as a measure of student achievement.

Test Administration

In the spring of 1997, our Texas students took the English language version of the TAAS in reading and math. A few weeks later, we administered the following three tests to these same students: the Stanford 9 multiple-choice science test, the Stanford 9 open-ended (OE) math test, and a "hands-on" (HO) science test developed by RAND (Stecher & Klein, 1996). The Stanford 9 OE math test asked students to construct their own answers and write them in their test booklets. In the HO science test, students used various materials to conduct experiments. They then wrote their answers to several open-ended questions about these experiments in a simulated laboratory notebook. Table 3 shows the means and standard deviations on each measure.

Some Expected and Unexpected Findings

We analyzed the data in two ways. First, we investigated whether the students who earned high scores on one test tended to earn high scores on the other tests. Next, we examined whether the schools that had a high average score on one test tended to have high average scores on the other tests. We also looked at whether the results were related to type of test used (i.e., multiple-choice or open-ended), subject matter tested (reading, math, or science), and whether a student was in a free or reduced-price school lunch program. The latter variable serves as a rough indicator of a student's socioeconomic status (SES). For the school-level analyses, SES was indicated by the percentage of students at the school who were in the subsidized lunch program.

Table 3
Means and Standard Deviations on Supplemental Study

Measures by Unit of Analysis

Variable	Students		Schools	
	Mean	Standard Deviation	Mean	Standard Deviation
TAAS math	37.97	13.62	38.84	3.80
TAAS reading	29.33	10.61	29.61	2.59
Stanford 9 science	29.01	5.40	28.55	1.94
Stanford 9 OE math	15.14	5.21	14.84	1.44
HO science	11.78	6.00	11.44	1.83
Percentage in lunch program (SES)	67.84	46.7	76.10	22.3

Notes: TAAS math had 52 items and TAAS reading had 40 items. Stanford 9 science had 40 items. The maximum possible scores on on Stanford 9 OE math and HO science were 27 and 30, respectively.

Some of our results were consistent with those in previous studies. Others were not. We begin with what was consistent and then turn to those that were anomalous.

The first column of Table 4 shows the correlation between various pairs of measures when the student (N approx. 2,000) is the unit of analysis. (Note 5) The second column shows the results when the school (N = 20) is the unit of analysis. The first set of rows show that the measures we administered correlated about .55 with each other when the student was the unit of analysis. These correlations were substantially higher when the school was the unit. For example, the correlation between Stanford 9 science and Stanford 9 OE math was .55 when the student was the unit, but it was .78 when the school was the unit. These results are very consistent with the general findings of other research on student achievement.

Table 4
Correlations Between Measures

Correlations between:	Unit of Analysis	
	Students	**Schools**
Non-TAAS tests		
• Stanford 9 science and HO science	.57	.88
• Stanford 9 science and Stanford 9 OE math	.55	.78
• Stanford 9 OE math and HO science	.53	.71
SES and non-TAAS tests		
• SES and Stanford 9 science	–.17	–.76
• SES and Stanford 9 OE math	–.10	–.72
• SES and HO science	–.18	–.66
SES and TAAS tests		
• SES and TAAS math	–.08	.13
• SES and TAAS reading	–.14	-.21
TAAS and non-TAAS tests		
• TAAS math and Stanford 9 science	.48	–.07
• TAAS math and Stanford 9 OE math	.46	.02
• TAAS math and HO science	.48	.03
• TAAS reading and Stanford 9 science	.52	.10
• TAAS reading and Stanford 9 OE math	.42	.21
• TAAS reading and HO science	.53	.13
TAAS math and TAAS reading	.81	.85

The second set of rows in Table 4 shows a strong negative correlation between the percentage of students at a school who were in the lunch program and that school's mean on the tests we administered. In other words, schools with more affluent students tended to earn higher mean scores on the non-TAAS tests than did schools with less wealthy students. This relationship is present regardless of test type (multiple-choice or open-ended) and subject matter (math or science). Again, these findings are very consistent with those found in other testing programs.

The correlation between SES and our test scores is much stronger when the school is used as the unit of analysis than when the student is the unit. This is a common finding and stems in part from the fact that it is difficult to get a high correlation with a dichotomous variable (i.e., in program versus not in program). The school-level analyses do not suffer from this problem because SES at the school level is measured by the percentage of students at the school who are in the program (i.e., a continuous rather than a dichotomous variable). School-level analyses also tend to produce higher correlations than individual-level analyses

because aggregation of scores to the school level reduces the percentage of error in the estimates.

The anomalies appear in the third and fourth sets of rows. In the third set, SES had an unusually small (Pearson) correlation with both of the TAAS scores even when the school was used as the unit of analysis. (Note 6) This result (which is opposite to the one we found with the non-TAAS tests) was due to a curvilinear relationship between SES and TAAS scores. Specifically, schools with a relatively low or high percentage of students in the lunch program tended to have higher mean TAAS math scores than did schools with an average percentage of students in this program (see Figure 5). Thus, the typical relationship between SES and test scores disappeared on the TAAS even though this relationship was present on the tests we administered a few weeks after the students took the TAAS. Figure 6 illustrates the more typical pattern by showing the negative, linear relationship between Stanford 9 math test scores and the percentage of students in the free or reduced-price lunch program.

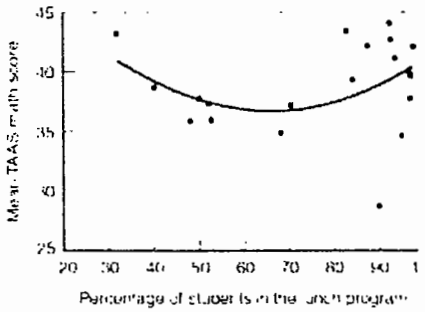


Figure 5
Relationship Between Mean TAAS Math Score and Percentage of Students in the Lunch Progra

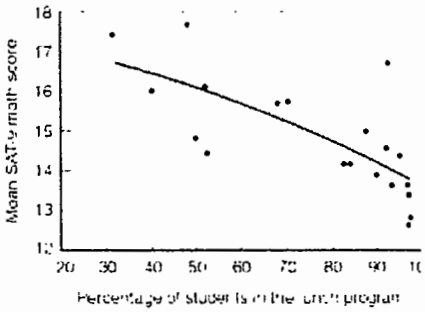


Figure 6
Relationship Between Mean SAT-9 Math Score and Percentage of Students in the Lunch Progra

The fourth set of rows in Table 4 shows that when the student is the unit of analysis, TAAS math and reading scores correlate well with the scores on the tests we gave. Although the correlations are somewhat lower than would be expected from experience with other tests (especially the .46 correlation between the two math tests), these differences do not affect the conclusions we would make about the relationships among different tests. However, the correlation between TAAS and non-TAAS tests essentially disappears when the school is the unit of analysis. This result is contrary to the one that would be expected by other studies and the results in the first block of rows.

The last row of Table 4 shows that TAAS math has a very high correlation with TAAS reading (despite being a different subject). In fact, TAAS math correlates much higher with TAAS reading than it does with another math test (namely: Stanford 9 OE math).

To sum up, the non-TAAS tests correlated highly with each other and with SES; and, as expected, this correlation increased when the school was used as the unit of analysis. Also as anticipated, the two TAAS tests had a moderate correlation with the non-TAAS tests, but unexpectedly, this only occurred when the student was used as the unit of analysis. Rather than getting larger, the correlation between TAAS

and non-TAAS tests essentially evaporated when the school was the unit. And finally, regardless of the unit of analysis, the two TAAS tests had an extremely high correlation with each other, but both had a virtually zero correlation with SES.

One of the reasons we were surprised that the TAAS and non-TAAS scores behaved so differently is that the latter tests were designed to measure some of the same kinds of higher-order thinking skills that the TAAS is intended to measure. However, our results could be due to the unique characteristics of the 20 schools in our study or other factors. We are therefore reluctant to draw conclusions from our findings with these schools or to imply that these findings are likely to occur elsewhere in Texas. Nevertheless, they do suggest the desirability of periodic administration of external tests to validate TAAS results. This procedure, which is sometimes referred to as "audit testing," could have been incorporated into the study of the Metropolitan Achievement Test discussed previously.

Conclusions

We are now ready to answer the questions that we posed at the beginning of this issue paper. Specifically, we found that the reading and math skills of Texas students improved since the full implementation of the TAAS program in 1994. However, the answers to the questions of how much improvement occurred, whether the improvement in reading was comparable to what it was in math, and whether Texas reduced the gap in scores among racial and ethnic groups depend on whether you believe the NAEP or TAAS results. They tell very different stories.

NAEP and TAAS results tell us very different stories.

According to NAEP, Texas fourth graders were slightly more proficient in reading in 1998 than they were in 1994. However, the country as a whole also improved to about the same degree. Thus, there was nothing remarkable about reading score gains in Texas. In contrast, the increase in fourth grade math scores in Texas was significantly greater than it was nationwide. However, the small improvements in NAEP eighth grade math scores were consistent with those observed nationally. The gains in scores between fourth and eighth grade in Texas also were consistent with national trends. In short, except for fourth grade math, the gains in Texas were comparable to those experienced nationwide during this time period.

In all the analyses, including fourth grade math, the gains on the TAAS were several times greater than they were on NAEP. Hence, how much a Texas student's proficiency in reading and math actually improved depends almost entirely on whether the assessment of that student's skills relies on NAEP scores (which are based on national content frameworks) or TAAS scores (which are based on tests that are aligned with Texas' own content standards and are administered by the classroom teacher).

The huge disparities between the stories told by NAEP and TAAS are especially striking in the assessment of (1) the size of the gap in average scores between whites and students of color and (2) whether these gaps are getting larger or smaller. According to NAEP, the gap is large and increasing slightly. According to TAAS, the gap is much smaller and decreasing greatly. We again quote Linn (2000, p. 14): "Divergence of trends does not prove that NAEP is right and the state assessment is misleading, but it does raise important questions about the generalizability of gains reported on a state's own assessment, and hence about the validity of claims regarding student achievement." Put simply, how different could "reading" and "math" be in Texas than they are in the rest of the country?

The data available for this report were not ideal. Limitations in the way NAEP is administered make it difficult to do the kinds of comparisons that would be most informative. For example, NAEP is not given every year and individual student or school scores are not available. And the supplemental study described above was limited to 20 schools in just one part of a very large state. Nevertheless, the stark differences between TAAS and NAEP (and other non-TAAS tests) raise very serious questions about the generalizability of the TAAS scores.

These concerns about TAAS do not condemn all efforts to increase accountability, nor should they be interpreted as being opposed to testing. On the contrary, we believe that some form of large-scale assessment, when properly implemented, is an essential tool to monitor student progress and thereby support state efforts to improve education. Moreover, the possible problems with the TAAS discussed earlier in this issue paper are probably not restricted to this test or state. For example, score inflation and unwanted test preparation have been found in a number of jurisdictions (Koretz & Barron, 1998; Linn, 2000; Stecher et al., 1998; Heubert & Hauser, 1999).

To sum up, states that use high-stakes exams may encounter a plethora of problems that would undermine the interpretation of the scores obtained. Some of these problems include the following: (1) students being coached to develop skills that are unique to the specific types of questions that are asked on the statewide exam (i.e., as distinct from what is generally meant by reading, math, or the other subjects tested); (2) narrowing the curriculum to improve scores on the state exam at the expense of other important skills and subjects that are not tested; (3) an increase in the prevalence of activities that substantially reduce the validity of the scores; and (4) results being biased by various features of the testing program (e.g., if a significant percentage of students top out or bottom out on the test, it may produce results that suggest that the gap among racial and ethnic groups is closing when no such change is occurring).

There are a number of strategies that states might try to lessen the risk of inflated and misleading gains in scores. They can reduce the pressure to "raise scores at any cost" by using one set of measures to make decisions about individual students and another set (employing sampling and third-party administration) to make decisions about

teachers, schools, and educational programs. States can replace their traditional paper-and-pencil multiple-choice exams with computer based "adaptive" tests that are tailored to each student's abilities, that draw on "banks" of thousands of questions, and that are delivered over the Internet into the school building (for details, see Bennett, 1998; Hamilton, Klein, & Lorie, 2000). States can also periodically conduct audit testing to validate score gains. They can study the positive and negative effects of the testing program on curriculum and instruction, and whether these effects are similar for different groups of students. For instance, what knowledge, skills, and abilities are and are not being developed when the focus is concentrated on preparing students to do well on a particular statewide, high-stakes exam? However, given the findings reported above for Texas, it is evident that something needs to be done to ensure that high-stakes testing programs, such as the TAAS, produce results that merit public confidence and thereby provide a sound basis for educational policy decisions.

Notes

RAND issue papers explore topics of interest to the policymaking community. Although issue papers are formally reviewed, authors have substantial latitude to express provocative views without doing full justice to other perspectives. The views and conclusions expressed in issue papers are those of the authors and do not necessarily represent those of RAND or its research sponsors.

1. It was beyond the scope of this issue paper to identify the specific similarities and differences in content coverage between NAEP and TAAS.
2. This estimate includes students who spent one to four years in small classes.
3. In Figures 2 and 3, the NAEP and TAAS trends cover different but overlapping years, due to the testing schedules of these measures.
4. The results in the 20-school study discussed later in this issue paper suggest that some topping-out occurred on the TAAS. For example, although about two-thirds of the 2,000 students in this study were in a free or reduced-price lunch program, 7 percent answered 95 percent of the TAAS reading questions correctly and 9 percent did so on the math test. Only a few students were able to do this on any of the tests we gave.
5. The correlation coefficient, which can range from -1.00 to +1.00, is a measure of the degree of agreement between two tests. A high positive correlation is obtained when the students (or schools) that have high scores on one test also tend to have high scores on the other test.
6. We also examined the relationships by splitting the schools into two groups, according to whether they had relatively high versus low percentages of students in the lunch program (e.g., those that had more than 70 percent versus those with less than 70 percent). This analysis produced results that were consistent with the data in

Figures 5 and 6. Specifically, schools with a high percentage of students in the lunch program had much lower scores on the three tests we gave than did schools with a relatively low percentage of students in this program whereas that was not the case with the TAAS scores.

Acknowledgments

The preparation of this issue paper benefited greatly from the many thoughtful suggestions and insights of our RAND colleagues, Dr. David Grissmer, Dr. Daniel Koretz, and Dr. James Thomson, and our external reviewers, Professor Richard Jaeger of the University of North Carolina at Greensboro and Professor Robert Linn of the University of Colorado at Boulder. We are also grateful to Miriam Polon and Christina Pitcher for editorial suggestions.

References

Barton, P. E., & Coley, R. J. (1998). *Growth in school: Achievement gains from the fourth to the eighth grade* (ETS Policy Information Report). Princeton, NJ: Educational Testing Service.

Bennett, R. E. (1998). *Reinventing assessment.* Princeton, NJ: Educational Testing Service.

Carnoy, M., Loeb, S., & Smith, T. L. (2000). *Do higher state test scores in Texas make for better high school outcomes?* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis,* 21, 97-109.

Grissmer, D., & Flanagan, A. (1998). *Exploring rapid achievement gains in North Carolina and Texas.* Washington, DC: National Education Goals Panel.

Grissmer, D., Flanagan, A., Kawata, J., & Williamson, S. (2000). *Improving student achievement: What state NAEP test scores tell us.* Santa Monica, CA: RAND, MR-924-EDU.

Hamilton, L. S., Klein, S. P., & Lorie, W. (2000). *Using web-based testing for large-scale assessment.* Santa Monica, CA: RAND, IP-196-EDU.

Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives,* 8 (41). Available at http://epaa.asu.edu/epaa/v8n41.

Hedges, L. V., & Nowell, A. (1998). Black-white test score convergence since 1965. In Jencks, C., & Phillips, M. (Eds.), *The Black-White Test Score Gap* (pp. 149-181). Washington, DC: Brookings.

Heubert, J. P., & Hauser, R. M. (Eds.) (1999). *High stakes: Testing for tracking, promotion, and graduation.* A Report of the National Research Council, Washington, DC: National Academy Press.

Hoff, D. J. (2000). As stakes rise, definition of cheating blurs. *Education Week,* June 21.

Hoffman, J. V., Assaf, L., Pennington, J., & Paris, S. G. (in press). High stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher.*

Johnston, R. C. (1999). Texas presses districts in alleged test-tampering cases. *Education Week,* March 17.

Koretz, D., & Barron, S. I. (1998). *The validity of gains on the Kentucky Instructional Results Information System (KIRIS).* Santa Monica, CA: RAND, MR-1014-EDU.

Koretz, D., Linn, R. L., Dunbar, S. B., & Shepard, L. A. (1991). The effects of high-stakes testing: Preliminary evidence about generalization across tests, in R. L. Linn (chair), *The Effects of High Stakes Testing,* symposium presented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago, April.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher,* 29 (2), 4-16.

Linn, R. L., Graue, M. E., & Sanders, N. M. (1990). Comparing state and district test results to national norms: The validity of claims that "everyone is above average." *Educational Measurement: Issues and Practice,* 9, 5-14.

Mabin, Connie (2000). State's students again improve on TAAS scores. *Austin American-Statesman,* May 18.

McLaughlin, D. (2000). *Protecting state NAEP trends from changes in SD/LEP inclusion rates.* Palo Alto, CA: American Institutes for Research.

McNeil, L., & Valenzuela, A. (2000). *The harmful impact of the TAAS system of testing in Texas: Beneath the accountability rhetoric.* Cambridge, MA: Harvard University Civil Rights Project.

Reese, C. M., Miller, K. E., Mazzeo, J., & Dossey, J. A. (1997). *NAEP*

1996 report card for the nation and the states. Washington, DC: National Center for Education Statistics.

Stecher, B. M., Barron, S., Kaganoff, T., & Goodwin, J. (1998). *The effects of standards-based assessment on classroom practices: Results of the 1996-97 RAND Survey of Kentucky Teachers of Mathematics and Writing* (CSE Technical Report 482). Los Angeles: Center for Research on Evaluation, Standards, and Student Testing.

Stecher, B. M., & Klein, S. P. (Eds.) (1996). *Performance assessments in science: Hands-on tasks and scoring guides.* Santa Monica, CA: RAND, MR-660-NSF.

Texas Education Agency (1999). *Texas Student Assessment Program: Technical digest for the academic year 1998-1999.* Available at http://www.tea.state.tx.us/student.assessment/techdig.htm.

Texas Education Agency (2000). *1999 Texas National Comparative Data Study.* Available at http://www.tea.state.tx.us/student.assessment/researchers.htm.

Texas Education Agency (2000). *Texas TAAS passing rates hit seven-year high; four out of every five students pass exam.* Press release, May 17.

About the Authors

Dr. Stephen P. Klein is a Senior Research Scientist at RAND, where for the past 25 years he has led studies on health, criminal justice, military manpower, and educational issues. His current projects include analyzing licensing examinations in teaching and other professions, delivering computer adaptive performance tests over the Web, and measuring the effects of instructional practices and curriculum on student performance.

Dr. Laura Hamilton is an Associate Behavioral Scientist at RAND where she conducts research on educational assessment and the effectiveness of educational reform programs. Her current projects include a study of systemic reforms in math and science, an investigation of the validity of statewide assessments for students with disabilities, and an analysis of the effectiveness of private governance of public schools.

Dr. Daniel F. McCaffrey is a Statistician at RAND where he works on studies of health and educational issues. His work on education includes studies on teaching practices and student achievement, the effects of class size reduction on the test scores of California's students, and the properties of hands-on performance measure of achievement in science.

Dr. Brian Stecher is a Senior Social Scientist in the Education program at RAND. Dr. Stecher's research emphasis is applied educational measurement, including the implementation, quality, and impact of state assessment and accountability systems and the cost, quality, and feasibility of performance-based assessments in mathematics and science.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
New York University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers

Jay D. Scribner

Ann Leavenworth Center
for Accelerated Learning

University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois—UC

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 2134 times since November 2, 2000

Education Policy Analysis Archives

Volume 8 Number 50 **November 2, 2000** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES.**
Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Student Evaluation of Teaching: A Methodological Critique of Conventional Practices

Robert Sproule
Bishop's University (Canada)

Abstract
The purpose of the present work is twofold. The first is to outline two arguments that challenge those who would advocate a continuation of the exclusive use of raw SET data in the determination of "teaching effectiveness" in the "summative" function. The second purpose is to answer this question: "In the face of such challenges, why do university administrators continue to use these data exclusively in the determination of 'teaching effectiveness'?"

I. Introduction

The original purpose of collecting data on the student evaluation of teaching (hereafter SET) was to provide student feedback to an instructor on her "teaching effectiveness" [(Adams (1997), Blunt (1991), and Rifkin (1995)]. This function is dubbed the "formative" function by some, and is viewed as non-controversial by most. In time, raw SET data have been put to another use—this is to provide student input into faculty committees charged with the responsibility deciding on the reappointment, pay, merit pay, tenure, and promotion of an individual instructor [Rifkin (1995), and Grant (1998)]. This second function, dubbed the "summative" function by some, is viewed as controversial by many. (Notes 1, 2)

The purpose of the present work is twofold. The first is to outline two arguments that challenge those who would advocate a continuation of the exclusive use of raw SET data in the determination of "teaching effectiveness" in the "summative" function. The first argument identifies two conceptual, and the second identifies two statistical, fallacies inherent in their methodology. Along the way, I shall also argue that while both conceptual fallacies cannot be remedied, one of the statistical fallacies can—this by means of the collection of additional data and the use of an appropriate statistical technique of the sort outlined in Mason et al. (1995). The second purpose of the present paper is to answer this question: In the face of such challenges, why do university administrators continue to use these data exclusively in the determination of "teaching effectiveness"?

The general motivation for the present work is located in three classes of statements. The first class is the many reports of the confusion and general disarray caused to the academic mission of many disciplines by the SET process. For example, Mary Beth Ruskai (1996), an associate editor of *Notices of The American Mathematical Society*, wrote:

> Administrators, faced with a glut of data, often find creative ways to reduce it (the SET process) to meaningless numbers. I encountered one who insisted that it sufficed to consider only the question on overall effectiveness, because he had once seen a report that, on average, the average on this question equaled the average of all other questions. He persisted in this policy even in cases for which it was patently false ... Advocates often cite a few superficial studies in support of the reliability of student evaluations. However, other studies give a more complex picture ... Many experienced faculty question the reliability of student evaluations as a measure of teaching effectiveness and worry that they may have counter- productive effects, such as contributing to grade inflation, discouraging innovation, and deterring instructors from challenging students.

The second concerns what constitutes admissible and inadmissible evidence in legal and quasi-legal proceedings related to the "summative" function. For example, over fifteen years ago, Gillmore (1984) wrote: "If student ratings are to qualify as evidence in support of faculty employment decisions, questions concerning their reliability and validity must be addressed" (p. 561). In recent times, it seems that the issue of admissibility has been clarified in the U.S. courts. For example, Adams (1997) wrote:

> Concerning questions about the legal basis of student evaluations of faculty, Lechtreck (1990) points out that, "In the past few decades, courts have struck down numerous tests used for hiring, and/or promotions on the grounds that the tests were discriminatory or allowed the evaluator to discriminate. The question, How would you rate the teaching ability of this instructor, is wide open to abuse" (p. 298). In his column, "Courtside," Zirkel (1996) states, "Courts will not uphold evaluations that are based on subjective criteria or data" (p. 579). Administrative assumptions to the contrary, student evaluations of faculty are not objective, but rather, by their very nature, must be considered subjective. (p. 2) (Note 3)

That said, the present work should be seen as an attempt to further reinforce two views: that SET data are not methodologically sound, and that they ought not be treated as admissible evidence in any legal or quasi-legal hearing related to the "summative" function.

And the third motivation stems from *the notion of academic honesty, or from the virtue of acknowledging ignorance when the situation permits no more or no less* – a notion and a virtue the academic community claims as its own. This motivation is captured succinctly by Thomas Malthus (1836) in a statement made over a century and half ago. He wrote:

> To know what can be done, and how to do it, is beyond a doubt, the most important species of information. The next to it is, to know what cannot be done, and why we cannot do it. The first enables us to attain a positive good, to increase our powers, and augment our happiness: the second saves us from the evil of fruitless attempts, and the loss and misery occasioned by perpetual failure. (p. 14)

This article is organized as follows. In the second section, I offer a characterization of the conventional process used in the collection, and processing, of the SET data. This is done for the benefit of those unacquainted with the same. This is then followed by an outline of fallacies inherent in the conventional SET process of the conceptual sort. Similarly, in the fourth section, I outline fallacies inherent in the same of the statistical sort. The next to last section addresses this question: In the face of such challenges, why do university

administrators continue to use these data exclusively in the determination of "teaching effectiveness"? Final remarks are offered in a concluding section.

II. The Conventional SET Process

The conventional process by which the SET data (on a particular instructor of a particular class) are collected and analyzed may be characterized as follows (Note 4)

1. The SET survey instrument is comprised of a series of questions about course content and teaching effectiveness. Some questions are open-ended, while others are closed-ended.
2. Those, which are closed-ended, often employ a scale to record a response. The range of possible values, or example, may run from a low of 1 for "poor," to a high of 5 for "outstanding."
3. In the closed-ended section of the SET survey instrument, one question is of central import to the "summative" function. It asks the student: "Overall, how would you rate this instructor as a teacher in this course?" In the main, this question plays a pivotal role on the evaluation process. For ease of reference, I term this question the "single-most-important question" (hereafter, the SMIQ).
4. In the open-ended section of the SET survey instrument, students are invited of offer short critiques of the course content and of the teaching effectiveness of the instructor.
5. The completion of the SET survey instrument comes with a guarantee to students; that is, the anonymity of individual respondents.
6. The SET survey instrument is administered: (i) by a representative of the university administration to those students of a given class who are present on the data- collection day, (ii) in the latter part of the semester, and (iii) in the absence of the instructor.
7. Upon completion of the survey, the analyst then takes the response to each question on each student's questionnaire, and then constructs question-specific and class-specific measures of central tendency, and of dispersion – this in an attempt to determine if the performance of a given instructor in a particular class meets a cardinally- or ordinally- measured minimal level of "teaching effectiveness." (Note 5)
8. It seems that, in such analyses, raw SET data on the SMIQ are used in the main. More likely than not, this situation arises from the fact that the SET survey instrument does not provide for the collection of background data on the student respondent (such as major, GPA, program year, required course?, age, gender, ...), and on course characteristics. (Note 6)

An example of the two-last features may prove useful. Suppose

there are three professors, A, B, and C, who teach classes, X, Y, and Z, respectively. And suppose that the raw mean of the SMIQ for A in X is 4.5, the raw mean value of the SMIQ for B in Y is 3.0; and the raw mean value of the SMIQ for C in Z is 2.5. Suppose too that the reference-group raw mean score for the SMIQ is 3.5 where the reference group could be either: (i) all faculty in a given department, or (ii) all faculty in the entire university. In the evaluation process, C's mean score for the SMIQ may be compared of with that of another [say A's], and will be compared with that of her reference group. The object of this comparison is the determination of the teaching effectiveness, or ineffectiveness, of C. The questions addressed below are: (a) are the data captured by the SMIQ a valid proxy of "teaching effectiveness," and (b) can the raw mean values of the SMIQ be used in such comparisons?

III. Fallacies Of A Conceptual Sort Inherent In The SET Process

In this section, I outline two fallacies of a conceptual sort inherent in the SET process. These are: (a) that students are a, or alternatively are the only, source of reliable information on teaching effectiveness, and (b) there exists a unique and immutable metric termed "teaching effectiveness."

III.1. Students As A, Or The Only, Source Of Reliable Information on Teaching Effectiveness

Let us return to the example of the three professors, A, B, and C, who teach classes, X, Y, and Z, respectively. There are two questions to be addressed here: (a) Would one be justified in believing that students provide reliable information on teaching effectiveness? (b) If yes, would one be justified in believing that students provide the only source of reliable information on teaching effectiveness? In my view, one would not be justified in holding either belief. There are four reasons:

- The Public-Good Argument: The advocates of the SET process would argue: The university is a business, and the student its customer. And since the customer is always right, customer opinion must drive the business plan. Mainstream economists would argue that this is a false analogy. Their reason is that these same advocates are assuming that the provision of tertiary education is a "private good." This (economists would argue) is not so: It is a "public good."(Note 7) As such, students are not solely qualified to evaluate course content, and the pedagogical style of a faculty member.

- The Student-Instructor Relationship Is Not One of Customer-Purveyor, And Hence Not A Relationship Between Equals: As Stone (1995) noted,

> Higher education makes a very great mistake if it permits its primary mission to become one of serving student "customers." Treating students as customers means shaping services to their taste. It also implies that students are entitled to use or waste the services as they see fit. Thus judging by enrollment patterns, students find trivial courses of study, inflated grades, and mediocre standards quite acceptable. If this were not the case, surely there would have long ago been a tidal wave of student protest. Of course, reality is that student protest about such matters is utterly unknown. Tomorrow, when they are alumni and taxpayers, today's students will be vitally interested in academic standards and efficient use of educational opportunities. Today, however, the top priority of most students is to get through college with the highest grades and least amount of time, effort, and inconvenience.

As Michael Platt (1993) noted:

> The questions typical of student evaluations teach the student to value mediocrity in teaching and even perhaps to resent good teachers who, to keep to high purposes, will use unusual words, give difficult questions, and digress from the syllabus, or seem to. Above all, such questions also conceive the relation of student and teacher as a contract between equals instead of a covenant between unequals. Thus, they incline the student, when he learns little, to blame the teacher rather than himself. No one can learn for another person; all learning is one's own .... (p. 31)

While the student-instructor relationship is not one of customer-purveyor, and hence not a relationship between equals, the SET process itself offers the illusion that it is. As Platt (1993) noted:

> Merely by allowing the forms, the teacher loses half or more of the authority to teach. (p. 32)

- Students Are Not Sufficiently Well-Informed To Pronounce On The Success Or Failure of the Academic Mission: Because of age and therefore relative ignorance, students are not sufficiently well-informed about societal needs for educated persons, and employers' needs for skill sets. Therefore, students are not in a position to speak for all vested interests (including their own long- term interests). For example, Michael Platt (1993) noted:

  > Pascal says: while a lame man knows he limps, a lame mind does not know it limps, indeed says it is we who limp. Yet these forms invite the limpers to judge the runners;

> non-readers, the readers; the inarticulate, the articulate; and non-writers, the writers. Naturally, this does not encourage the former to become the latter. In truth, the very asking of such questions teaches students things that do not make them better students. It suggests that mediocre questions are the important questions, that the student already knows what teaching and learning are, and that any student is qualified to judge them. This is flattery. Sincere or insincere, it is not true, and will not improve the student, who needs to know exactly where he or she stands in order to take a single step forward. (p. 32)

In the same vein, Adams (1997) noted,

> Teaching, as with art, remains largely a matter of individual judgment. Concerning teaching quality, whose judgment counts? In the case of student judgments, the critical question, of course, is whether students are equipped to judge teaching quality. Are students in their first or second semester of college competent to grade their instructors, especially when college teaching is so different from high school? Are students who are doing poorly in their courses able to objectively judge their instructors? And are students, who are almost universally considered as lacking in critical thinking skills, often by the administrators who rely on student evaluations of faculty, able to critically evaluate their instructors? There is substantial evidence that they are not. (p. 31)

- The Anonymity of The Respondent: As noted above, the SET process provides that the identity of the respondent to the SET questionnaire would or could never be disclosed publicly. This fact contains a latent message to students. This is, in the SET process,

> there are not personal consequences for a negligent, false, or even malicious representation. There is no "student responsibility" in student evaluations. It is as if the student was being assured: "We trust you. We do not ask for evidence, or reasons, or authority. We do not ask about your experience or your character. We do not ask your name. We just trust you. Your opinions are your opinions. You are who you are. In you we trust." Most human beings trust very few other human beings that much. The wise do not trust themselves that much. [Platt (1993, p. 34)]

III.2. Opinion Misrepresented As Fact Or Knowledge

A major conceptual problem with the SET process is that opinion is misrepresented as fact or knowledge, not to mention the

unintended harm that this causes to all parties. As Michael Platt (1993) noted:

> I cannot think that the habit of evaluating one's teacher can encourage a young person to long for the truth, to aspire to achievement, to emulate heroes, to become just, or to do good. To have one's opinions trusted utterly, to deliver them anonymously, to have no check on their truth, and no responsibility for their effect on the lives of others are not good for a young person's moral character. To have one's opinions taken as knowledge, accepted without question, inquiry, or conversation is not an experience that encourages self-knowledge. (pp. 33-34)

He continued:

> What they teach is that "Opinion is knowledge." Fortunately, the student may be taught elsewhere in college that opinion is not knowledge. The student of chemistry will be taught that the periodic table is a simple, intelligible account of largely invisible elements that wonderfully explains an enormous variety of visible but heterogeneous features of nature. (p. 32)

This misrepresentation of opinion as fact or knowledge raises problems in statistical analysis of the SET data in that any operational measure of "teaching effectiveness" will not be, by definition, a unique and immutable metric. [This is one of the concerns raised in the next section.] In fact, I claim that the metric itself does not exist, or the presumption that it does is pure and unsubstantiated fiction. The assessment of these claims is the next concern.

To initiate discussion, return to the example of the three professors, A, B, and C, who teach classes, X, Y, and Z, respectively. From data extracted from the SMIQ, recall that A in X scored 4.5, B in Y scored 3.0; and C in Z scored 2.5. Two premises of the conventional SET process are: (i) there exists a unique and an immutable metric, "teaching effectiveness," and (ii) the operational measure of this metric can be gleaned from data captured by the SMIQ, or by a latent-variable analysis (most commonly, factor analysis) of a number of related questions. The question to be addressed here is: Would one be justified in believing that these two premises are true?

In my view, neither premise is credible. The first premise is not true because to assume otherwise is to contradict both the research literature, and casual inspection. There are three inter-related aspects to this claim:

> 1. The first premise contains the uninspected supposition that through introspection, any student can "know" an unobservable metric called "teaching effectiveness," and can then be relied upon to accurately report her measurement of

it in the SET document. (Note 8)

2. The literature makes quite clear that within any group of students one can find multiple perceptions of what constitutes "teaching effectiveness" (e.g., Fox (1983)). (Note 9)

3. If a measure is unobservable, its metric cannot be claimed to be also unambiguously unique and immutable. (Note 10) To argue otherwise is to be confronted by a bind: A measure cannot be subjective, and its metric objective.

That said, what could account for the subjective nature of the term, "teaching effectiveness"? One explanation arises from the existence of two distinct motivations for attending university, or alternatively for enrolling in a given program. The details are these:

1. One motivation is the "education-as-an-investment-good" view. This is tantamount to the view that "going to university" will enhance one's prospects of obtaining a high-paying and/or an intellectually-satisfying job upon graduation. Latent in this view is the fact or belief that many employers take education as a signal of the productive capability of a university graduate as a job applicant [Spence (1974), and Molho (1997, part 2)]. (Note 11)

2. The other motivation is the "education-as-a-consumption-good" view. This view is tantamount to some mix of these five views: (a) that education is to be pursued for education's sake, (b) that "going to university" must be above all else enjoyable, (c) higher education is a democracy, and (d) in this democracy, learning must be fun, and (e) to be educated, students must like their professor. (Notes 12, 13)

Thus, any student can be seen holding some linear combination of these two views. What differentiates one student from the next (at any point in time) is the weighting of this combination.

Next, consider the second premise. It states that the operational measure of the metric, "teaching effectiveness," can be gleaned from data captured by the SET data in general, and by the SMIQ in particular. In my view, one is not justified in assuming the second premise is true because the metric, "teaching effectiveness," is unobservable and subjective. (Note 14) As such, the data captured by the conventional SET process in general, and the SMIQ in particular, can at best measure "instructor popularity" or "student satisfaction" [Damron (1995)]. An example of this subjectiveness can be found in the following passage from Cornell University's (1997) *Science News*,

> Attention teachers far and wide: It may not be so much what or how you teach that will reap high student evaluations, but something as simple as an enthusiastic tone of voice and beware, administrators, if you use student ratings to judge teachers: Although student evaluations may be systematic and reliable, a Cornell university study has found that they can be totally invalid. Yet many schools use them to determine tenure, promotion, pay hikes and awards.
> These warnings stem from a new study in which a Cornell professor taught the identical course twice with one exception—he used a more enthusiastic tone of voice the second semester—and student ratings soared on every measure that second semester.
> Those second-semester students gave much higher ratings not only on how knowledgeable and tolerant the professor was and on how much they say they learned, but even on factors such as the fairness of grading policies, text quality, professor organization, course goals and professor accessibility.
> And although the 249 students in the second-semester course said they learned more than the 229 students the previous semester believed they had learned, the two groups performed no differently on exams and other assessment measures.
> "This study suggests that factors totally unrelated to actual teaching effectiveness, such as the variation in a professor's voice, can exert a sizable influence on student ratings of that same professor's knowledge, organization, grading fairness, etc.," said Wendy M. Williams, associate professor of human development at Cornell. Her colleague and co-author, Stephen J. Ceci, professor of human development at Cornell, was the teacher evaluated by the students in a course on developmental psychology that he has taught for almost 20 years.

The assertion that the data captured by the conventional SET process in general, and the SMIQ in particular, measure at best "instructor popularity" or "student satisfaction" is echoed by Altschuler (1999). He wrote:

> At times, evaluations appear to be the academic analogue to "Rate the Record" on Dick Clark's old "American Bandstand," in which teen-agers said of every new release, "Good beat, great to dance to, I'd give it a 9." Students are becoming more adjectival than analytical, more inclined to take faculty members' wardrobes and hairstyles into account when sizing them up as educators.

IV. Fallacies Of A Statistical Sort Inherent In The SET Process

In this section, I outline potential fallacies of a statistical sort inherent in the SET process. There are two: (a) under all circumstances, the SMIQ provides a cardinal measure of "teaching effectiveness" of an instructor, and (b) in the absence of statistical controls, the SMIQ provides an ordinal measure of "teaching ef. ctiveness" of an instructor. (Notes 15,16)

IV.1. Ascribing A Cardinal Measure of Teaching Effectiveness To An Instructor Based on The SMIQ

Return to the example of the three professors, A, B, and C, who teach classes, X, Y, and Z, respectively. Recall that A in X scored 4.5, B in Y scored 3.0; C in Z scored 2.5, and the reference group scored 3.5. A premise of the SET process is that these averages are *cardinal measures* of "teaching effectiveness." The question to be addressed here is: Would one be justified in believing that this premise is true? That is, would one be justified in believing that A is 50% "more effective" than B, that B is 20% "more effective" than C, or that A is 28% "more effective" than the average? (Note 17)

In my view, one would not be justified in believing any such claim simply because of the argument outlined in the previous section; that is, a unique and an immutable metric, "teaching effectiveness," does not exist.

IV.2. The Rank Ordering Of Instructors By Teaching Effectiveness Based on The SMIQ

Return again to the example of three professors, A, B, and C, who teach classes, X, Y, and Z, respectively. An alternative premise of the conventional SET process is that the averages of the data captured by the SMIQ serve as a basis for an *ordinal measure* of "teaching effectiveness." The question to be addressed here is: Would one be justified in believing that this premise is true? That is, would one be justified in believing that A is "more effective" than B, or that B is "more effective" than C? In my view, this belief could be seen as justifiable: (a) if the SMIQ captures an unequivocal reading of "teaching effectiveness" (see above), and (b) if the subsequent analysis controls for the many variables which confound the data captured by the SMIQ.(Note 18)

What are these confounding variables that require control? To answer this question, two studies are worthy of mention. One, in a review of the literature, Cashin (1990) reports that (in the aggregate) students do not provide SET ratings of teaching performance uniformly across academic disciplines. (Note 19)

Two, in their review of the literature, Mason et al. (1995, p. 404) note that there are three clusters of variables, which affect student perceptions of the teaching effectiveness of faculty members. These clusters are: (a) student characteristics, (b) instructor characteristics, and (c) course characteristics. (Note 20) They also note that only one

of these clusters ought to be included in any reading of "teaching effectiveness." This is the cluster, "instructor characteristics." Commenting on prior research, Mason et al. (1995, p. 404) noted:

> A ...virtually universal problem with previous research is that the overall rating is viewed as an effective representation of comparative professor value despite the fact that it typically includes assessments in areas that are beyond the professor's control. The professor is responsible to some extent for course content and characteristics specific to his/her teaching style, but is unable to control for student attitude, reason for being in the course, class size, or any of the rest of those factors categorized as student or course characteristics above. Consequently, faculty members should be evaluated on a comparative basis only in those areas they can affect, or more to the point, only by a methodology that corrects for those influences beyond the faculty member's control.
>
> By comparing raw student evaluations across faculty members, administrators implicitly assume that none of these potentially mitigating factors has any impact on student evaluation differentials, or that such differentials cancel out in all cases. The literature implies that the former postulate is untrue.

The true import of the above is found again in Mason et al. (1995). Using an ordered-probit model, (Note 21) they demonstrate that student characteristics, instructor characteristics, and course characteristics do impact the response to the SMIQ in the SET dataset. They wrote:

> Professor characteristics dominated the determinants of the summary measures of performance, and did so more for those summary variables that were more professor-specific. However, certain course- and student- specific characteristics were very important, skewing the rankings based on the raw results. Students consistently rewarded teachers for using class time wisely, encouraging analytical decision making, knowing when students did not understand, and being well prepared for class. However, those professors who gave at least the impression of lower grades, taught more difficult courses, proceeded at a pace students did not like, or did not stimulate interest in the material, fared worse. (p. 414)

Mason et al. (1995) then wrote:

> Based on the probit analysis, an alternative ranking scheme was developed for faculty that excluded influences beyond the professor's control. These rankings differed to some

> extent from the raw rankings for each of the aggregate questions. As a result, the validity of the raw rankings of faculty members for the purposes of promotion, tenure, and raises should be questioned seriously. ... Administrators should adjust aggregate measures of teaching performance to reflect only those items within the professors' control, so that aggregates are more likely to be properly comparable and should do so by controlling for types of courses, levels of courses, disciplines, meeting times, etc. ... Administrators failing to do this are encouraged to reconsider the appropriateness of aggregate measures from student evaluations in promotion, tenure, and salary decisions, concentrating instead on more personal evaluations such as analysis of pedagogical tools, peer assessments, and administrative visits. (p. 414)

It may be useful to ask: To what extent are the findings of Mason et al. (1995) unique? Surprisingly, they are not; they echo those of other studies, some recent, and some more than a quarter-century old. For example, Miriam Rodin and Burton Rodin (1972) writing in Science present a study in which they correlated an objective measure of "good teaching" (viz., a student's performance on a calculus test) with a subjective measure of "good teaching" (viz., a student's evaluation of her professor) holding constant the student's initial ability in calculus. What they found is that these two measures were not orthogonal or uncorrelated as some might expect, but something more troublesome. These two variables had a correlation coefficient less than –0.70, and these two accounted for more about half of the variance in the data. How did they interpret their findings? The last sentence in their paper states: "If how much students learn is considered to be a major component of good teaching, it must be concluded that good teaching is not validly measured by student evaluations in their current form." How might others interpret their findings? They suggest the individual instructor is in a classic double-bind: If she attempts to maximize her score on the SMIQ, then she lowers student performance. Alternatively if she attempts to maximize student performance, then her score on the SMIQ suffers. This begs the question: In such a dynamic, how can one possibly use SET data to extract a meaningful measure of "teaching effectiveness?"

In a different study (one concerned with the teaching evaluations for the Department of Mathematics at Texas A&M University, and one which entails the analysis of the correlation coefficients for arrays of variables measuring "teaching effectiveness" and "course characteristics"), Rundell (1996) writes: "(T)he analysis we have performed on the data suggests that the distillation of evaluations to a single number without taking into account the many other factors can be seriously misleading" (p. 8).

V. Why Has The Conventional SET Process Not Been Discarded?

Given that the likelihood of deriving meaningful and valid inferences from raw SET data is nil, the question remains: Why is the conventional SET process (with its conceptual and statistical shortcomings) employed even to this day, and by those for who highly revere the power of critical thinking?

To my mind, there are three answers to this question. The first answer concerns political expediency; that is, while fatally flawed, raw SET data can be used as a tautological device; that is, to justify most any personnel decision. As a professor of economics at Indiana University and the Editor of *The Journal of Economic Education* noted:

> End of term student evaluations of teaching may be widely used simply because they are inexpensive to administer, especially when done by a student in class, with paid staff involved only in the processing of the results...Less-than-scrupulous administrators and faculty committees may also use them ... because they can be dismissed or finessed as needed to achieve desired personnel ends while still mollifying students and giving them a sense of involvement in personnel matters. [Becker (2000, p. 114)]

The second is offered by Donald Katzner (1991). He asserted that in their quest to describe, analyze, understand, know, and make decisions, western societies have accepted (for well over five hundred years) the "myth of synonymity between objective science and measurement" (p. 24). (Note 22) He wrote:

> [W]e moderns, it seems, attempt to measure everything.... We evaluate performance by measurement.... What is not measurable we strive to render measurable, and what we cannot, we dismiss it from our thoughts and justify our neglect by assigning it the status of the "less important." ... A moment's reflection, however, is all that is needed to realize that measurement cannot possibly do everything we expect it to do. ... by omitting from our considerations what cannot be measured, or what we do not know how to measure, often leads to irrelevance and even error. (p. 18)

The third reason is offered by Imre Lakatos (1978) in his explanation as to why prevailing scientific paradigms are rarely replaced or overthrown. This contains these elements:

1. What ought to be appraised in the philosophy of the sciences is not an isolated individual theory, but a cluster of interconnected theories, or what he terms "scientific research programs" (hereafter SRP).
2. An SRP protects a "hard core" set of unquestioned and untestable statements. These statements are accepted as "fact."

3. Stated differently, the hard core of a SRP is surrounded by a "protective belt" of "auxiliary hypotheses."
4. One or more of the hard core statements cannot be refuted without dismantling the entire cognitive edifice, which happens in practice only very rarely. That said, it follows that any departure from the hard core of a SRP is tantamount to the creation of a new and different SRP.

Thus, in my view, the conventional SET process is the artifact of an SRP. Judging from the substance of its protective belt, and from the disciplinary affiliations of its proponents or advocates, this is an SRP defined and protected by a cadre of psychologists and educational administrators. (Notes 23,24)

VI. Conclusion

In the present work, I have advanced two arguments, both of which question the appropriateness of using raw SET data (as the only source of data) in the determination of "teaching effectiveness." The first argument identified two types of fallacies in this methodology. One is conceptual, and the other statistical. Along the way, I argued by implication that the conceptual fallacies cannot be remedied, but that one of the statistical fallacies can – this by means of the collection of additional data and the use of an appropriate statistical technique of the sort outlined in the study of Mason et al. (1995), which I also discussed.

The second argument is centered on the question, why do the current practices used in the determination of the "teaching effectiveness" ignore these two fallacies? I offered three answers to this question. These are: (a) that the conventional SET process offers to any university administration a politically-expedient performance measure, and (b) that the conventional SET process may be seen as an example of: (i) Katzner's (1991) "myth of synonymity between objective science and measurement," and (ii) Lakatos' (1978) general explanation of the longevity of SRPs.

Two implications flow from these arguments, and the related discussion. These are as follows: One, the present discussion should not be seen as tantamount to an idle academic debate. On the contrary, since the SET data have been entered as evidence in courts of law and quasi-legal settings [Adams (1997), Gillmore (1984), and Haskell (1997d)], and since the quality and the interpretation of these data can impact the welfare of individuals, it is clear that the present paper has import and bearing to the extent that: (i) it explicates the inadequacies, and unintended implications, of using raw SET data in the "summative" function, and (ii) it explains the present resistance of the conventional SET process to radical reform.

Two, given the present assessment of the conventional SET process, and given the legal repercussions of its continued use, the question becomes: What to do? Here, the news is both good and bad. The bad news is that nothing can be done to obviate the conceptual

fallacies outlined in the above pages. The inescapable truth is that the SMIQ in particular, and the SET dataset in general, do not measure "teaching effectiveness." They measure something akin to the "popularity of the instructor," which (it must be emphasized) is quite distinct from "teaching effectiveness." [Recall the discussion of Rodin and Rodin (1972) in the above.] The good news is that one of the statistical fallacies inherent in the conventional SET process can be overcome – this by capturing and then using background data on student, instructor, and course characteristics, in the mold of Mason et al. (1995). That said, I leave the last word to what (in my opinion) amounts to a classic in its own time. Mason et al. (1995) state, and I repeat:

> Administrators should adjust aggregate measures of teaching performance to reflect only those items within the professors' control, so that aggregates are more likely to be properly comparable and should do so by controlling for types of courses, levels of courses, disciplines, meeting times, etc. ... Administrators failing to do this are encouraged to reconsider the appropriateness of aggregate measures from student evaluations in promotion, tenure, and salary decisions, concentrating instead on more personal evaluations such as analysis of pedagogical tools, peer assessments, and administrative visits. (p. 414)

Notes

This article was prepared during the winter semester of 2000 while the author was on a half-year sabbatical at the University of Manitoba (Winnipeg, Canada). Without implicating them for any remaining errors and oversights, the author thanks Donald Katzner, Paul Mason, Stuart Mckelvie, and three anonymous referees, for many useful comments and critiques.

1. For reviews of the literature that are essentially supportive of the SET process, see d'Apollonia and Abrami (1997), Greenwald and Gilmore (1997), Marsh (1987), Marsh and Roche (1997), and McKeachie (1997). And for reviews of the literature that are highly critical of some mix of the conceptual, statistical, and legal foundations of the SET process, see Damron (1995), and Haskell (1997a, 1997b, 1997c, and 1997d).
2. The terms "formative" and "summative" are due to Scriven (1967).
3. On such matters, the position of the Canadian Association of University Teachers on the admissibility of SET data appears unambiguous in light of statements like these: "Appropriate professional care should be exercised in the development of questionnaires and survey methodologies. Expert advice should be sought, and reviews of the appropriate research and scientific evidence should be carried out. Comments from faculty and

students and their associations or unions should be obtained at all stages in the development of the questionnaire. Appropriate trials or pilot studies should be conducted and acceptable levels of reliability and validity should be demonstrated before a particular instrument is used in making personnel decisions" [Canadian Association of University Teachers (1998, p. 3)]. In a footnote to this passage, this document continues, "Most universities require at least this standard of care before investigators are permitted to conduct research on human subjects. It is unacceptable that university administrations would condone a lesser standard in the treatment of faculty, particularly when the consequences of inadequate procedures and methods can be devastating to teachers' careers."

4. The present characterization represents an amalgam of three sources: (a) first-hand knowledge of the SET documents used at three Canadian universities; (b) a small, non-random sample of SET documents for four universities taken from the internet [viz., University of Minnesota, University of British Columbia, York University (Toronto), and University of Western Ontario]; and (c) non-institutional-specific comments made in the voluminous literature on the SET process.
5. The phrase "a cardinally- or ordinally- measured minimal level of "teaching effectiveness"" requires four comments. One, examples of cardinal measures are: The heights of persons A, B, and C are 6'1", 5'10", and 5'7" respectively. And using the same data, examples of ordinal measures are: A is taller than B, B is taller than C, and A is taller than C. Two, the present measurement terminology is used in economics [Pearce (1992)], and (it can be said) is distinct from that used in other disciplines [e.g., Stevens (1946), Siegel (1956, p. 30), and Hands (1996)]. Three, it is the existence of a unique and an immutable metric (in the above examples, distance or length) that makes both cardinal and ordinal measures meaningful. Four, as the above examples make clear, an ordinal measure can be inferred from a cardinal measure, but not the reverse.
6. An example of this statement is the instrument used by York University (Toronto). An exception to this statement is that used by the University of Minnesota.
7. The distinction between a "private good" and a "public good" can be rephrased in several, roughly equivalent ways. These are: (i) tertiary education has externalities; (ii) that the net social benefits of tertiary education differ from the net private benefits, (iii) that the benefits of tertiary education do not accrue to, nor are its costs borne by, students solely, and (iv) that students do not pay full freight. Because of this, one could argue that (in the evaluation of "teaching effectiveness") the appropriate populations of opinion to be sampled are all groups who share in the social benefits and social costs. These would include not only students, but also members of the Academy, potential employers, and other members of society (such as taxpayers). In sum, because tertiary

education is not a private, but a public good, students are not solely qualified to evaluate course content, and the pedagogical style of a faculty member.

8. A personal vignette provides some insight into the potential seriousness of the inaccuracy of self-reported data. In the fall of 1997, I taught an intermediate microeconomics course. The mark for this course was based solely on two mid-term examinations, and a final examination. Each mid-term examination was marked, and then returned to students and discussed in the class following the examination. Now, the course evaluation form has the question, "Work returned reasonably promptly." The response scale ranges from 0 for "seldom," to 5 for "always." Based on the facts, one would expect (in this situation) an average response of 5. This expectation was dashed in that 50% of the sample gave me a 5, 27.7% gave me a 4, and 22.2% gave me a 3. The import of this? If self-reported measures of objective metrics are inaccurate (as this case indicates), how can one be expected to trust the validity of subjective measurements like "teaching effectiveness?"
9. Indeed, it appears that students and professors can hold different perceptions as to what constitutes "appropriate learning," and hence "appropriate teaching," in tertiary education. For example, Steven Zucker (1996), professor of Mathematics at Johns Hopkins University, laments the gulf between the expectations of students and instructors. He writes: "The fundamental problem is that most of our current high school graduates don't know how to learn or even what it means to learn (a fortiori to understand) something. In effect, they graduate high school feeling that learning must come down to them from their teachers. That may be suitable for the goals of high school, but it is unacceptable at the university level. That the students must also learn on their own, outside the classroom, is the main feature that distinguishes college from high school." (p. 863).
10. Alternatively, Weissberg (1993, p. 8) noted that one cannot measure what one cannot define.
11. These assertions have been borne out empirically under the rubric, "sheepskin effect." The interested reader is directed to Belman and Heywood (1991 and 1997), Heywood (1994), Hungerford and Solon (1987), and Jaeger and Page (1996).
12. Some of these views contradict the *raison d'être* and the *modus operandi* of tertiary education. For example, Frankel (1968) wrote: "Teaching is a professional relationship, not a popularity contest. To invite students to participate in the selection or promotion of their teachers exposes the teacher to intimidation." (pp. 30-31) In fact, the Canadian Association of University Teachers (1986) speaks of the irrelevance of "popularity" as a gauge of professional performance by stating: "The university is not a club; it is dedicated to excellence. The history of universities suggests that its most brilliant members can sometimes be difficult, different from their colleagues, and

unlikely to win a popularity contest. 'The university is a community of scholars and it is to be expected that the scholars will hold firm views and wish to follow their convictions. Tension, personality conflicts and arguments may be inevitable by-products.'"

13. As Crumbley (1995) noted: "There is another universal assumption that students must like an instructor to learn. Not true. Even if they dislike you and you force them to learn by hard work and low grades, you may be a good educator (but not according to SET scores). SET measures whether or not students like you, and not necessarily whether you are teaching them anything. Instructors should be in the business of educating and teaching students--not SET enhancement. Until administrators learn this simple truth, there is little chance of improving higher education."
14. It seems that some psychologists would argue that latent measures of "teaching effectiveness" can be uncovered by a factor analysis of the SET data [e.g., d'Apollonia and Abrami (1997)]. Also, it seems that the motivation for such a claim is the intellectual appeal and success of studies of a completely different ilk. A case in point is Linden (1977) who uses factor analysis to uncover dimensions, which account for event-specific performances of athletes in the Olympic decathlon. However, the expectation that the success found in studies such as Linden (1977) can be replicated in the factor analysis of SET data is unwarranted in that this expectation ignores the fact that the SET data (unlike Linden's data) are opinion based or subjective, have measurement error, and are in need of statistical controls. In brief, it is my view that the use of factor analysis on SET data to uncover latent measures such as "teaching effectiveness" is analogous to trying to "unscramble an egg" in that it just cannot be done. Besides, as the authors of a popular text on multivariate statistics observe, "When all is said and done, factor analysis remains very subjective" [Johnson and Wichern (1988, p. 422)].
15. The terms, ordinal and cardinal measures, are defined in a footnote above. In conjunction with that, it should be noted that the type of a variable governs the statistical manipulations permissible [Hands (1996, pp. 460-62)], and "(T)he use of ordinally calibrated variables as if they were fully quantified .. results in constructions that are without meaning, significance, and explanatory power. Treating ordinal variables as cardinal ... can mislead an investigator into thinking the analysis has shed light on the real world" [Katzner (1991, p. 3)]. This latter point captures an important dimension of the present state of research on SET data, and of the present paper.
16. For reasons of brevity, I have concentrated on only two of several statistical problems. These are "measurement error" and "omitted variables." By doing so, I have overlooked other statistical problems inherent in the SET data like the unreliability of self- and anonymous-reporting, inadequate sample size, sample-selection bias, reverse causation, and teaching to tests.

The reader interested in a more complete treatment of some of these issues may wish to consult readings such as Aiger and Thum (1986), Becker and Power (2000), Gramlich and Greenlee (1993), and Nelson and Lynch (1984).

17. As Rundell (1996) noted, in actual practice, this would mean: "...'Jones had a 3.94 mean on her student evaluations, and since this is 0.2 above the average for the Department, we conclude she is an above average instructor as judged by these questionnaires' is a statement that appears increasingly common" (p. 1).
18. Statistical controls are needed to the extent that they eliminate "observational equivalence." In this connection, two comments are warranted here. One, observational equivalence is said to exist when "alternative interpretations, with different theoretical or policy implications, are equally consistent with the same data.. No analysis of the data would allow one to decide between the explanations, they are observationally equivalent. Other information is needed to identify which is the correct explanation of the data" [Smith (1999, p. 248)]. Two, Sproule (2000) has identified three distinct forms of observational equivalence in the interpretation of raw data from the SMIQ.
19. Cashin (1990) reports, for example, professors of fine arts and music receive high scores on the SMIQ, and professors of chemistry and economics receive lower scores, all things being equal.
20. Mason et al. (1995) contend that those variables which fall under the "student-characteristics" rubric include: (i) reason for taking the course, (ii) class level of the respondent, (iii) student effort in the course, (iv) expected grade in the course, and (v) student gender. Those variables which fall under the "instructor-characteristics" rubric include: (i) the professor's use of class time, (ii) the professor's availability outside of class, (iii) how well the professor evaluates student understanding, (iv) the professor's concern for student performance, (v) the professor's emphasis on analytical skills, (vi) the professor's preparedness for class, and (vii) the professor's tolerance of opposing viewpoints and questions. Those variables which fall under the "course-characteristics" rubric include: (i) course difficulty, (ii) class size, (iii) whether the course is required or not, and (iv) when the course was offered.
21. For an elementary discussion of the ordered-probit model, see Pindyck and Rubinfeld (1991, pp. 273-274.).
22. Katzner (1991) also states that this "blind pursuit of numbers" can lead to unintended, and unjust, outcomes. For example, "(W)hen the state secretly sterilizes individuals only because their 'measured intelligence' on flawed intelligence tests is too low, then bitterly dashed hopes and human suffering becomes the issue." (p. 18). That said, it would not be too difficult to claim that the "blind pursuit of numbers" by those responsible for the "summative" function has also led to unintended, and unjust, outcomes. [In fact, see Haskell (1997d) for details.]

23. Three comments seem warranted here. One, the enterprise of science can been seen as a "market process" [Walstad (1999)]. Two, the SRP of this cadre of psychologists and educational administrators could be viewed as barrier to entry (of the epistemological sort) into the marketplace of ideas. Three, that said, perhaps the recommendation of Paul Feyerabend (1975) applies in this instance; that competition between epistemologies, rather than the monopoly of a dominant epistemology, ought to be encouraged.
24. While it is clear from the above that the protective belt of the SRP associated with the SET has survived many types of logical appraisals (or epistemological attacks), the question remains: Can this protective belt, and this SRP itself, continue to withstand such repeated attacks? I would hazard the opinion that, no, it cannot.

References

Adams, J.V. (1997), Student evaluations: The ratings game, *Inquiry 1 (2)*, 10-16.

Aiger, D., and F. Thum (1986), On student evaluation of teaching ability, *Journal of Economic Education*, Fall, 243-265.

Altschuler, G. (1999), Let me edutain you, *The New York Times*, Education Life Supplement, April 4.

Becker, W. (2000), Teaching economics in the 21st century, *Journal of Economic Perspectives 14 (1)*, 109-120.

Becker, W., and J. Power (2000), Student performance, attrition, and class size, given missing student data, *Economics of Education Review*, forthcoming.

Belman, D., and J.S. Heywood (1991), Sheepskin effects in the returns to education: An examination on women and minorities, *Review of Economics and Statistics 73 (4)*, 720-24.

Belman, D., and J.S. Heywood (1997), Sheepskin effects by cohort: Implications of job matching in a signaling model, *Oxford Economic Papers 49 (4)*, 623-37.

Blunt, A. (1991), The effects of anonymity and manipulated grades on student ratings of instructors, *Community College Review 18*, Summer, 48-53.

Canadian Association of University Teachers (1986), What is fair? A guide for peer review committees: Tenure, renewal, promotion, Information Paper, November.

Canadian Association of University Teachers (1998), Policy on the use of anonymous student questionnaires in the evaluation of teaching, CAUT Information Service Policy Paper 4-43.

Cashin, W. (1990), Students do rate different academic fields differently, in M. Theall and J. Franklin, eds., *Student Ratings of Instruction: Issues for Improving Practice,* New Directions for Teaching and Learning, No. 43 (San Francisco, CA: Jossey-Bass).

Cornell University (1997), Cornell study finds student ratings soar on all measures when professor uses more enthusiasm: Study raises concerns about the validity of student evaluations, *Science News*, September 19th.

Crumbley, D.L. (1995), Dysfunctional effects of summative student evaluations of teaching: Games professors play, *Accounting Perspectives 1 (1)*, Spring, 67-77.

Damron, J.C. (1995). The three faces of teaching evaluation, unpublished manuscript, Douglas College, New Westminster, British Columbia.

d'Apollonia, S., and P. Abrami (1997), Navigating student ratings of instruction, *American Psychologist 52 (11)*, 1198-1208.

Feyerabend, P. (1975), *Against Method* (London: Verso).

Fox, D. (1983), Personal theories of teaching, *Studies in Higher Education 8 (2),* 151-64.

Frankel, C. (1968), *Education and the Barricades* (New York: W.W. Norton).

Gillmore, G. (1984), Student ratings as a factor in faculty employment decisions and periodic review, *Journal of College and University Law 10*, 557- 576.

Gramlich, E., and G. Greenlee (1993), Measuring teaching performance, *Journal of Economic Education*, Winter, 3-13.

Grant, H. (1998), Academic contests: Merit pay in Canadian universities, *Relations Industrielles / Industrial Relations 53 (4)*, 647-664.

Greenwald, A., and G. Gilmore (1997), Grading leniency is a removable contaminant of student ratings, *American Psychologist 52 (11)*, 1209-17.

Hands. D.J. (1996), Statistics and the theory of measurement, *Journal of the Royal Statistical Society – Series A 159 (3)*, 445-473.

Haskell, R.E. (1997a), Academic freedom, tenure, and student evaluations of faculty: Galloping polls in the 21st century, *Education Policy Analysis Archives 5 (6),* February 12.

Haskell, R.E. (1997b), Academic freedom, promotion, reappointment, tenure, and the administrative use of student evaluation of faculty (SEF): (Part II) Views from court, *Education Policy Analysis Archives 5 (6),* August 25.

Haskell, R.E. (1997c), Academic freedom, promotion, reappointment, tenure, and the administrative use of student evaluation of faculty (SEF): (Part III) Analysis and implications of views from the court in relation to accuracy and psychometric validity, *Education Policy Analysis Archives 5 (6),* August 25.

Haskell, R.E. (1997d), Academic freedom, promotion, reappointment, tenure, and the administrative use of student evaluation of faculty (SEF): (Part IV) Analysis and implications of views from the court in relation to academic freedom, standards, and quality of instruction, *Education Policy Analysis Archives 5 (6),* November 25.

Heywood, J.S. (1994), How widespread are sheepskin returns to education in the U.S.?, *Economics of Education Review 13 (3),* 227-34.

Hungerford, T., and G. Solon (1987), Sheepskin effects in the returns to education, *Review of Economics and Statistics 69 (1),* 175-77.

Jaeger, D., and M. Page (1996), Degrees matter: New evidence on sheepskin effects in the returns to education, *Review of Economics and Statistics 78 (4),* 733-40.

Johnson, R., and D. Wichern (1988), *Applied Multivariate Statistical Analysis*, Second Edition (Englewood Cliffs: Prentice-Hall).

Katzner, D. (1991), Our mad rush to measure: How did we get there?, *Methodus 3 (2),* 18-26.

Lakatos, I. (1978), *The Methodology of Scientific Research Programmes* (Cambridge: Cambridge University Press).

Linden, M. (1977), A factor analytic study of Olympic decathlon data, *Research Quarterly 48 (3),* 562- 568.

Malthus, T. (1836), *Principles of Political Economy*, 2nd Edition.

Marsh, H. (1987), Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research, *International Journal of Educational Research 11,* 253-388.

Marsh, H., and L. Roche (1997), Making students' evaluations of teaching effectiveness effective: The central issues of validity, bias, and utility, *American Psychologist 52 (11)*, 1187-97.

Mason, P., J. Steagall, and M. Fabritius (1995), Student evaluations of faculty: A new procedure for using aggregate measures of performance, *Economics of Education Review 12 (4)*, 403-416.

McKeachie, W. (1997), Student ratings: The validity of use, *American Psychologist 52 (11)*, 1218- 1225.

Molho, I. (1997), *The Economics of Information: Lying and Cheating in Markets and Organizations* (Oxford: Blackwell).

Nelson, J., and K. Lynch (1984), Grade inflation, real income, simultaneity, and teaching evaluations, *Journal of Economic Education*, Winter, 21-37.

Pearce, D.W., ed. (1992), *The MIT Dictionary of Modern Ecoonomics*, 4th Edition (Cambridge, MA: MIT Press).

Pindyck, R. and D. Rubinfeld (1991), *Econometric Models & Economic Forecasts* (New York: McGraw-Hill).

Platt, M. (1993), What student evaluations teach, *Perspectives In Political Science 22 (1)*, 29-40.

Rifkin, T. (1995), The status and scope of faculty evaluation, *ERIC Digest.*

Rodin, M., and B. Rodin (1972), Student evaluations of teaching, *Science 177*, September, 1164- 1166.

Rundell, W. (1996), On the use of numerically scored student evaluations of faculty, unpublished working paper, Department of Mathematics, Texas A&M University.

Ruskai, M.B. (1996), Evaluating student evaluations, *Notices of The American Mathematical Society 44 (3)*, March 1997, 308.

Scriven, M. (1967), The methodology of evaluation, in R. Tyler, R. Gagne, and M. Scriven, eds., *Perspectives in Curriculum Evaluation* (Skokie, IL: Rand McNally).

Siegel, S. (1956), *Nonparametric Statistics For The Behavioral Sciences* (New York: McGraw-Hill).

Smith, R. (1999), Unit roots and all that: The impact of time-series methods on macroeconomics, *Journal of Economic Methodology 6 (2).*

239-258.

Spence, M. (1974), *Market Signaling* (Cambridge, MA: Harvard University Press).

Sproule, R. (2000). The underdetermination of instructor performance by data from the student evaluation of teaching,*Economics of Education Review* (in press).

Stevens, S.S. (1946), On the theory of scales of measurement, *Science 103*, 677-680.

Stone, J.E. (1995), Inflated grades, inflated enrollment, and inflated budgets: An analysis and call for review at the state level, *Education Policy Analysis Archives 3 (11).*

Walstad, A. (1999), Science as a market process, unpublished paper, Department of Physics, University of Pittsburgh—Johnstown.

Weissberg, R. (1993), Standardized teaching evaluations, *Perspectives In Political Science 22 (1),* 5-7.

Zucker, S. (1996), Teaching at the university level, *Notices of The American Mathematical Society 43 (8),* August, 863-865.

About the Author

Robert Sproule
Department of Economics
Williams School of Business and Economics
Bishop's University
Lennoxville, Québec, J1M 1Z7, Canada
The author can be reached via e-mail at rsproule@ubishops.ca

Robert Sproule is a Professor of Economics at Bishop's University. He received his Ph.D. in Economics at the University of Manitoba (Winnipeg, Canada). His research interests include statistics, econometrics, and decision making under uncertainty. His research appears in journals such as the *Bulletin of Economic Research, Communications in Statistics: Theory and Methods, Economics Letters,* the *Economics of Education Review*, the *European Journal of Operations Research, Metroeconomica, Public Finance,* and *The Statistician.*

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Greg Camilli
Rutgers University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher
Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

This article has been retrieved 1478 times since November 15, 2000

Education Policy Analysis Archives

Volume 8 Number 51 November 15, 2000 ISSN 1068-2341

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Findings from the Teaching, Learning, and Computing Survey: Is Larry Cuban Right?

Henry Jay Becker
University of California, Irvine

Abstract

Cuban (1986; 2000) has argued that computers are largely incompatible with the requirements of teaching, and that, for the most part, teachers will continue to reject their use as instruments of student work during class. Using data from a nationally representative survey of 4th through 12th grade teachers, this paper demonstrates that although Cuban correctly characterizes frequent use of computers in academic subject classes as a teaching practice of a small and distinct minority, certain conditions make a big difference in the likelihood of a teacher having her students use computers frequently during class time. In particular, academic subject-matter teachers who have at least five computers present in their classroom, who have at least average levels of technical expertise in their use, and who are in the top quartile on a reliable and extensive measure of constructivist teaching philosophy are very likely to have students make regular use of computers during class. More than 3/4 of such teachers have students use word processing programs regularly during class and a majority are regular users of at least one other type of software besides skill-based games. In addition, other factors-such as an orientation towards depth rather than breadth in their teaching(perhaps caused by limited pressures to cover large amounts of content) and block scheduling structures that provide for long class periods-are also associated with greater use of computers by students during class. Finally, the paper provides evidence that certain approaches to using computers result in students taking greater initiative in using computers outside of class time-approaches consistent with a constructivist teaching philosophy, rather than a standards- based, accountability-oriented approach to teaching. Thus, despite their clear minority status as a primary resource in academic subject classroom teaching, computers are playing a major role in at least one major direction of current instructional reform efforts.

Introduction

For about 15 years, Larry Cuban has argued that computers, as a medium of instruction and as a tool for student learning, are largely incompatible with the requirements of teaching. Cuban points out that teachers have so many students to teach (or, in the elementary grades, so many different subjects to cover) that, along with the increasing accountability demanded of them, it is simply too hard for most teachers to incorporate student computer use as a regular part of their instructional practice. Moreover, computers are hard to master, hard to use, and often break down; therefore, investing effort into having students use them frequently is hardly worthwhile, and we should not expect many teachers to make this effort. Finally, all too often, district or school administrators have placed computers in teachers' rooms with the expectation that computers will become part of the teacher's instructional repertoire, even though the teachers did not ask for them and did not have specific plans for using them (Cuban, 1986; Cuban, 2000). (Note 1)

Yet, although Cuban's argument may have applied in the mid-1980's, is it correct today? The capabilities and functionality of what we call personal computers have changed by orders of magnitude since Cuban first wrote about desktop microcomputer technology. What passed for classroom computers fifteen years ago seem like primitive toys today. Because the early "8-bit" computers that dominated schools' installed base in 1985 stored, processed, and displayed information at a tiny fraction of the capacity and speed of today's computers, they required much more patience and personal interest in the technology itself than current technology demands. For example, in the mid-1980's, a serious computer-using teacher would have had to keep track of programs and student files on dozens of different floppy disks, but today the widespread use of hard disks and local area networks has eliminated much of that shuffle of materials. Software applications that in earlier years were frustratingly slow or markedly limited in their functionality have matured a great deal, providing much more in the way of on-line user help, even as they have come to provide more functionality. Moreover, the instructional possibilities that computers provided to teachers were much narrower then than now. New applications have evolved that hardly existed ten or fifteen years ago—electronic mail, the World Wide Web, software for presenting digital slide shows, student-created multimedia authoring environments, and digital video-editing, just to name some. Today, advocates for teachers using computers regard these new applications, embedded in current computer and communications technology infrastructures, as learning resources of a totally different sort from what pioneering teachers bravely attempted to use a decade and a half ago.

So, have computers become more compatible with the conditions of teaching? Have their richer capabilities made them more relevant to teaching objectives? Do they now constitute resources with potential for significantly changing and improving the nature of school learning? Have teachers themselves become more skilled and knowledgeable about using computer software and hardware with their

students? Or is Cuban right even today: Are computers really a mismatch with the requirements and conditions of teaching?

The Teaching, Learning, and Computing Survey

Data from the 1998 national survey of teachers, Teaching, Learning, and Computing (TLC), suggests that Cuban's argument that teachers' "intractable workplace conditions" do still limit widespread classroom use of computers. However, under the right conditions—where teachers are personally comfortable and at least moderately skilled in using computers themselves, where the school's daily class schedule permits allocating time for students to use computers as part of class assignments, where enough equipment is available and convenient to permit computer activities to flow seamlessly alongside other learning tasks, and where teachers' personal philosophies support a student-centered, constructivist pedagogy that incorporates collaborative projects defined partly by student interest—computers are clearly becoming a valuable and well-functioning instructional tool.

In the TLC survey, more than 4,000 teachers in over 1,100 schools across the U.S. described their educational philosophies and characteristic teaching practices, their uses of computers in teaching, and various aspects of their school's environment. The survey included a nationally representative sample of 2,251 4th through 12th grade teachers as well as more than 1,800 other teachers from two targeted samples of schools—schools with the greatest presence of computer technology and schools that participate in one of more than 50 identified national or regional educational reform programs. Roughly 75% of the schools sampled for the study participated and nearly 70% of the teachers sampled within those schools completed 20-page survey questionnaires. (Note 2)

In this article, I discuss some of the findings of this survey as they relate to the questions raised by Cuban's critique: Are teachers using computers with their students? Which teachers are doing so? What are their teaching objectives for students' computer use? How are those objectives met by using computers? Do certain approaches to using computers have an impact on students and on their teaching in general? What types of teachers are making these changes, and what conditions permit this to happen?

The Most Common Frequent Uses of Computers Are in Computer Classes and Business Classes

Although computers in schools by now number over 10 million, frequent student experiences with school computers occur primarily in four contexts--separate courses in computer education, pre-occupational preparation in business and vocational education, various exploratory uses in elementary school classes, and the use of word processing software for students to present work to their

teachers. The one area where one might imagine learning to be most impacted by technology—students acquiring information, analyzing ideas, and demonstrating and communicating content understanding in secondary school science, social studies, mathematics, and other academic work—involves computers significantly in only a small minority of secondary school academic classes.

Figure 1 shows the proportion of teachers, by subject, who reported that a typical student in one of their classes used computers on more than 20 occasions during class over roughly a 30-week period. (Note 3) Apart from computer education teachers, a majority of only one other group—business education teachers—reported computer use occurred that frequently in their classes. About two-fifths of vocational education teachers and elementary teachers of self-contained classes also reported frequent (i.e., roughly weekly) use. Among secondary academic subject teachers, the highest rate of frequent use was reported by English teachers (24%). Only one out of six science teachers, one out of eight social studies teachers, and one out of nine math teachers said students used computers that often during their class. Given the distribution of course-taking patterns in high school, it turns out that a majority of students' intensive computer experiences occur outside of academic work, as part of computer education or occupational preparation.

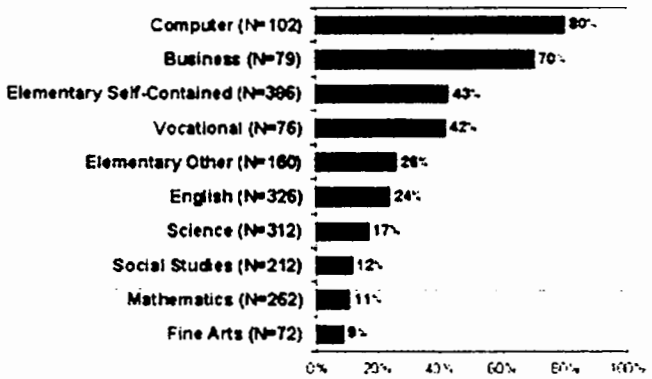


Figure 1. Frequent Student Use of Computers by Subject
(i.e., Typical Student Used Computers in Class More Than 20 Times Over Most of School Year)
[Sample: National probability sample. Three groups of teachers omitted: secondary foreign language teachers (N less than 50), secondary teachers of mixed academic subjects (no subject taught for a majority of the school week), and secondary teachers of other applied subjects.]

Why is this the case? From the survey's findings, there appear to be at least five elements to an explanation.

One problem is scheduling. Most secondary students have a continuous block of less than one hour's duration to do work in any one class. That time limit constrains the variety of learning modalities their teachers can orchestrate. As a result, fewer teachers plan

BEST COPY AVAILABLE

computer activities on a regular basis. In the TLC survey, secondary academic teachers who work in schools with schedules involving longer blocks of time (e.g., 90-120 minute classes) were somewhat more likely to report frequent (i.e., roughly weekly) student use during class (19% vs. 15%), even though they met those classes on perhaps half the number of days as teachers who taught in traditional 50-minute periods.

A second issue is the pressure of curriculum coverage. Teachers of academic subjects are strong believers in transmitting a large amount of information or skills during the course of a year. Our data show that secondary mathematics and social studies teachers and high school science teachers believe more strongly than other teachers of the importance of broad content coverage of their curriculum. In addition, many teachers feel pressured by administrator expectations for content coverage, particularly content to be covered on high-stakes tests. Those pressures are strongest among elementary teachers, math teachers, middle school social studies teachers, and high school English teachers. Computer use is often seen as inhibiting the coverage of topics. In fact, the relatively few academic teachers whose pedagogy involves "a small number [of topics] covered in great depth" (only one out of every thirteen academic secondary teachers in the study) are twice as likely as those who report covering a large number of topics to assign computer activities to their students on a frequent basis (29% vs. 14%).

A third issue has to do with convenient access to computers. This factor is so important, it deserves special consideration.

Classroom Access to Clusters of Computers: More Frequent Use Than Labs Produce

Across the various subjects taught in school, there is a strong relationship between how frequently students use computers during class time and whether their classroom has a substantial number of computers present. Those school subjects where teachers are more likely to have a 1:4 ratio of computers to students (that is, one computer for every four students) are the same subjects where frequent use of computers is more likely. Figure 2 shows this quite clearly: the subjects where frequent student use is common (the long bars coming from the left edge to the 100% bar in the middle) are the subjects where clusters of classroom computers are also more common (the long bars coming from the right edge to the middle). The only real discrepancy in the pattern is that elementary teachers of self-contained classes have students use computers more frequently than one would predict solely based on how many computers they have in their classroom. The obvious explanation is that elementary teachers have their students for most of a school day rather than 50 minutes at a time. Thus, they have a greater opportunity to provide frequent computer experiences for each student. However, at the secondary level, where 50-minute instructional periods are the norm, the pattern is very strong: in math, social studies, and foreign languages, the subjects

where students use computers the least often, very few teachers have more than one or two computers in their classroom.

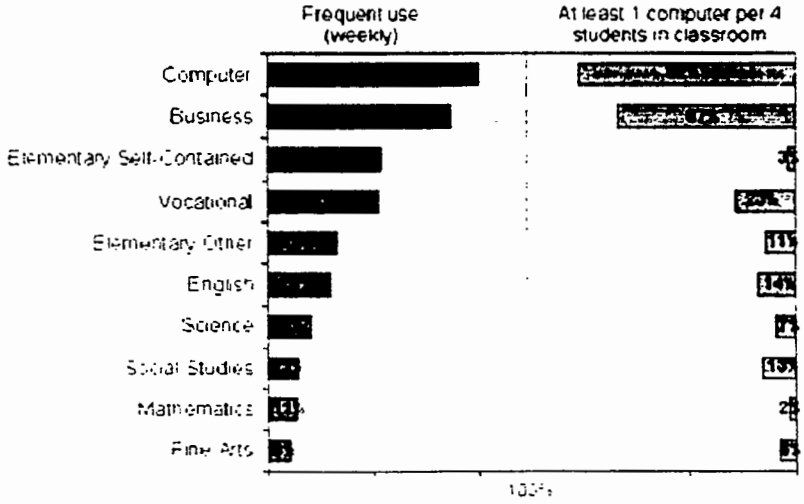


Figure 2. Frequent Use and Classroom Access by Subject
[Sample: For statistics on computer to student ratios of 1:4, 50% random subsample of teachers in the national probability sample. For statistics on frequent use of computers, see Figure 1 for additional details.]

Of course, most teachers have the option of using computers in shared spaces such as computer labs or media centers, where large numbers of computers may be present. (The typical computer lab has 21 computers.) However, despite such settings having so many more computers than in most classrooms (the typical number of computers in classrooms that have any at all is still only 2), teachers with a reasonable number of computers available in their own class are much more likely to provide frequent opportunities for students to use computers than when they have to make use of a computer lab. Specifically, we found that secondary academic subject teachers who have 5 to 8 computers in their classroom are twice as likely to give students frequent computer experience during class than teachers of the same subjects whose classes use computers in a shared space with a minimum of 15 computers present. (See Figure 3.) This may seem counter-intuitive since being in a lab with three times as many computers as these classrooms have would seem to be preferable. However, the scheduling of whole classes of students to use computers, at wide intervals determined well in advance of need (i.e., weekly or every-other-week use scheduled weeks in advance) makes it almost impossible for computers to be integrated as research, analytic, and communicative tools in the context of the central academic work of an academic class.

BEST COPY AVAILABLE

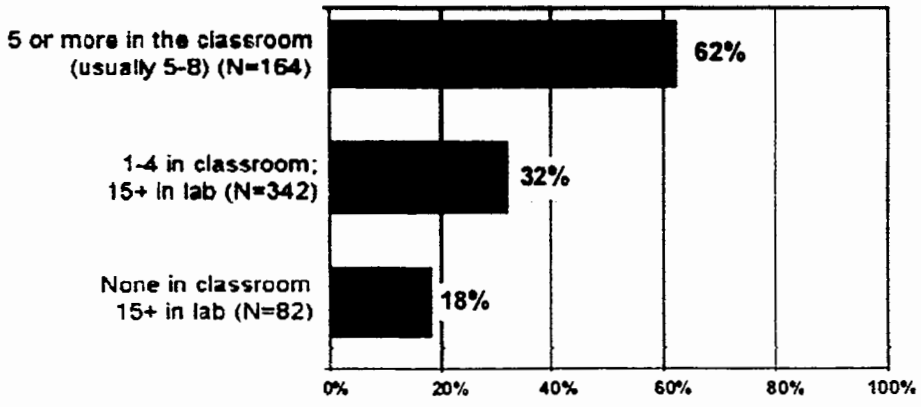


Figure 3. Frequent Computer Use by Location and Number of Computers Available (Selected Combinations), For Secondary Academic Teachers

[Sample: 50% random subsample of teachers who used computers with their selected class in both probability and purposive samples. A fourth access category is not shown—teachers with 0-4 computers in classroom and under 15 in a lab or other outside location, if available.]

This analysis does not take into account the economies that centralized placement of computers involve. In other words, if a school's 12 science teachers, for example, each had five computers in their classrooms, this would require twice as many computers than if they all shared one computer lab with 30 computers in it. Instead, what we are examining is the relative likelihood that students will receive a substantial computer experience during instructional time. If the 12 science teachers each taught five classes of students, the 60 classes would have at most only one opportunity to use computers in the lab every two weeks. On the other hand, if the computers were constantly present in every student's science classroom, one would expect them to have more opportunities to use computers for doing scientific work, particularly if their teachers' instructional practice enabled different students to be using different resources at the same time. (Note 4) If centralized placement of computers does not result in students getting a substantial experience with using computers in doing academic work, the apparent economies of scale are not likely to be cost-effective in the end.

Teacher Expertise and Comfort in Using Computers Professionally

Besides inconvenient access to clusters of computers, besides problems of overly-scheduled secondary schools, and besides problems related to having a large amount of curriculum to "cover," another element that prevents more teachers from using computers frequently with their students is their own limited skill and expertise in using computers themselves.

Many teachers have learned information technology skills and put them to use over the past five to ten years. A majority of the teachers in the nationally representative TLC sample said they know how to use a World Wide Web search engine, more than a third said they would be able to create a new database and establish fields and screen layouts, and one-

fourth said they could prepare a slide show using presentation software. Nearly one-third report using either camcorders, digital cameras, or scanners at least occasionally, and many teachers have even posted ideas, lesson plans, or student work on the World Wide Web. (Note 5) On the other hand, the most widespread professional uses of software by teachers are fairly routine—preparing handouts, writing lesson plans, and recording and calculating grades. And although most teachers *do* report using the Web to get information to use in their lessons, most do so on a relatively infrequent basis. At least that was the case in 1998, when the survey was conducted.

But do the teachers who have those skills and who regularly use computers for their own purposes use computers more frequently with students or do so in a different way than less computer-knowledgeable teachers? Cuban (2000) argues that insufficient technical skills is not holding back teachers' classroom use of computers. However, our data suggests that they are. Teachers who have an above- average amount of technical skill and who use computers for their own professional needs use computers in broader and more sophisticated ways with students than teachers who have limited technical skills and no personal investment in using computers themselves. (Note 6)

To conduct this analysis, we divided teachers into equal- sized groups based on an index measuring the variety of their self-reported computer skills, the different ways they used computers professionally, and how extensive their experience was on different computer platforms. (Note 7) The teachers in the top 25% on that Computer Knowledge index, on average, had students use three times the number of types of software as did teachers in the bottom 25%. (Note 8) Figure 4 shows that the pattern is even stronger for teachers of individual secondary academic subjects. The biggest difference is between teachers in the upper 25% and the rest of the teachers; that is, the math, science, English, and social studies teachers who are most skilled and involved in using computers themselves account for most of the situations where students use a variety of software to do work for their academic classes.

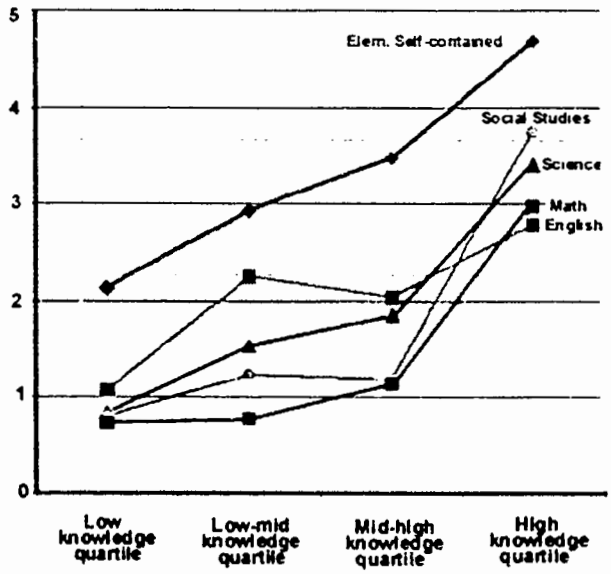


Figure 4. Breadth of Student Software Use (Number of types of software used by students in 3 or more lessons) by Teacher's Computer Knowledge by Subject Taught
[Sample: All teachers in probability sample. Vertical axis indicates the mean number of different types of software (out of 10) which the teacher reported having students in her selected class use in at least 10 lessons during the school year.]

BEST COPY AVAILABLE

Several types of software were much more likely to be used in classes taught by the more computer-knowledgeable teachers: (1) presentation software such as Powerpoint, (2) World Wide Web browsers, (3) electronic mail, (4) spreadsheets and database software, and, (5) in English, social studies and elementary classes, multimedia authoring software. The one type of software that was clearly NOT used by students of these computer-knowledgeable teachers more than by students of other teachers is skills-practice software, i.e., traditional computer-assisted-instruction. (The more knowledgeable teachers didn't have students use skills practice software *less* than other teachers; they just used other types of software much more.) Table 1 shows, subject by subject, the correlation coefficients between the Computer Knowledge index and how extensively teachers in that subject used different types of software with their students. (Note 9)

Table 1
Correlations Between Teacher Computer Knowledge-Professional Use and Extent of Instructional Use of Different Types of Software, By Subject Taught

	English	Social Studies	Science	Math	Other Secondary	Elementary
Skill Games	0.14	-0.01	0.02	-0.08	-0.01	0.08
Simulation/Exploratory	0.09	0.28	0.23	0.14	0.19	0.21
CD-ROM Reference	0.16	0.23	0.21	0.23	0.10	0.21
Word Processing	0.24	0.29	0.21	**0.32**	0.22	0.29
Presentation Software	**0.38**	**0.32**	**0.34**	0.25	**0.36**	0.27
Graphics Oriented	0.28	0.11	0.05	0.24	0.25	0.23
Spreadsheet/Database	0.21	0.28	0.28	**0.32**	**0.31**	0.19
Multimedia Authoring	0.25	**0.31**	0.16	0.16	**0.34**	**0.32**
WWW Browser	**0.30**	**0.45**	0.15	**0.36**	0.27	**0.31**
E-Mail	0.25	**0.31**	0.27	0.20	0.21	0.24

[Sample: All teachers in probability and purposive samples. Boldface numbers indicate correlations of .30 or above.]

One might ask, however, why the differences in Figure 4 and Table 1 are not even greater than they are. Our evidence suggests that a powerful limitation on broadening teachers' use of computers with students derives from teachers' personal philosophical beliefs about the basic nature of student learning and what type of instruction is optimal given their own implicit theory of learning.

Teaching Philosophy and Objectives for Computer Use

Traditionally, teaching practice has been characterized by an emphasis on skill and knowledge transmission from teacher to students. This usually involves

1. the use of an externally prescribed curriculum of discrete skills and factual knowledge;
2. direct presentation and explanation to students of that procedural and factual knowledge;
3. frequent assignment of written exercises to students aimed at their remembering factual knowledge and accurately performing skills; and then
4. evaluation of students' mastery of skills and knowledge by giving them written tests that prompt students to recognize factual statements and to apply learned algorithms and other skills to produce correct answers.

Transmission pedagogy derives from a conventional theory of learning in which understanding arises from carefully planned direct instruction on a narrowly defined skill or content topic and guided practice on questions related to that topic. Such a pedagogy is similar to conventional (i.e., culturally normative) beliefs about learning, and is therefore part of most teachers' own schooling experiences. Moreover, assessment of factual knowledge and specific skills can be accomplished with a fair degree of reliability and validity, both through teacher-constructed tests and in the kinds of large-scale external assessments on which teachers are increasingly judged. Using such tests as measures of academic accomplishment, transmission pedagogy has been supported by a good deal of evidence from studies of reading, language, and arithmetic instruction, particularly in the elementary grades.

However, transmission pedagogy and the tests which certify its accomplishment are primarily oriented towards only a narrow range of academic competencies, those emphasizing isolated mental processing on tasks with only a surface resemblance to deep understanding of a domain. Even the most recently constructed large-scale assessments of student achievement may have a built-in bias towards a transmission model of instruction and fail to capture a range of important competencies. Take, for example, the challenge of extracting from a large, messy collection of information and ideas a subset of evidence that is most relevant to constructing a good argument about a controversial issue; developing an argument that addresses the issue in consultation with other classmates, outside resources, and using analytic tools available; and then making the most cogent presentation possible to an audience that personally cares about this issue. Most "standards-based" assessments would not even attempt to judge students' abilities to give such a "performance of understanding" (Perkins, 1998), in part because the "standardized" nature of such an assessment would not permit students to employ any analytic tools or information resources that they happened to have experience with, such as computer software, that might be relevant to accomplishing the task.

At any event, our data suggests that academic subject-matter teachers who use computers most productively in grades 4-12 are not very comfortable with a transmission-oriented pedagogy, even though that is the approach which may satisfy policy-makers and large portions of the public through its assumed ability to result in higher standardized test scores. The most computer-engaged teachers, instead, appear to endorse an alternative philosophy of teaching, which might be explained as including two pedagogical emphases:

1. attending to the "meaningfulness" of instructional content for each student—for example, by developing examples connected to students' own personal experience or by providing opportunities for students to present detailed explanations of their reasoning; and
2. developing students' capacities to understand a subject deeply enough, and see the

interrelationships of different ideas and issues, so they are able to know how and when to apply their knowledge to particular contexts and communicate their understandings to others.

Both of those emphases require substantial amounts of time and teaching expertise to put into practice, and both usually conflict with the objective of covering large amounts of curriculum.

These two emphases are associated with the theory of learning called "constructivism." Constructivist theory claims that understanding comes from a person's effortful activity to integrate newly communicated claims and ideas with his own prior beliefs and understandings. In that view, understanding cannot be transmitted, nor does skills-practice result in understanding which can be automatically applied as needed. Instead, effective teaching involves creating environments in which students take mindful effort towards developing their understanding and have opportunities to learn *how* to apply their knowledge and *when* to do so. Instruction is particularly valued that gets students to articulate their understandings and defend them against contrary points of view. Many ways of using computers lend themselves to instruction based on a constructivist model of learning—for example, presentations to a critical audience, integrating different perspectives in a report or multimedia document, or examining contrary assumptions using a spreadsheet model.

The way that a teacher uses computers gives an indication of her underlying pedagogical philosophy. Of course, any computer application *could* be used in a transmission- oriented pedagogy. That is, a teachers *could* focus students' use of multimedia, word processing, or spreadsheet software by teaching them a set of technical skills primarily so they can master the software itself. However, apart from school subjects where such skills are expected to be taught—computer education courses or business education courses—teachers would generally not have students use complex software unless they found that it facilitated learning in the subject they teach. Thus, in academic subjects, we would predict that teachers who believe in a more traditional transmission-oriented approach will find most applications of computer technology incompatible with their instructional goals, and will therefore use a more limited range of computer applications.

To examine this argument empirically, the TLC survey asked teachers a relatively extensive set of questions designed to measure their philosophical preference between transmission- oriented teaching and constructivist-compatible teaching. We found clear relationships between teaching philosophy and (a) *whether* a teacher used computers with students; (b) the particular objectives for computer use the teacher had; and (c) the types of software used frequently with students. Moreover, constructivist-compatible teaching objectives for computer use (i.e., those most associated with constructivist teaching philosophies) were also found to be associated with a greater amount of school-related computer activity by students, before- or after-school or at home—that is, on the students' own time. Finally, teachers who used computers in a constructivist way reported making more general changes to their characteristic pedagogy than did teachers who used computers in a more limited way or not at all. The remaining set of figures and tables illustrate those findings.

Teachers' Philosophical Positions

Survey questions about teachers' philosophy were of several types. In one type, teachers were given two alternative statements of teaching philosophy—for example, a statement that argued for structured presentation and explanation of information versus a statement that argued for the teacher being a provider of resources for

students "to construct concepts for themselves." In another set of questions, two teachers' contrasting practices of conducting recitations were described. One teacher asked a rapid series of direct questions, designed to keep students attentive and on-task. The other teacher encouraged questions from students, and used these as springboards for suggesting student-initiated research activities.

Overall, teachers' responses reflected quite varying philosophies. For example, about 40% of teachers felt that the teacher acting as facilitator was preferable to giving structured explanations, while 30% felt the reverse was true and 30% gave the middle or ambivalent response. (Note 10) Slightly more teachers felt that rapid-fire direct-questioning teaching resulted in students gaining more knowledge than the opposite approach, but a majority of teachers felt that "skills" would be learned more in the class where teachers led students towards their own investigations into their own questions. (Note 11)

Other survey questions suggesting a transmission-oriented philosophy dealt with the value of a quiet classroom for learning, the importance of background knowledge and basic reading and math skills for "meaningful" subject- matter learning, having the teacher be the sole determinant of classroom activities, and building instruction around problems with clear, easily found, single correct answers. Questions (and responses) suggesting a constructivist philosophy argued for the value of "sense-making" over curriculum-coverage, the utility of organizing a class with multiple activities occurring simultaneously, the value of student interest and effort in academic work over the particular content covered in subject textbooks, and having students play a role in establishing criteria for evaluating student work.

To analyze these competing philosophical viewpoints about teaching, we created an index combining answers to these 13 different prompts (alpha = .83). We divided teachers into four equal-sized groups, from the quartile who most valued a transmission approach to the quartile who most valued a constructivist approach. Not surprisingly, elementary teachers turn out to be more constructivist than secondary teachers, with 32% of the elementary teachers in the "high constructivist" quartile compared to 21% of secondary (middle and high-school) teachers. (Middle school academic subject teachers are about half-way between the high school and elementary group.)

Computer-using teachers—that is, teachers who have their students do any computer work during class at all—are distinctly more constructivist than non-using teachers. Among elementary teachers, relatively infrequent users are no less constructivist than teachers who have students use computers a lot. However, among secondary academic subject teachers, the teachers who assign computer work frequently are much more constructivist than those who make computers are less central part of their pedagogy. (See Figure 5, lower panel.)

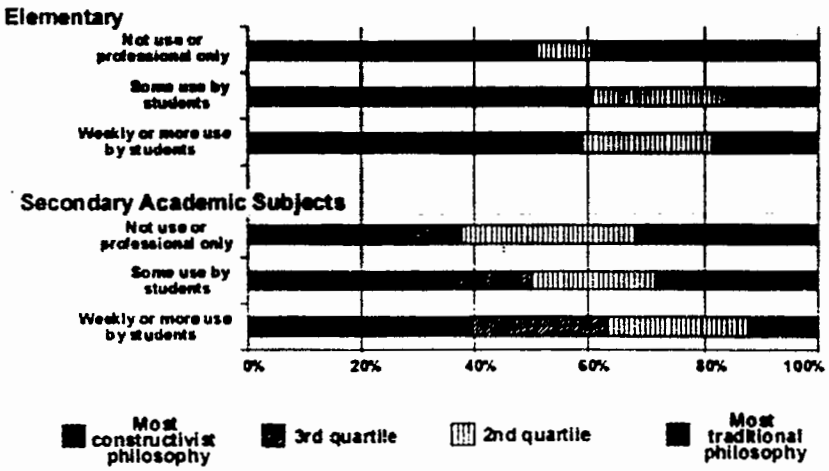


Figure 5. Frequency of Computer Use by Teacher Philosophy By General Teaching Responsibility
[Sample: All teachers in probability sample.]

Computer-Using Teachers' Objectives for Student Computer Use

There is a strong relationship between teachers' general philosophical viewpoint about what constitutes good teaching and the particular objectives they view as most central to their use of computers with students. The survey asked teachers to select three objectives from a list of ten that were their most important objectives for student computer use. The objectives most commonly supported by computer- using teachers were "getting information or ideas" and "expressing themselves in writing." Mastering skills, both academic skills and computer skills, were less often cited, but "skills" as objectives were much more often cited than such objectives as "presenting information to an audience" or "communicating electronically with other people." (See Figure 6.)

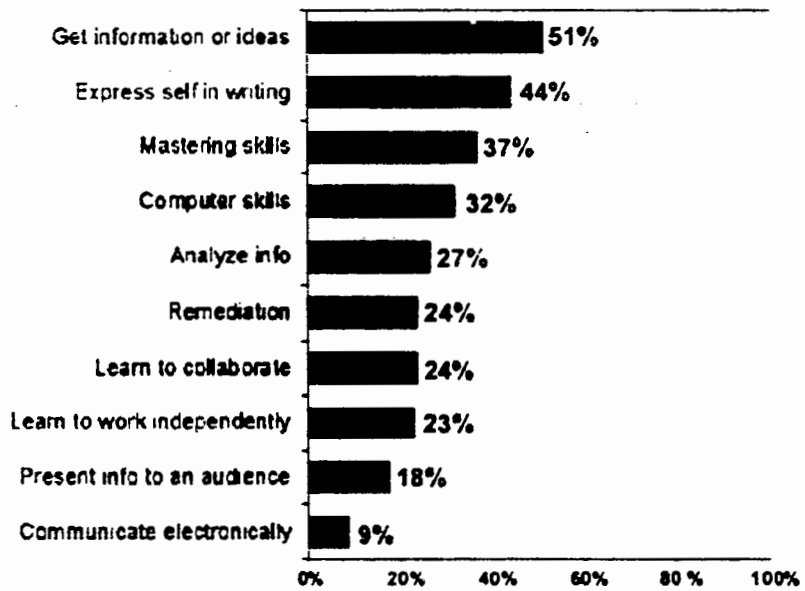


Figure 6. Teachers' Primary Objectives For Computer Use (Percent of teachers who report the objective as being among their 3 most important ones).
[Sample: Probability sample; teachers who used computers with their selected class.]

The relationship between objectives and teaching philosophy is shown in Figure 7, where objectives for computer use are ordered according to how "constructivist" teachers were in terms of their survey answers to questions about teaching philosophy. (Note 12) Figure 7 shows that the relatively small minority of computer-using teachers who selected having students "communicate electronically with other people" (only 9% of all computer-using teachers) had, overall, the most constructivist philosophies. The next-most philosophically constructivist teachers were those who chose "presenting information to an audience" and "learning to work collaboratively" as their main objectives for student computer use. Teachers who selected "getting information or ideas" or "expressing themselves in writing" were also more constructivist than most teachers overall, but about average when just considering teachers who used computers with students.

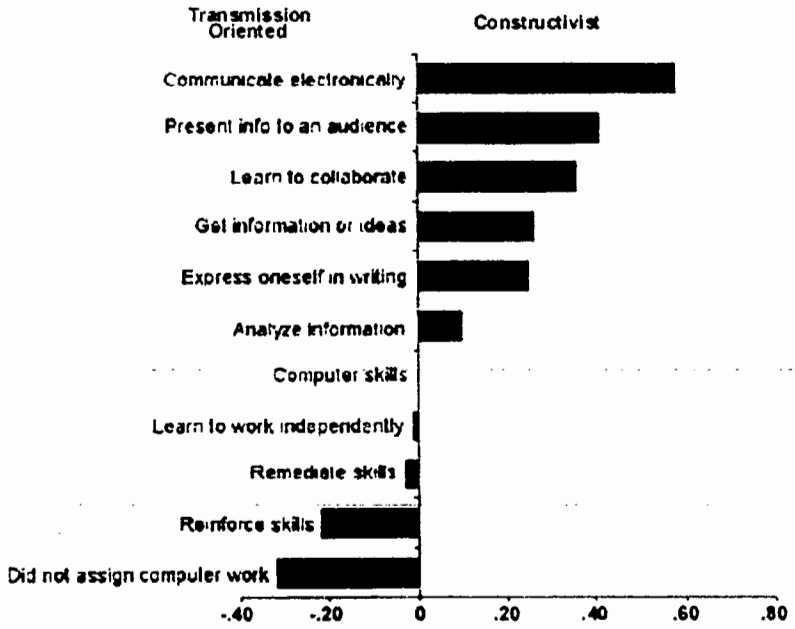


Figure 7. Objectives For Computer Use Are Also Linked To Teaching Philosophy (mean z-score on Teaching Philosophy Index)
[Sample: Probability sample; teachers who used computers with their selected class.]

In contrast to those teachers, the 36% of computer-using teachers who selected skills reinforcement as one of their top three objectives ("mastering skills just taught") reported much more transmission-oriented philosophies than teachers who chose other objectives. However, even the skills-reinforcement-valuing teachers

BEST COPY AVAILABLE

were somewhat more constructivist (i.e., less transmission- oriented) than the teachers who didn't have students use computers at all.

Types of Software Used by Teachers Who Assign Computer Work Frequently

The rapid progress of computer technology over the past decades has meant an increasing variety of software has become available for teachers to use with students. During the 1980's, teachers could have students program in BASIC or LOGO, use drill-and-practice software, simple word processing programs, or some inventive problem-solving puzzles and simulations, but not much else. The range of possibilities has grown enormously since then. Our survey asked teachers to name the software that has been most valuable in their teaching—the best computer programs their students have used. Table 2 shows that general office tool software clearly dominates the list of the programs most commonly named as "most valuable."

Table 2
Specific Software Reported As "Best" or "Most Valuable" For Students by Computer-Assigning Teachers, by Subject & Level of Teacher

	Percent of All Computer-Assigning Teachers (naming at least one program as "best')*			
	20%+	**15-19%**	**10-14%**	**5-9%**
Elementary Self-contained	ClarisWorks		Hyperstudio	**Accelerated Reader****, Encarta, Groliers, M. Word, Netscape, **Oregon Trail**, Writing-Pub. Center
Elementary Other	ClarisWorks	**Accelerated Reader**	Hyperstudio	Groliers, M. Works, Netscape, Writing-Pub. Center
English		ClarisWorks, M.Works	M. Word, Netscape	**Accelerated Reader**, Powerpoint
Science		ClarisWorks, Netscape		**Hyperstudio**, M.Office, M.Word,

				M.Works
Math	**Geometer's Sketchpad**	ClarisWorks		**Excel, Math Blaster,** M.Word, Netscape
Social Studies		ClarisWorks, Netscape	Hyperstudio	Encarta, Groliers, I.E., M.Word, M.Works, Powerpoint
Foreign Language	ClarisWorks, M.Word	**M.Publisher**	Netscape	M.Works, Powerpoint
Misc. Academic Secondary	ClarisWorks		Encarta, M.Word, Netscape	Groliers, M.Office
Computers		M.Office, Netscape	ClarisWorks, M.Word, M.Works, Word Perfect	Excel, **Hyperstudio,** Powerpoint
Business	M.Works, Word Perfect	M.Office	M.Word	ClarisWorks, Excel, Netscape
Vocational	AutoCAD	Netscape	ClarisWorks, Word Perfect	M.Office, M. Works
Fine Arts	ClarisWorks	PhotoShop	Netscape	**Hyperstudio,** M.Word, M.Works, **PageMaker**
Other Applied Secondary	ClarisWorks	M.Word, M.Works, Netscape, Powerpoint		**Hyperstudio,** M.Office, Word Perfect
Elementary	ClarisWorks		Hyperstudio	**Accelerated Reader,** M.Word, Netscape, Encarta, Groliers, M.Works, **Oregon Trail**
Middle School	ClarisWorks		Netscape	M.Works, M.Word, **Hyperstudio**

High School		Netscape, M.Works ClarisWorks, MWord	M. Office, Powerpoint, Word Perfect
All comp.-assigning teachers	ClarisWorks	Netscape	M.Word, M.Works, **Hyperstudio**, M.Office

[Probability and purposive samples; teachers who assigned computer work to selected class and who named at least one program.
* One-half of teachers responded to a question about the "best computer programs students in this class have used." The other one-half responded to a question about their most valuable software in each of the past five years. Data from the two most recent years were taken from this latter group, and only if the software did not seem to be named primarily because of its value for the teacher's own professional use.
** Software in bold are applications other than office software, Internet access software, or CD-ROM encyclopedias. They are primarily subject-specific applications or authoring tools.]

Clarisworks was by far the software title most frequently named by teachers. Three of the five next-most commonly named were *Microsoft Works*, *Microsoft Word*, and *Microsoft Office*. (The other two were *Netscape*, reflecting the importance of Web activity; and *Hyperstudio*, the primarily Macintosh-based multimedia student authoring environment named primarily by elementary teachers and middle school social studies teachers.) Some software titles focusing on specific curricular areas were frequently named as well, including *Geometer's Sketchpad* in mathematics, the inquiry- oriented conjecturing tool; *Autocad* in vocational education, which has dominates the growing field of computer-aided-design; *PhotoShop*, in fine arts classes; and *Accelerated Reader*, the computer-based test library used in off-line tradebook reading programs in elementary and middle-grades reading and language arts programs.

But overall, what is the balance of different types of software that teachers use on a frequent basis with students, and what teaching philosophies and instructional objectives do these types of software reflect?

Although many teachers have students use a variety of software at least occasionally, the only type of software which commands both broad use (across subjects) and frequent use (used by students for at least 10 lessons) is word processing. Frequent use of all other applications is limited to at most one or two specific subjects (usually computer education). Table 3 shows the percent of teachers, by subject, who reported having had students use each of ten types of software for at least 10 lessons during the school year. Highlighted in Table 3 are the types of software where at least one-fourth of all teachers of a given subject reported that level of frequent use. Word processing reaches the "frequent-use one-quarter penetration" criterion for elementary teachers and secondary English, computer education, and business teachers and nearly approaches that level for science and social studies teachers. Nearly half of all computer education and business teachers also report having students use spreadsheet and database software frequently.

Table 3
Percent of Teachers Reporting Frequent Student Use (Use In At Least 10 Lessons), by Type of Software and Subject & Level Taught

	Word Proc.	CD-ROM	WWW	Skill practice games	Simulations/ Exploratory Environments	Graphics	Spread-sheets/ Database	Presen-tation	Multi-media	Email
Second.										
English	38%	10%	12%	4%	2%	6%	2%	4%	2%	2%
Science	24	15	22	3	5	6	8	5	2	4
Math	4	2	4	13	8	1	4	2	1	1
Soc. Stud.	20	16	14	2	3	5	4	5	6	6
Comp.	76	15	38	20	22	23	43	29	13	9
Bus.	78	3	14	16	19	20	47	22	2	5
Vocat.	15	12	15	1	21	16	7	6	3	3
Fine Arts	10	4	7	0	1	15	0	1	2	1
Elem.	46	28	12	32	11	10	2	3	5	3
All teachers	32	16	13	14	7	8	7	5	4	3

[Sample: All teachers in probability sample.]

There are only three remaining combinations of subject and at least one-fourth penetration of frequent use: the World Wide Web in computer education classes, and skills-practice software and CD-ROM software among elementary teachers. (Note 13) For many other combinations where it is reasonable to think that a given type of software ought to be relevant to learning in a particular subject, we find that fewer than 10% of the teachers of that subject actually have students use that type of software on a frequent basis—for example, spreadsheets in math instruction (4% of all secondary math teachers use it in 10 or more lessons), simulations in science (5%), presentation software in English (4%), multimedia authoring software in social studies (6%), and electronic mail in business education (5%). Clearly, there are large gaps between the potential penetration of many types of software in academic classes and the current proportion of teachers who are actually making use of that software in their classes. With overall patterns of software use like these numbers suggest, Cuban's major claim is clearly supported. Frequent use of most computer applications is still a minority teaching practice.

Constructivist Philosophy and Teachers' Frequent Use of Computers with Students

But what of the minority of teachers who *do* make substantial use of different types of software as part of the way they orchestrate student activity during their class time? Do users of only *some* types of software stand out as being constructivist, or are *most* types of software use associated with having a constructivist philosophy? (Note 14) And how different in philosophy, overall, do these teachers look from the "average" teacher who might have her students use software only occasionally?

Our data suggest that teachers of academic subjects, both elementary and secondary, who use *most* types of software on a frequent basis have consistently more constructivist philosophies than the average teacher. Electronic mail assigning-teachers (that is, the 3% of academic subject teachers who have students use electronic mail on a regular basis) and the almost as small percentage of teachers whose students often use presentation software like Powerpoint (4%) have the most constructivist philosophies of all, with roughly half of them being in the "high constructivist"

quartile of teachers, as shown in Figure 8. (Note 15) But, in fact, frequent users of *most* types of software are more constructivist in philosophy than more typical teachers are. All categories of frequent software-users are except those who use *only* skill games frequently. Even skill games users are more constructivist than average if the games are part of a practice that uses other types of software frequently as well. The teachers 3rd-ranked in terms of constructivist philosophy (the 5% who are frequent users of multimedia authoring software) and the 9th-ranked category (the 13% who assign students to do Web work frequently) are closer in philosophy to one another than either is to the larger number of teachers who only occasionally have students use computers. Again, Cuban appears to be correct that technology integration has been accomplished by a relatively small group of academic subject-matter teachers who are significantly different than their peers in terms of teaching philosophy.

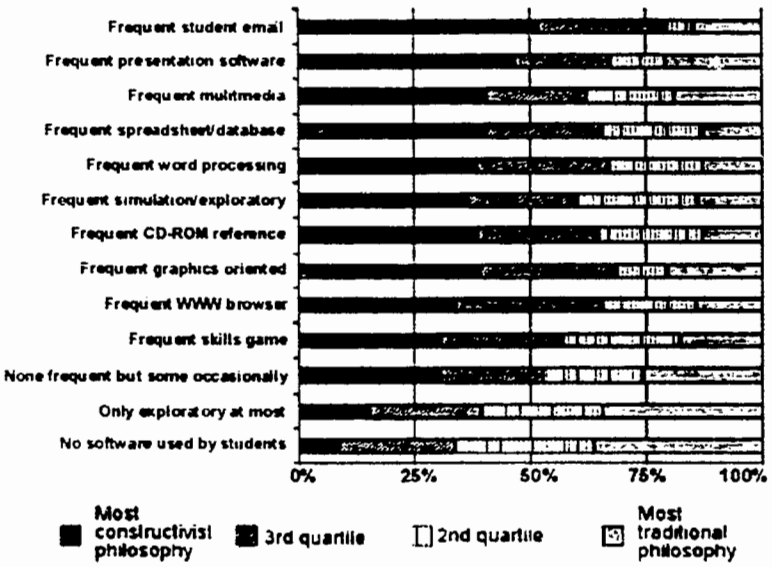


Figure 8. Frequent Use of Software (In 10+ Lessons) by Teaching Philosophy
[Sample: Probability sample; academic secondary and elementary teachers only.]

When Favorable Conditions are in Place: Compatible Philosophy, Access, and Expertise

If the teachers whose students use software frequently have substantially more constructivist philosophies than most teachers, does it follow that *most* constructivist teachers are computer users? Our data show that, by itself, a constructivist philosophy raises the chance that an academic subject-matter teacher will use many types of software frequently with students, but rarely is a compatible philosophy itself sufficient to boost a majority of teachers into assigning a certain type of computer work frequently. For example, consider middle and high school science teachers. Of *all* science teachers, only 5% reported having students use *simulations or exploratory environments* in at least 10 lessons during the year (shown previously in Table 3). Among the most constructivist quartile of teachers, proportionally twice as many did, but that is still only 10% of the science teachers in that group (see Table 4). In addition, overall, 24% of science teachers had students use *word processing* frequently, but 39% of the high-constructivist science teachers did—nearly two out of every five, but still not a majority. To take another example, in social studies, no type of software was used frequently by at least one- fourth of all social studies teachers (shown in Table 3). For the high-constructivist social studies teachers, though, three types of software had that level of penetration—word processing, CD-ROM reference materials, and World Wide Web browsers. Nevertheless, the boost was modest, at best; none of those types of software involved even one-third of the high-constructivist social studies teachers on a frequent basis. The only type of

BEST COPY AVAILABLE

software to be used frequently by a majority of high- constructivist teachers was word processing, by elementary grade teachers (55%; see Table 4). In sum, having a compatible teaching philosophy makes frequent use of computers more likely, but by itself is insufficient to make frequent computer use a modal teaching practice.

Table 4
Percent of High Constructivist Teachers (Academic Subjects Only) Reporting Frequent Computer Use

	Word Proc.	CD-ROM	WWW	Skill practice games	Simulations/ Exploratory Environments	Graphics	Spread-sheets/ Database	Presen-tation	Multi-media	Email
English	49%	15%	22%	6%	2%	13%	3%	14%	5%	7%
Science	39	23	24	7	10	7	12	10	3	8
Math	11	4	6	9	8	3	4	2	2	3
Social Studies	28	28	25	8	11	6	9	8	3	11
Elem.	55	35	14	31	14	12	3	6	9	5
All Teachers of Academic Subjects	42	21	19	13	10	11	8	10	6	6

[Sample: All academic teachers in probability and purposive samples. "Frequent Use" defined as students in any of the teacher's classes having used that type of software on 10 or more occasions.]

However, when we add in two other facilitating conditions—convenient access to a cluster of computers and the teacher having at least average levels of computer knowledge—the story changes. For this analysis, we have to combine teachers of the various academic secondary and elementary subjects together because otherwise the number of survey respondents to be analyzed becomes too small. We present data regarding the use of two categories of software: (1) word processing, because it so clearly dominates frequent computer use; and (2) any other type of software besides skill games, the latter being excluded because of the clearly distinct pedagogical approach it reflects. Figure 9 shows the percentage of teachers reporting frequent use of these two categories of software according to progressively more enabling conditions. Overall, 29% of all academic secondary and elementary teachers reported using word processing frequently and 28% reported using at least one other type of software frequently. When we restrict ourselves to the high constructivist quartile of teachers from the same subjects, the percentages rise somewhat, to 44% and 37% respectively. (Note 16) However, when we specify the other two important facilitating conditions—that the teacher has a cluster of five or more computers available in her own classroom and also has at least average computer skill and breadth of professional computer use—the percentages climb to well over a majority. More than three-fourths of such teachers (76%) had students use word processing in at least 10 lessons during the year, and 56% had them use some other type of software that often. (Note 17)

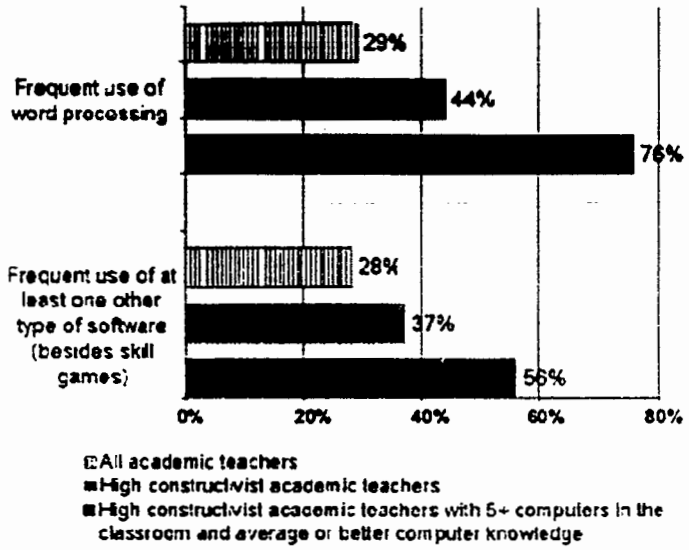


Figure 9. Frequent Use of Software by Facilitating Condition
[Sample: All academic teachers in probability and purposive samples.]

Figure 10 shows that for this group of academic subject matter teachers—that is, those with a highly constructivist philosophy who also have a cluster of computers in their classroom and at least average computer competencies and professional use themselves—not only did three-fourths have students use word processing frequently, but about one-third had students use presentation software frequently, one-third had students use the Web in 10 different lessons, a majority had students use CD-ROM reference materials on at least 3 occasions during the year, and similarly a majority had students use exploratory or simulation software at least that often. For this group, skill-based software is used less often than any of *those* applications, but it is still more common than spreadsheet work, student e-mail, or student authoring of multimedia documents.

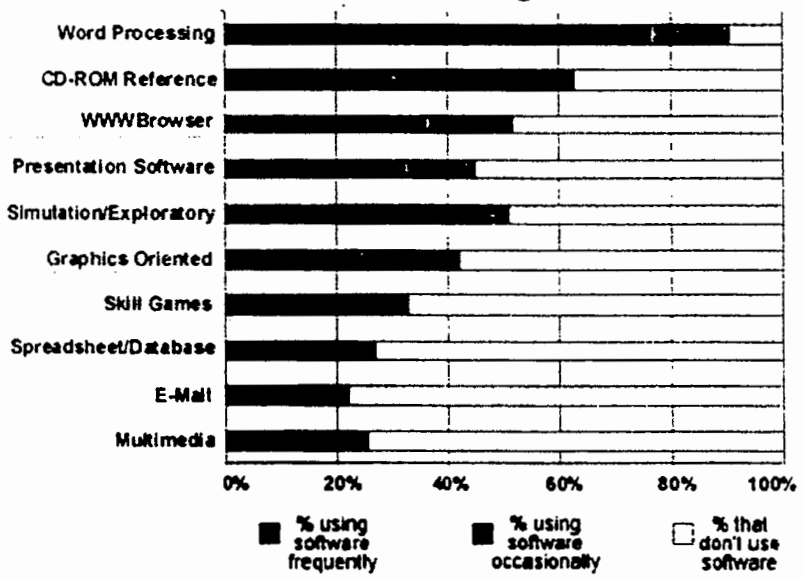


Figure 10. Software Use Among Teachers With Favorable Facilitating Conditions
[Sample: Probability and purposive samples. Teachers from the most constructivist quartile of secondary academic and elementary teachers, who have at least five computers in their classroom, and average or better computer knowledge.]

The statistics in the previous paragraph are *critical*. They demonstrate that under the right

conditions, teachers of academic subjects will make substantial use of a wide range of computer software, going well beyond routine drill-and-practice. Nevertheless, not every computer application has yet found its niche in the practice of academic subject teachers, even when many of the facilitating conditions are in place.

Outcomes of Constructivist Uses of Computers: Effects on Student Out-of-Class Effort

Demonstrating that under propitious conditions, a large fraction of teachers of academic subjects are having their students use a variety of computer applications does not necessarily prove that students are better off for this as a result. Our Teaching, Learning, and Computing survey did focus more on the "teaching" and "computing" aspects of computer use in schools than on the "learning" part, but we do have some modest empirical evidence on one interesting student outcome—students' use of computers for doing class work on their own time.

Why should simply measuring student out-of-class-time use of computers for schoolwork be considered an important outcome? For one thing, although public evaluation of schools tends to focus on the substantive facts and skills that students are being taught, a widely acknowledged goal of schooling is to foster in students a disposition to undertake learning activities on their own initiative, over the long-term. If students take initiative in doing academic work outside of the time they are being directly supervised in class, the strategies that teachers use to increase the likelihood of that happening may be as important as what they do to help students learn more during class time. Although we have a very weak measure of the out-of-class computer-use outcome—teachers' own estimates of the proportion of their students who use computers for class work at other times during the school day and the proportion that do so while at home—we can report some interesting findings related to teachers' different patterns of computer use.

We found that computer-using teachers who prioritize certain objectives for their students' computer use are much more likely than those emphasizing other objectives to report that their students use computers for class assignments during other times of the day and week. Figure 11 shows the general result and highlights four outcomes associated with greater than average out-of-class-time work and three outcomes associated with below-average levels. (Note 18) The teachers who report by far the highest proportion of students doing computer work outside of class were those whose primary objectives were having students *present information to an audience.* Asking students to prepare an oral talk before an audience seems to generate a strong motivation for students to be deeply engaged in their schoolwork—enough to keep them working after school or even at lunch. The other three objectives whose advocates reported more than average out-of-class computer work being done were these: (a) having students communicate electronically with other people, (b) having them obtain information or ideas from computer sources, and (c) having them express themselves in writing. When we distinguished the extra time spent by students while they were still at school from their efforts at home, it was clearly the time at-home which was being affected by teachers emphasizing the objectives of communications (i.e., through e-mail), information acquisition (Web), and writing (word processing). Not surprisingly, e-mail, Web browsers, and word processing programs, along with games, are the most common software applications available to students on their home computers. In contrast, where students followed their teachers' aspirations for them to prepare presentations to an audience by spending extra effort, disproportionately they did so while at school. This may be due to many assignments like this requiring collaboration among classmates, and the convenience of being able to get together as a group while at school.

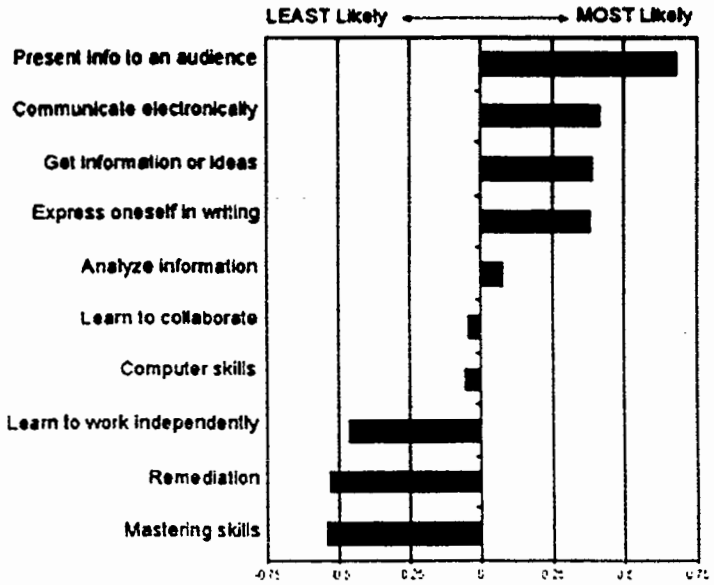


Figure 11. Which Teachers Report Student Use Computers for Class Work Outside of Class Time? (Effect Sizes)
[Sample: 50% random subsample of national probability sample; teachers who used computers with students in their selected class.]

The fact that at-home differences in students' out-of-class efforts (i.e., for teachers with different objectives) were generally greater than at-school differences is a reminder of the important role that private access to computing facilities plays in some of the types of computer work which may be most beneficial for students. We did not have information on the presence of home computers among the students of each teacher, but we did analyze the effects of teacher objectives on out-of-class effort after taking into account the socio-economic-status (SES) of the school's students and the student ability levels reported by teachers, two factors that are closely associated with home computer access. (Note 19) Table 5 shows that although class ability and school socio-economic-status are each strongly associated with student out-of-class computer work (and more strongly with at-home effort than at-school effort), teacher objectives still have effects that are independent of student characteristics. Thus, teachers whose objectives for student computer work were skills- related or "learning to work independently" (i.e., not bothering other students) reported less out-of-class computer work than teachers having other objectives, even after controlling statistically for school SES and class ability level. This was particularly true for students' doing computer work for class while at home. Similarly, at the positive end, the same objectives shown in Figure 11 remain important. In particular, teachers with presentation objectives for their students' computer work have more students doing computer work on their own time at school, and teachers with writing, information gathering, and electronic communications objectives have students who do more computer for class while at home, even after socio- economic and scholastic achievement factors are considered. (Note 20)

Table 5
Teachers' Objectives For Student Computer Use Related To Fraction of Students Reported To Use Computers For Classwork Outside of Class Time

	Use in school, outside of class	Use outside of school	All non-class time use
Correlation of non-class time use with...			
Average Ability of Students in Class (teacher estimate)	+.21	+.34	+.32
School Socio-Economic Status	+.14	+.38	+.30
Multiple correlation coefficient (control variables only; includes school level also)	.22	.44	.38
Standardized regression coefficients controlling on class ability, SES, & school level (elem., MS, HS). (each objective in separate equation)			
Present Information to an Audience	+.22	+.14	+.21
Express Oneself in Writing	+.10	+.20	+.17
Get Information and Ideas	+.06	+.21	+.16
Communicate Electronically	-.03	+.17	+.08
Learn to Work Collaboratively	+.02	-.01	+.00
Improve Computer Skills	-.01	-.03	-.02
Analyze Information	-.03	-.08	-.04
Remediate Skills	-.09	-.16	-.15
Learn to Work Independently	-.05	-.21	-.15
Master Skills Taught (reinforcement)	-.16	-.20	-.22

[Sample: Probability sample only; teachers who used computers with their selected class.]

Effects of Computer Use on Teachers: Changing Towards a Constructivist Practice

Although most discussion of the outcomes of teachers' use of computers in instruction focuses on *student* outcomes, it is important to consider how teachers' experiences with using computers might be changing their teaching practice as a whole. In particular, examination of our survey data showed us that teachers are much more constructivist in philosophy than they typically are in actual practice—no doubt the result of the many difficulties involved in doing constructivist sorts of things; e.g., having students' interests affect the topics of their classwork, orchestrating classes so that multiple activities can occur simultaneously, or having students do serious group work including engaging one another in authentic exchanges of ideas and opinions (Ravitz, Becker, and Wong, 2000).

In previous research, Becker and Ravitz (1999) proposed that when circumstances were favorable, sustained and thoughtful use of computers

as learning resources could actually help teachers implement a teaching practice that was as constructivist as their teaching philosophy would permit. In a study of 441 teachers at 152 schools of the National School Network, we found that teachers at these schools who used computers with students regularly over a three year period were roughly twice as likely to report having made a number of constructivist-oriented changes in their teaching practice as were teachers who did not use computers with their students. In particular, more than 70% reported they were now more willing "to be taught by students" than three years previously, compared to fewer than 30% among non-computer-assigning teachers. Similarly, they were much more likely to report increased skill in conducting multiple parallel activities during class time, engaging students in long projects, and giving students choices in the tasks they undertook. (See Figure 12.) In addition, supporting the argument made earlier, teachers were twice as likely to report seeing students take more initiative outside of class time. It is important to note that the schools of the National School Network were not "typical" schools. First, they had significantly more technology per-capita than average. Second, they were schools where leadership had developed strong associations with outside organizations supporting educational reform through the use of computer technology, organizations such as museums, university research projects, and private businesses. And third, the schools provided a climate supportive of curricular and instructional change.

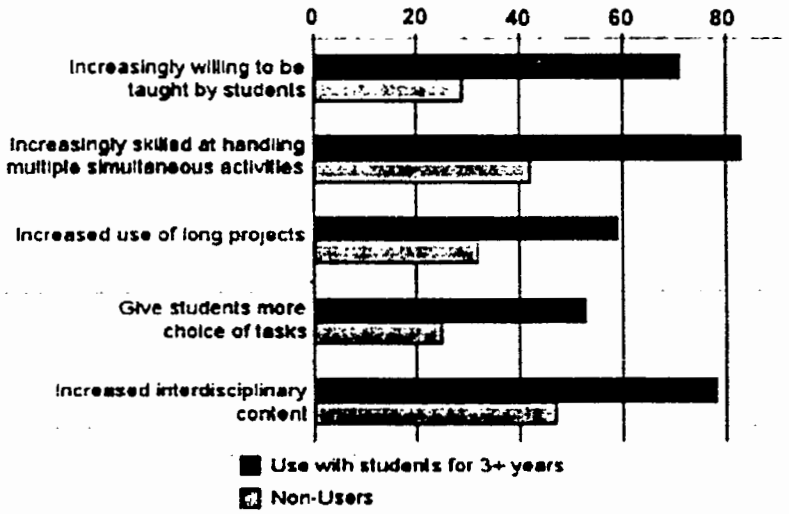


Figure 12. Changes In Their Teaching Practice Over 3 Years Reported by Computer-Using Teachers and Non-Users In the National School Network, Spring, 1997

In the Teaching, Learning, and Computing survey, we have explored similar relationships between teachers' computer use and changes in instructional practices towards a more constructivist approach to teaching. We have found, for example, that across all schools (as opposed to the relatively homogeneous schools of the National School Network) teachers who were *the least* knowledgeable about computers were also less likely than other teachers to report having become more

constructivist over the previous three years. (However, no differences have been found between teachers who were "average" and those who were "high" on our index of computer knowledge.) On the other hand, constructivist change seems to have occurred more often than typically among teachers who used a large variety of software in their teaching practice, those who used the World Wide Web a great deal in their teaching, and those whose primary objectives for computer use were having students learn to work collaboratively or to write better. (Note 21)

Those are results that generalize to all schools. However, the theory proposed in the National School Network study was that the schoolwide environment with respect to technology and instructional reform is a conditioning variable (i.e., either facilitates or impedes) the effects of computer use on pedagogical practice more generally found. That hypothesis is supported by our initial analysis of the several different independently drawn samples in the Teaching, Learning, and Computing survey.

In addition to the national probability sample of schools, the TLC survey included several different "purposive" samples—schools selected either individually or sampled from larger sets of schools specifically because of either having a large presence of leading-edge computer technology or being closely involved with programs of instructional reform, including 50 of the major national and regional reform programs (e.g., Coalition of Essential Schools, Accelerated Schools, two NSF systemic reform programs). We are finding that teachers in three groups of schools seem to have made more changes towards a constructivist teaching practice than teachers in the national probability sample: (a) teachers in the leading- edge schools with high levels of technology per capita, (b) teachers in schools with both a schoolwide emphasis on instructional reform *and* an emphasis on using computer technology in those reforms, and (c) participating teachers (and *only* participating teachers) in schools where one or two such teachers are involved in an externally organized program of technology-based instructional reform. Significantly, one group of schools does *not* show greater movement towards constructivist practices by their teachers—schoolwide reform programs that *do not* emphasize computer technology. Teachers in those schools reported, at best, the same pattern of pedagogical change as did the national probability sample of teachers. (See Figure 13.)

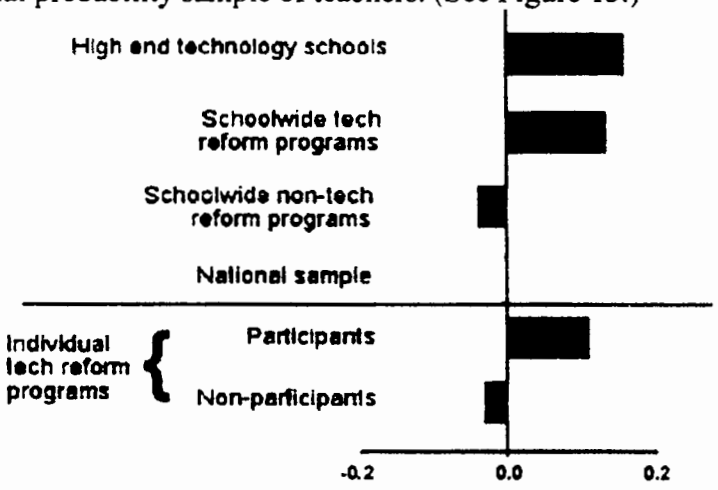


Figure 13. Constructivist Change in Teaching Compared to the TLC National Sample (Effect Sizes)

[Sample: All teachers in probability and purposive samples. Preliminary findings.]

These findings suggest that both *teacher-level* characteristics (i.e., how much they use certain computer applications and their objectives for that use) and *school-level* characteristics, such as the central role of computers in the school's character, help teachers move towards a constructivist pedagogy.

Conclusion

In response to Cuban's projection that computers are likely to continue to play a minor role in student learning of academic subjects in elementary and secondary schools, this article has presented an examination of related evidence.

On the issue of whether computers are generally a central vehicle of instructional activities in classrooms, the data suggest that Cuban remains correct up to the present time. Although a substantial fraction of teachers are having students do word processing during class time, most in-class use of computers occurs as part of separate skills-based instruction about computers, in occupationally-oriented courses such as business and vocational education, and as one of many explorations of different learning modalities that occur in the 6-hour-long days of self-contained elementary classes.

We have also found that the teachers who have students use non-skills-oriented computer software in academic classes have fairly distinctive teaching philosophies, being disproportionately supportive of constructivist pedagogies such as developing student responsibility for selecting and carrying out learning tasks, emphasizing group work involving discourse, and the use of projects, products, and performances for outside audiences.

However, this data also suggests that when constructivist- oriented teachers have sufficient resources in their classroom (i.e., clusters of 5 or more computers in a typical sized class) and have come to have a reasonable level of experience and skill in using computers themselves, a *majority* of such teachers will have their students make active and regular use of computers during their class period. That use will be principally word processing but will typically involve at least one other type of software as well, most often either CD-ROM or Internet-based information retrieval or exploratory simulation software. Other facilitating factors, such as extending the secondary classroom period from 50 minutes to significantly longer blocks of time and not only removing curriculum coverage mandates from teachers but encouraging them to teach fewer subjects in depth also can increase the number of teachers who make frequent use of computers in their plans for student class work.

Furthermore, we found that when teachers emphasize communication and information-oriented objectives for their students'

software use (i.e., publishing for an audience, communicating electronically, writing, and finding information), they expand students' academic effort from class time to free time, suggesting that a non-skill, tool-application focus to using computers in class results in greater student engagement in their academic assignments.

Finally, our data suggest that certain approaches to using computer technology (i.e., broad use of different types of software, an emphasis on student writing and on exploiting Web-based sources of information) as well as a schoolwide emphasis on technology, particularly in the context of supporting instructional reform, are forces that help teachers realize significant changes in their pedagogy more generally, enabling them to put into practice a pedagogy that is more constructivist and more attuned with their teaching philosophy.

Thus, in a certain sense Cuban is correct—computers have *not* transformed the teaching practices of a majority of teachers, particularly teachers of secondary academic subjects. However, under the right conditions—where teachers are personally comfortable and at least moderately skilled in using computers themselves, where the school's daily class schedule permits allocating time for students to use computers as part of class assignments, where enough equipment is available and convenient to permit computer activities to flow seamlessly alongside other learning tasks, and where teachers' personal philosophies support a student-centered, constructivist pedagogy that incorporates collaborative projects defined partly by student interest—computers are clearly becoming a valuable and well-functioning instructional tool.

Moreover, where implemented in a responsible way, that tool is having an impact, not only on students' performance in class, but on their academic effort outside of class as well. In addition, many teachers, emphasizing the use of computers for student outcomes such as improved writing and research competencies, along with other teachers who are lucky enough to work in school environments where computer technology and instructional reform are cultural values, are being helped by technology to accomplish the goals of most current instructional reform efforts. They are creating classrooms where both they and their students are engaged in authentic efforts at increasing academic understanding rather than going through the more superficial traditional practice of schooling: surface coverage of a massive and externally mandated curriculum, even when anointed under a label of "standards-based reform.

Notes

Revision of a paper written for the January, 2000 School Technology Leadership Conference of the Council of Chief State School Officers, Washington, D.C. The author wishes to thank four anonymous reviewers for their critiques and suggestions.

1. Cuban recognizes that most teachers use computers professionally, for example, to prepare their lessons or to provide materials for student work, and that a small minority do have their students use computers regularly during class. However, he continues to maintain

that "deeply embedded factors...will continue to retard widespread classroom use of technology" (Cuban, forthcoming; undated manuscript p. 281).

2. Except where indicated by text or footnotes, statistical results are based solely on the weighted nationally representative sample of teachers and schools. The survey was fielded in the Spring of 1998, with most teacher questionnaires being returned in April or May of that year. For more details on the sampling and study methodology, see Becker, Ravitz, and Wong (1999), Appendix B. Online at http://www.crito.uci.edu/tlc/findings/computeruse/html/startpage.htm

3. The survey question read "On how many days since September has a typical student in this particular class used a computer while you were teaching their class?" The fourth and fifth choices in the list were "21-40 times (weekly)" and "41+ times (twice/week)." The class selected for questioning was the class selected by the teacher as the one where the teacher was "most satisfied with your teaching—where you accomplish your teaching goals most often." Subject-coding of teachers was based on the subject area in which the teacher taught for a majority of his or her classes.
4. Just a few computers in a classroom would not seem to make much sense. However, numbers like 5, 6, or 8 can be used quite efficiently for many kinds of classroom activity plans.
5. Although 18% of the survey respondents reported publishing on the World Wide Web, that estimate does seem inordinately high, given other data reported in the survey. Some frequency of misunderstanding of the survey question is probably responsible.
6. Means (2000) provides examples of how professional computer knowledge does not *always* translate into effective pedagogy with the same software.
7. Three sub-indices contributed equally to this index of computer knowledge (by standardizing the variance of each one). One measured the number of technical computing skills a teacher reported having (out of seven skills; for example, copying files from one disk to another, preparing a slide show using presentation software, using a Web search engine). The second measured the number of ways the teacher reported using computers for professional functions (out of eight, including corresponding with parents, exchanging computer files with other teachers, and making handouts for students). The third reported the teachers' self- assessments of the level of their experience with each of the two major computer platforms—Macintosh and Windows/DOS. The correlations among the three subindices ranged from r=.43 (professional uses with platform experience) to r=.60 (technical computing skills with platform experience).
8. Teachers were asked to estimate in how many lessons did they have students use each of ten types of software in their selected class. The "types" of software included "games for practicing skills," "simulations or other exploratory environments," "encyclopedias and other references on CD-ROM," "word processing," "software for

making presentations," "graphics-oriented printing (e.g., Print Shop)," "spreadsheets or database programs (creating files or adding data)," "Hyperstudio, Hypercard, or other multimedia authoring environment," "World Wide Web browser," and "electronic mail." The number of types of software used was the mean number reportedly used in at least three lessons during the year.

9. In this survey measurement context, correlations above .20 generally indicate differences worth paying attention to; correlations above .30 are "substantial"; and those above .40 would be considered very large. The table excludes teachers who don't use computers with their classes at all, but includes teachers from the special samples of schools in reform programs or with high-end technology presence in addition to the nationally representative sample.
10. These were five-point scales, with the extreme and moderate positions combined in the percentages provided in the text. The wording of the two choices were as follows: (A) "I mainly see my role as a facilitator. I try to provide opportunities and resources for my students to discover or construct concepts for themselves." (B) "That's all nice, but students really won't learn the subject unless you go over the material in a structured way. It's my job to explain, to show students how to do the work and to assign specific practice."
11. The validity of teachers' philosophical statements is somewhat problematic. Like reports of their actual practice, they may be subject to "social desirability" effects—i.e., wanting to give an answer perceived as desirable by others. However, prior to this national survey, we validated a set of statements about teaching philosophy through extensive interviews with 72 teachers in 24 schools in three parts of the U.S. The items selected (or modified) for this study were the items that correlated most strongly with the interviewers' judgments about the teachers' actual teaching philosophies. See Becker and Anderson (1998). Moreover, the primary use of the philosophy items in this study, however, is not to determine on an absolute scale how constructivist teachers are but whether those who are relatively more constructivist in philosophy than others respond more strongly to the option of using computers in their teaching.
12. Figure 7 uses a continuous measure of teaching philosophy, from most transmission-oriented to most constructivist, rather than the quartiles shown in Figure 5.
13. The CD-ROM item was described as CD-ROM *Reference* software but probably many teachers interpreted the survey question to include skills-games and exploratory software on CD-ROMs.
14. Chris Dede, in a recent paper (Dede, 2000), discusses how a wide range of software provides opportunities for students to engage in knowledge construction activities.
15. The analysis in this paragraph concerns teachers of secondary academic subjects and elementary teachers. It omits teachers of applied secondary subjects like computer education, business education, vocational education and fine arts.
16. Comparison based on probability plus purposive sample data. These two groups differ very little on gross measures; however, the

purposive sample is needed in these comparisons because the restriction to high constructivist philosophy teachers limits the number of teachers available by subject.

17. Teacher reports of frequent computer use by their students in class may be subject to upward bias due to the same social desirability factor noted in an earlier footnote with respect to reports of constructivist teaching philosophies. However, the data show *huge differences* in frequent student computer use between *all* teachers and teachers whose conditions are favorable (i.e., philosophy, computer knowledge, etc.). If social desirability was inflating teacher reports of frequent computer use substantially, we would not see such low percentages for all teachers combined with such high percentages for teachers with facilitating conditions. Moreover, random error in the measurement of the facilitating conditions (e.g., "adequate computer knowledge" is measured by a simple index of self-reports) tends to diminish the size of differences found. This would suggest that the true percentage of frequent users in the "all facilitating conditions present" category is even higher than reported.
18. The measure used in Figure 11 is the effect size between teachers who selected a given objective as primary versus those who did not. The effect size is the difference in the mean responses by the two groups of teachers divided by the standard deviation of teacher responses on the measure. The two items averaged in the measure (computer use at other times of the day while at school; and computer use at home) were each scored on a scale from 1 to 5 representing the poles of "none or few" students doing this on at least several occasions to "all students" doing this.
19. See Becker (2000) for evidence on the relationship between student SES and basic home computer access as well as the level of functionality of home computers owned by families of students of different economic and educational circumstances.
20. It is also possible that weak measurement of control variables—class SES was measured by school-level SES indicators and student ability was estimated by teachers, and home presence of computers was not measured directly—might leave us to ascribe some variation to teacher objectives that ought to be ascribed to student background factors. However, the SES and school level controls reduced the associations for objectives only to a small degree. Further discussion of the findings concerning student out-of-class computer use can be found in Becker (in press a).
21. The findings regarding changes in pedagogy over the previous three years are presented here only as preliminary. They will be the subject of a future TLC report.

References

Becker, H. J. (2000). Pedagogical Motivations for Student Computer Use That Lead to Student Engagement. *Educational Technology*. 40:5 (Sept.-Oct.), 5-17.

Becker, H. J. (in press). Who's Wired and Who's Not. *The Future of Children*. Center for the Future of Children, The David and Lucille Packard Foundation. 10:2 (Fall).

Becker, H. J. & Anderson, R. E. (1998). Validating Self-Report Measures of the 'Constructivism' of Teachers' Beliefs and Practices. Paper presented at the 1998 meetings of the American Educational Research Association.

Becker, H. J. & Ravitz, J. L. (1999). The Influence of Computer and Internet Use on Teachers' Pedagogical Practices and Perceptions. *Journal of Research on Computing in Education*. 31:4 (Summer), 356-384.

Becker, H. J., Ravitz, J. L., & Wong, Y-T (1999). *Teacher and Teacher-Directed Student Use of Computers and Software*. Teaching, Learning, and Computing: 1998 National Survey, Report 3. Center for Research on Information Technology and Organizations, University of California, Irvine (November).

Cuban, L. (1986). *Teachers and Machines: The Classroom Use of Technology Since 1920*. New York: Teachers College Press.

Cuban, L. (2000). So Much High-Tech Money Invested, So Little Use And Change In Practice: How Come? Paper prepared for the Council of Chief State School Officers' annual Technology Leadership Conference. Washington, D.C. (January).

Cuban, L. (forthcoming). *Oversold and Underused: Reforming Schools Through Technology, 1980-2000*. Manuscript in review.

Dede, C. (2000). Implications of Emerging Information Technologies for States' Education Policies. In *2000 State Educational Technology Conference Papers*. Washington, D.C.: State Leadership Center, Council of Chief State School Officers. (May).

Means, B. (2000). Accountability in Preparing Teachers to Use Technology. In *2000 State Educational Technology Conference Papers*. Washington, D.C.: State Leadership Center, Council of Chief State School Officers. (May).

Perkins, D. (1998). What is Understanding. In Wiske, Martha Stone, ed., *Teaching for Understanding: Linking Research with Practice*. San Francisco: Jossey-Bass.

Ravitz, J. L., Becker, H. J., & Wong, Y-T (2000). *Constructivist-Compatible Beliefs and Practices Among U.S. Teachers*. Teaching, Learning, and Computing: 1998 National Survey, Report 4. Center for Research on Information Technology and Organizations, University of California, Irvine (July).

About the Author

Henry Jay Becker
University of California, Irvine

Email: hjbecker@uci.edu

Henry Jay (Hank) Becker is a Professor of Education, University of California, Irvine. His research focuses on instructional and organizational reforms associated with the use of computer technologies. He is now analyzing data from Teaching, Learning, and Computing: 1998, the fourth in a series of national surveys of teachers and schools and their instructional use of computers, a series that stretches back to 1983. This survey focuses on teachers' pedagogical beliefs and practices and their relationship to teachers' use of technology. Besides these national surveys, he has conducted studies of the National School Network, a collaboration of curriculum reform projects at the leading edge of Internet use, and studies of Integrated Learning Systems. In the 1980s, he conducted a national field experiment on the effectiveness of typical practices of technology use in 50 pairs of classrooms across 13 states. Professor Becker holds a Ph.D. in Sociology from the Johns Hopkins University where he also worked as a Research Scientist at the Center for Social Organization of Schools between 1977 and 1992.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Greg Camilli
Rutgers University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher
Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

Este artículo ha sido consultado 223 veces desde el 16 de noviembre de 2000

Education Policy Analysis Archives

Volume 8 Number 52 **novembre 16, 2000** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Apoyo a la participación de padres en las escuelas primarias: Un estudio etnográfico sobre un grupo latinoamericano en Canadá

Judith K. Bernhard
Ryerson Polytechnic University

Marlinda Freire
Hospital for Sick Children, University of Toronto

Veronica Pacini-Ketchabaw
Ontario Institute for Studies in Education
University of Toronto

Resumen

Este artículo describe cómo un grupo de padres latinoamericanos aprenden a manejar en forma más efectiva sus experiencias con el sistema educacional canadiense. La etnicidad y variables como sexo y clase social son determinantes críticos en las interacciones sociales, en las cuales los recién llegados constituyen el grupo minoritario. Durante ocho meses, doce padres latinoamericanos compartieron una vez al mes sus experiencias con la escuela. Finalizado el estudio se descubre que estos padres no sólo aprendieron a colaborar con los maestros sino que también los enfrentaron y validaron ante ellos sus diferencias etnoculturales. Esta interacción condujo a ganancias inesperadas más allá de lo relativo al proceso educacional. A través de este examen los padres develaron lo que el sistema escolar considera como intervención familiar ideal, sin menoscabo del bagaje cultural del estudiante. Este estudio puede ser un modelo de adaptación para grupos de recién llegados que intentan integrarse al sistema educativo.

Abstract

This article describes how a group of Latin American parents became more effective in their dealings with their children's schools, a mainstream Canadian institution. Ethnicity, along with race, gender, and social class, is a critical determinant in of the interactions between schools and any group of newcomers to a society, particularly when those newcomers are an ethnic minority. Over an eight-month period, twelve Latin American parents met monthly to discuss aspects of their children's experience with the Canadian educational system. These parents learned to collaborate with teachers and expressed their needs, but also affirmed their ethno-cultural differences. The positive feedback on their activities led to unforeseen gains, not just in relation to education and the schools. This exploratory study focuses on how the experience helped the parents to better comprehend what is expected of them in the support of their children's schooling while retaining their own cultural assets. This study may serve as a possible model of adaptation for newcomer groups in their efforts to integrate in the school system.

Introducción

Gobiernos, ministerios y escuelas en Canadá, como en otros países del mundo, enfatizan la importancia de la colaboración entre la

familia y la escuela. Nuestro trabajo con un grupo de padres latinoamericanos residentes en Canadá, efectuado entre octubre de 1997 y junio de 1998 demuestra que, desafortunadamente, en este país existen dificultades para la implementación de la legislación pertinente. En este artículo, basado en un estudio etnográfico de un grupo latinoamericano en Toronto, exploramos las experiencias de las familias latinoamericanas y las dinámicas observadas dentro del contexto educacional canadiense. Documentamos además, desde su propia perspectiva, la naturaleza de las desventajas institucionales a las que estos padres se enfrentan. Asimismo, discutimos cómo el proceso de grupo facilitó la participación eficaz de los padres en las experiencias educacionales de sus hijos. Examinamos las dificultades de los padres en su adaptación al sistema educacional como un aspecto más de las dificultades globales que los inmigrantes experimentan en sus intentos de integración a la sociedad mayoritaria. (Freire, 1993). Esta investigación y la información obtenida en este trabajo permitieron encarar las inquietudes siguientes y demostrar cómo los padres, trabajando colectivamente, resolvieron estas inquietudes de una manera más efectiva:

1. ¿Cómo las familias latinoamericanas perciben la práctica escolar?
2. ¿De qué manera los padres latinoamericanos perciben el sistema escolar y cuáles son los roles sociales que les han sido asignados?
3. ¿De qué manera los padres latinoamericanos enfrentan un proceso institucional que los pone en desventaja por ser emigrantes y tener un manejo limitado del idioma dominante del país que los recibe?

Además de intentar aclarar la dinámica de interacción minoría-mayoría dentro de un encuadre escolar, este trabajo describe el proceso de grupo y la transformación que ulteriormente permite a los padres enfrentar, más preparados, el futuro académico de sus hijos, y a ellos ejercer más eficazmente su potencial de poder sociopolítico y económico en la nueva sociedad, basándose en sus interacciones con el sistema escolar.

De acuerdo con la última estadística disponible, la proporción de población que habla español como primera lengua en Canadá ha aumentado en más del doble en la última década. De 70,000 en 1981 a 187,000 en 1996 (Statistics Canada, 1998). De acuerdo con un estudio reciente, los inmigrantes latinoamericanos integran uno de los dos grupos étnicos con más probabilidades de vivir en la pobreza en Canadá (Halli & Kazemipur, 1997). Aunque el latinoamericano es uno de los grupos con mayor crecimiento en este país, al estar diseminado a través de numerosas escuelas, pareciera representar una pequeña minoría.

Marco teórico

Nuestro marco teórico es social y ante todo estructural. Las desventajas sistémicas son consideradas múltiples, basadas en factores

como clase social, raza, género y tipo de inmigración (Apple, 1992; Ng 1993). El poder es visto como esparcido en instituciones oficiales y en las acciones cotidianas de la sociedad. (Bourdieu, 1986; Cannella, 1997; Looker, 1994). Ng (1987) propone que clase social, raza, género y condiciones de inmigración están constituidas por relaciones y prácticas sociales institucionalizadas. En particular, el origen étnico no está considerado como una característica inherente al grupo, sino ante todo, como una continua imputación e interpretación por aquellos que se constituyen a sí mismos como grupo dominante. Estos mismos argumentos se aplican para la categoría de "razas" (Dei, 1993a, 1993b; Miles,1989).

Usamos el concepto de "capital cultural" de Bourdieu (1986) al referirnos al modo de ser, conocimientos, habilidades, disposiciones y capacidades que establecen a una persona en un contexto y estrato social determinados dentro de un ambiente de relaciones sociales. A través de esas relaciones, la misma persona es vista de acuerdo con su posición social previamente estructurada. En nuestro trabajo anterior con familias latinoamericanas encontramos que frecuentemente hay una disparidad de valoración entre el capital cultural que las familias traen a la nueva situación (país, sociedad, escuela, etc.) y aquel (implícitamente) requerido por la escuela. La valoración de los profesores sobre el "apoyo de los padres" al desarrollo académico de sus hijos resultó estar altamente determinada por modelos específicos de colaboración padre- hijo en la cultura mayoritaria. Sin embargo, el quehacer de los padres, incluyendo los consejos y la orientación que le dan a los niños, está basado en su propia visión cultural y en lo que a ellos les corresponde hacer en dichas situaciones.

Aunque podríamos llamar a esto una simple situación de desequilibrio en las habilidades y la manera de ver un mismo proceso por dos agentes diferentes, tal "desequilibrio" representa, en nuestra opinión, un fenómeno mucho más complejo, ya que está relacionado con diferencias de poder a nivel sociopolítico y económico.

Lareau (1989), en su trabajo con comunidades en Estados Unidos, encontró que los padres de clase social alta sabían utilizar el sistema y conseguían que los maestros ajustaran el programa escolar de acuerdo con las necesidades de sus niños. En cambio, las familias de clase socio económica baja no logran tales ajustes a pesar de que las necesidades de estos niños son mayores. En vista del bajo rendimiento escolar de los niños latinoamericanos en Canadá (Bernhard & Freire, 1996; Brown, 1994, Drever, 1996), consideramos que es importante que los padres entiendan cómo funciona el sistema educacional canadiense y la habilidades que se requieren para ser más eficaces en sus interacciones con el mismo.

El segundo marco de referencia en este trabajo es la teoría ecológico-cultural. Spindler (1990) ha considerado la escolaridad como un "proceso cultural obligatorio". De acuerdo con esta posición teórica, los profesores, reflejando una posición etnocéntrica, transmiten los valores dominantes e inadvertidamente debilitan la identidad cultural de los estudiantes de grupos minoritarios. Al sentir su propia identidad en peligro, los estudiantes responden con

conductas defensivas que contribuyen a perpetuar su marginación (Trueba, 1993). Las teorías de deficiencia cultural invierten la relación entre causa y efecto (Barrera, 1997) y por lo tanto refuerzan posiciones de victimización de grupos minoritarios.

Nuestro tercer marco teórico es la teoría anti-racista (e.g., Dei, 1993a). Este marco referencial es fundamental en el análisis de nuestro trabajo. Es imposible cambiar el rendimiento académico de los niños sin antes entender el contexto en el cual ellos viven sus experiencias, incluyendo las determinadas por el color de la piel. La primera serie de variables contextuales que identifican subordinación son raza, clase social y género. El problema de la interacción entre raza y clase social es sumamente complejo y ha sido explorado por numerosos autores (Bernhard, Freire & Pacini-Ketchabaw, in press; Dei, 1993a; Ng, 1993). En el contexto del presente estudio, vemos la subordinación como la continuidad de un proceso histórico, complejo e impersonal con implicaciones mayores de factores como raza, etnicidad, clase social y género.

Algunas de las dificultades que los padres encuentran se deben, en parte, a la exclusión social por parte del grupo dominante y al hecho de que no existe una red informal de comunicación a través de la cual los padres puedan compartir información importante y adecuada sobre la práctica escolar en el nuevo país. Los padres asumen que el sistema educacional canadiense funciona en forma similar al de su país de origen. Más aún, los padres se sienten intimidados por las autoridades escolares basados en sus experiencias previas con instituciones oficiales en su país de origen. Estas dificultades son agravadas por las barreras idiomáticas.

Es posible, entonces, tratar de identificar, analizar y ayudar a vencer las dificultades escolares de los niños, formando grupos de padres con un bagaje cultural y lingüístico similar y a través de facilitadores de grupo que funcionen como mediadores culturales. Es fundamental que estos mediadores compartan la cultura y el idioma nativo de las familias, entiendan cómo funciona el nuevo sistema escolar y conozcan las estrategias que permiten una colaboración efectiva de los padres en su tarea de apoyar a los niños en la escuela. Es de primordial importancia que los padres comprendan que adquieren una gran ventaja al participar activamente en el quehacer escolar. Posiblemente esto último fue una meta mayor implícita en la elaboración de este proyecto.

Método

Durante un período de ocho meses, un grupo de doce padres latinoamericanos se reunieron una vez por mes. Aunque hubieron padres que asistieron a algunas de las reuniones, fueron las madres las que participaron regularmente y es por eso que en el resto de este artículo usamos solamente el término "madre". El grupo estaba compuesto por cuatro madres chilenas, una argentina, tres salvadoreñas, dos uruguayas, una mexicana y una nicaragüense. Nueve de las madres eran de clase trabajadora y solamente tres de clase

media. Las edades fluctuaban entre los 33 y los 42 años. Todas tenían algún nivel de educación secundaria y tres de ellas, educación universitaria. Todas estas madres constituían familias intactas (madre y padre). En este grupo, las madres sudamericanas tenían un promedio de 15 años de residencia en Canadá, mientras que las centroamericanas eran de inmigración más reciente, con un promedio de 8 años, y una de ellas había residido menos de un año en el país. Todas tenían, por lo menos, un niño en edad escolar primaria. Las edades de ellos fluctuaban entre los 18 meses y los 16 años. El trabajo se enfocó en la experiencia escolar de los niños entre los 4 y 14 años, con el supuesto adicional de que cualquier aprendizaje de las madres sobre el sistema escolar las ayudaría con los niños menores de 4 y mayores de 14 años. Estos niños asistían a escuelas con un perfil étnico muy diverso y donde la población latinoamericana no era, en la mayoría de los casos, más del 2 o 3 por ciento.

Solamente dos de las madres tenían hijos en una misma escuela en que la población latina constituía el 15 por ciento del estudiantado. En esta escuela en particular había un profesor latinoamericano encargado del programa de "Herencia Lingüística" (español). Las madres que asistieron fueron contactadas a través de asociaciones comunitarias e invitadas a participar en esta serie de reuniones de grupo. Se les informó que el proyecto constituía un trabajo de investigación y cuáles serían los posibles beneficios para los padres participantes y para la comunidad latinoamericana en general. Las reuniones fueron de dos horas de duración, conducidas en español y se ofreció cuidado de niños y refrigerios.

Los dos investigadores principales asistieron a todas las reuniones. La facilitadora, entre otras funciones, iniciaba cada sesión, hacía un resumen de los temas discutidos en las sesiones previas, invitaba a los participantes a comentar sobre este resumen y finalmente, iniciaba la discusión de la nueva sesión con una pregunta neutra sobre un tema abierto pero pertinente. La sesión podía dedicarse a temas previamente discutidos que necesitaban más elaboración o podían ser temáticas nuevas que reflejaban otras áreas de preocupación de las madres con respecto a la educación de sus niños y el sistema educacional. De este modo las participantes fueron capaces de reflexionar de manera continua durante el periodo que duró el proyecto, sobre sus contribuciones y cuestionamientos y al mismo tiempo, de incorporar nuevos elementos que les permitían, no solamente entender el sistema escolar, sino reafirmarse frente al mismo. En forma rotativa, un co-facilitador, elegido entre las madres del grupo, estaba encargado de dirigir la discusión. Inicialmente las madres se resistían a participar como co-facilitadoras pero pronto se notó un cambio de actitud con un aumento en la confianza para ejercitar dicho rol. Cada sesión fue grabada, transcrita y codificada.

El análisis preliminar de este trabajo fue presentado al grupo para su verificación. En la segunda etapa, los temas identificados como centrales en el trabajo con los padres fueron discutidos informalmente con dos expertos en el área para corroborar la validez y la relevancia de los mismos. La transcripción de las grabaciones permitió hacer un

informe sobre las historias y versiones de las madres en la descripción de las experiencias de sus hijos y de las propias con los maestros y autoridades del sistema escolar. Estas madres constituyen una muestra de conveniencia, por lo tanto cualquier intento de generalizar nuestras conclusiones debería hacerse (y lo hacemos) con toda cautela. El propósito del trabajo fue documentar cómo este grupo particular de madres latinoamericanas consiguió entender mejor el sistema escolar canadiense, lo que los profesores y la escuela valoran y esperan del niño y de la familia, expresar con más confianza las necesidades de sus niños, y la forma más efectiva de intervenir a nivel escolar en favor de ellos.

Resultados

A. Descripción del grupo y las experiencias de las madres con el sistema educacional

Al comienzo del estudio las madres estaban aisladas y cada una vivía sus experiencias en forma individual. El grupo sirvió como medio para conectarlas a través de los problemas comunes en su contacto con el nuevo sistema educacional. Estas madres aprendieron a comunicarse más efectivamente con los maestros y descubrieron las expectativas tácitas que los maestros esperan encontrar en los "padres eficaces". En el proceso de reunirse para discutir la educación de sus hijos, estas madres comenzaron a reconocer los aspectos de la situación educacional de sus niños que reflejaban el funcionamiento del sistema. Alcanzando más allá de sus experiencias individuales, comenzaron a apoyarse mutuamente. A decir lo que pensaban, (inicialmente en el grupo y más tarde en las escuelas) y a exigir que algunos de los programas educacionales se adaptaran a las necesidades de sus hijos. Por ejemplo, la señora Morales acota: "le dije al maestro que quería que mi hijo fuera al programa de francés, respondió que no porque no sabía suficiente inglés. Pero le dije al maestro, ¿cuál es la diferencia si inglés tampoco es su primera lengua?. Entonces el maestro arregló una reunión especial con el director, un representante del Consejo de Educación, una sicóloga y una trabajadora social. Cuando llegué a la reunión tenía miedo pero de todas formas les dije lo que quería y ellos lo aceptaron". Las madres en forma paralela experimentaron un proceso de organización de un grupo de autoayuda que reforzó su autoestima y les ayudó a vislumbrar la posibilidad real de comenzar a entender el sistema escolar con su agenda oficial y las estructuras que permiten mantener el status quo de las clases privilegiadas.

Como grupo empezaron a entender la situación de riesgo implícita en su posición de nuevos inmigrantes o refugiados de países en desarrollo. Las madres comenzaron a experimentar un proceso de aprendizaje que afirmó la base para un proceso de concientización como el descrito por Paulo Freire (1972). Además, se sintieron capaces de iniciar cambios y se fueron dando cuenta de su potencial y de su efectividad ante situaciones que antes les parecían impenetrables. A

partir de entonces, empezaron a difundir este conocimiento con otras madres latinoamericanas que no asistían al grupo pero que compartían "desconocimientos" y preocupaciones parecidas. Por ejemplo, la señora Díaz dijo: "Mi comadre no puede asistir a estas reuniones pero yo le hago saber todo lo que aquí discutimos. Entre las dos tratamos de entender más aún. Ella quiere pedirle a su hermana que también se junte con nosotras porque sus hijos son chiquitos pero ya tienen problemas en la escuela. Yo no sé cómo podemos tratar de defendernos mejor de las cosas que les pasan a nuestros niños en la escuela sin saber lo que estamos aprendiendo en este grupo".

Durante la participación de las madres en el grupo vimos varios casos de madres que cuando entraban a las escuelas se sentían incómodas, ignoradas y consideraban que asistir a la escuela no era efectivo para las necesidades de ellas y de sus hijos. Por ejemplo, la señora Blanco dijo al comienzo del proyecto: "Cuando voy a las reuniones sobre calificaciones, salgo con muchas dudas, me voy de la reunión tal como llegué, sin entender nada. Los profesores asumen que entendí todo. Uno siente que en lo que dicen los maestros no hay nada que discutir, especialmente si ellos consideran que tu hijo no tiene problemas serios. Ellos te entregan las notas, asumen que tu entendiste y no dan oportunidad de preguntar nada. Entonces me voy como si hubiera entendido lo que dijeron pero en realidad no entendí nada". La señora Castillo describió su experiencia con el siguiente comentario: "En la escuela de mi hijo, los padres no van a la sala de clases. La maestra recibe a los niños en la mañana y los lleva a la sala de clases. Yo no tengo oportunidad de hablar con ella. Me gustaría saber por qué mi niño está en una clase donde ella enseña dos cursos, pero no sé si preguntarlo. No me atrevo a decir: quiero que lo pongan en una clase regular solamente con niños del mismo curso. La maestra sabe por qué lo puso en esa clase. No quiero tener problemas con ella, pero me preocupa bastante". A medida que el grupo progresaba, estas mismas madres comenzaron a interactuar con los maestros, a reafirmarse a sí mismas y a saber que sus opiniones tenían valor y podían ser escuchadas. La señora Rojas dijo: "Esta es la primera vez que participo en un grupo con esta dinámica. Tenemos la oportunidad de tratar los temas en profundidad. Como no somos muchos tenemos la oportunidad de hablar, lo que me ha ayudado a sentirme parte del grupo. Hemos compartido temas que son nuestros".

En el proceso de reunirse, las madres empezaron a valorarse a sí mismas y lo que ellas, u otros, pudieron haber considerado deficiencias, comenzaron a desaparecer y a transformarse en fortalezas y capital instrumental. Seis semanas más tarde, en algún momento del proceso, la señora Blanco dijo: "Este grupo me ha servido para ayudar a mis amigos que tienen problemas con sus niños en la escuela. Yo les digo: mira, tú puedes venir aquí, o puedes hacer esto o aquello. Les da fuerza. Esto me ha ayudado a compartir con otros padres ya que hay tanta gente que no sabe como enfrentar este tipo de situaciones". La señora López orgullosamente anunció en una de las reuniones: "He decidido que voy a empezar a ir a la escuela y molestar a los profesores hasta que me escuchen. Estoy en eso y si todos los padres

hacemos lo mismo terminarán escuchándonos. El otro día la profesora me pidió ir a la escuela para ayudar a los niños con la lectura". también fue interesante observar que las madres describieron cómo el proceso que se daba en el grupo les ayudó a generalizar en otras tareas sobre sus propios derechos y los de sus hijos. La señora González dijo: "El grupo me ayudó a entender a mi hija. Me ayudó a ver que tenía ciertos derechos y que eran legítimos. Me ayudó a entender que la persona que estaba enseñando a mi hija no era un ser superior que yo no podía alcanzar ni hablarle. El maestro y yo tenemos la misma tarea, la educación de mi hija".

B. Subordinación y reconocimiento de las desventajas

Otros factores que las madres presentaron como problemáticos fueron las diferencias que percibían en las áreas de género y raza. En cuanto a las diferencias de género las madres habían comprobado que la mejor estrategia para ser escuchadas por las autoridades escolares era ir acompañadas por el esposo u otro varón allegado a la familia. Algunas madres describieron cómo, a pesar de ser ellas las que verbalizaban el problema, el personal escolar normalmente se dirigía y hacía contacto visual con el varón. La señora Godoy dijo: "Han habido circunstancias en que he ido a la escuela con mi esposo para hablar con los profesores y he notado una mejor disposición de parte de ellos porque él estaba presente. Una vez le preguntaron su opinión y le dieron opciones, lo que no pasa si voy sola. Cuando voy a las reuniones con mi esposo los profesores tienden a dirigirse solo a él y no a los dos".

En cuanto a raza, en la mayoría de los casos, las participantes no elaboraron sus dificultades en términos de "problemas raciales". Las madres no se veían a sí mismas correspondiendo a un encuadre racial negro/caucásico u otro. Las madres pertenecían a diferentes grupos raciales y poseían distintos tonos de piel. Sin embargo, muchas de ellas se referían continuamente a la población latinoamericana como una raza y no como un grupo étnico. Algunas madres de piel más oscura mencionaron como factor adicional en las dificultades que los niños o ellas habían tenido, el racismo que habían experimentado en las instituciones escolares y que ellas atribuían, particularmente, al color de su piel. La señora Godoy dijo: "Pienso que nuestra posición en la sociedad es clara, tenemos desventajas debido al color de nuestra piel y a nuestro lenguaje y que nuestros niños tienen las mismas desventajas en la escuela. Un día con mi hija nos encontramos con dos nativos canadienses y ella les preguntó si eran peruanos ya que se veían como nosotros.... Muchas veces nos sucede que la gente piensa que somos nativos canadienses cuando en realidad somos latinos. Aquí sabemos lo que maltratan a los nativos canadienses".

La raza, de la manera presentada por las madres, es considerada en este trabajo como una construcción social, más que biológica o antropológica. Para estas familias, la cuestión del color de la piel no es vista como un indicador primario de la raza. De la misma manera, el ancestro latinoamericano y el bagaje de experiencias son identificados

por el grupo como raza. Aunque en los Estados Unidos el discurso político reconoce los distintos grupos étnicos, particularmente negros y blancos, nuestros resultados demuestran que en Canadá la situación de los latinoamericanos pareciera ser conceptualizada en forma diferente. Las madres no se consideraban a sí mismas como pertenecientes a un determinado grupo racial, pero a medida que el trabajo de grupo progresó, ellas comenzaron a ver sus semejanzas y sus ancestros comunes, y lo empezaron a definir como "raza".

Los latinoamericanos en Canadá recién comienzan a tomar conciencia de grupo étnico. Basados en el informe de las madres, no existen evidencias de que las escuelas canadienses hayan reconocido a las familias latinoamericanas como un grupo étnico particular, pero las madres describen experiencias discriminatorias en relación con diferentes aspectos del quehacer escolar que relacionan con raza. Por ejemplo la señora Pérez dice: "a veces la discriminación es muy sutil, especialmente con los niños que están empezando a hablar inglés y los ponen en niveles básicos aún cuando ellos son capaces para un nivel más avanzado. Como padres, nosotros sabemos que nuestros hijos son capaces pero los maestros insisten en que deben ir a los niveles básicos. A veces, solamente porque nos vemos y hablamos diferente. Entonces nosotros tenemos que luchar por sus derechos". La señora Mendoza agregó: "La escuela de mi hijo tiene principalmente una población portuguesa, italiana e hispana, sin embargo la mayoría de los maestros no tienen ancestro latino, entonces, se puede ver favoritismo con los niños de piel mas clara". Esto va en contraste con la situación que se vive en los Estados Unidos donde la población de color, incluyendo los latinoamericanos, está claramente definida como un grupo étnico reconocido. Como resultado, ciertos grupos étnicos sufren una invalidación del sentido de sí mismos que los puede llevar a un racismo internalizado como un factor adicional en la perpetuación de un desequilibrio en el poder socio-político y económico ya establecido por el sistema dominante.

Cuando el grupo dominante determina cuál es el capital cultural predominante y por lo tanto, el que se valora preferencialmente, las diferencias culturales se vuelven "deficiencias" de acuerdo con las medidas estándar de lo que se considera "normal y valioso" en ese medio. Los profesores, inadvertidamente, están incapacitados para evaluar el conocimiento y el capital cultural que las familias de clase trabajadora y de grupos minoritarios poseen en el momento de tratar de insertarse en el medio dominante. Dentro de este contexto y de acuerdo con el criterio institucional, no se reconoce la voluntad de colaboración y los esfuerzos de las familias en apoyar el desarrollo académico de los niños en la forma que ellos culturalmente lo entienden.

Los profesores, sin suficientes recursos personales ni instrumentales para tener contacto individualizado con sus estudiantes o con las familias, corren el peligro de convertirse en simples agentes transmisores del punto de vista etnocéntrico dominante. En estas circunstancias, las familias se enfrentan con obstáculos estructurales prácticamente insalvables que no les permiten remover las barreras

discriminatorias para lograr una verdadera colaboración y un trabajo en conjunto con las escuelas. Las limitaciones lingüísticas del uso del segundo idioma, el desconocimiento de los programas escolares y de la operación y funcionamiento del nuevo sistema escolar, son tres aspectos básicos que tienden a magnificar y perpetuar los obstáculos estructurales.

A través del proceso de crecimiento que las madres experimentaron, las ganancias parecieron mayores que las esperadas inicialmente para el proyecto. Por ejemplo la señora Rojas dijo: "Estaba preocupada por el rendimiento de mi hijo en matemáticas. Sé que necesita apoyo, entonces fui a la escuela y la profesora me dijo que no necesitaba clases especiales porque alguien en su curso lo estaba ayudando. después de eso me dejé de preocupar, pero ahora que terminó la escuela me mandaron un libro entero para que trabaje en el verano ya que estaba atrasado en matemáticas. Ella debería haberle dado tareas diarias y no esperar hasta el final. No me parece correcto. Si hubiera sabido habría insistido pero la profesora me dijo que no me preocupara. La próxima vez no me voy a quedar tan tranquila, voy a insistir hasta que yo me convenza que mi hijo lo está haciendo bien". De la señora Castillo escuchamos lo siguiente: "Nos juntamos con Esther y fuimos a una charla política en la universidad... Era la primera vez que participábamos en algo así. No fue fácil pero la participación en este grupo nos dio confianza para enfrentar nuevos niveles de actividad". La señora Morales acotó: "Una vez la secretaria de la escuela me llamó para que vaya a buscar a mi hijo porque se había pintado la cara y estaba distrayendo a los otros niños de la clase. Dijo que mi hijo estaba fuera de control y que lo tenía que ir a buscar inmediatamente. Me dijo que ellos no sabían que otra cosa hacer. Cuando la escuché sentí como si alguien me hubiera vaciado un balde de agua fría en la cabeza. Me sentí choqueada y avergonzada. No pude contestar nada y colgué. En otro momento habría salido corriendo a hacer lo que la secretaria me pedía. Esta vez fue distinto, me senté en la cama, estaba temblando, respiré profundo, llamé a la secretaria de vuelta y le dije no iría a recoger a mi hijo, que no pensaba que lo que el niño estaba haciendo fuera tan terrible".

Discusión general

Durante nuestras primeras reuniones, las madres estaban frustradas por el bajo rendimiento de sus hijos y porque además sentían que sus inquietudes y cuestionamientos no eran bien recibidos. Ellas, hasta el momento de empezar a participar en el grupo, habían reaccionado pasivamente como respuesta a su percepción de los maestros como figuras de expertos, autoritarias e inaccesibles. Por su desconocimiento del modus operandis del nuevo sistema escolar y por la falta de éxito en las iniciativas que ellas habían tomado frente a las problemáticas presentadas.

Como resultado del trabajo de grupo, las madres comenzaron a tomar iniciativas efectivas y a hacerse oír a medida que cumplían exitosamente con varios objetivos relacionados con las escuelas. Esta

retroalimentación positiva fue fundamental, ya que la mayoría manifestaron una necesidad de participar en forma más activa en las organizaciones latinoamericanas con representación oficial frente al sistema escolar, lo cual podría, a más largo plazo, ayudar a operar los cambios estructurales que sean necesarios. Consiguieron entender mejor el rol de los consejos escolares y se sintieron más preparadas para participar en forma activa y eficaz.

Los ejemplos presentados en este artículo demuestran la complejidad subyacente de las interacciones entre las familias latinoamericanas y las escuelas canadienses. Por ejemplo, si por una parte, existía una "falta de participación de los padres" de acuerdo al personal de la escuela, esto no debería ser entendido, automáticamente, como una falta de interés o motivación por parte de ellos en el desarrollo académico de sus niños. Más bien deberíamos entender estas acciones como el resultado de la percepción del nuevo sistema escolar por parte de los padres y de la interpretación que los maestros dan a su rol en la educación de todos sus estudiantes, incluyendo los hijos de emigrantes. Es nuestra opinión que éste, como parte de un proceso global de integración de los recién llegados, es un proceso bidireccional en el cual la mayor responsabilidad le corresponde al sistema ya establecido (sistema escolar).

Si bien el presente proyecto fue dirigido al nivel de educación primaria, la información obtenida indica que, según la perspectiva de los padres, ciertos modelos de comportamiento entre los padres y el sistema escolar han sido establecidos antes que el niño comience su educación primaria. La marginación del capital cultural de los padres es, realmente, un fenómeno que tiene sus orígenes en interacciones anteriores al comienzo de la educación escolar formal. En un estudio con pre- escolares latinoamericanos que asistían a guarderías asociadas a programas de aprendizaje de inglés para sus padres, se encontró, de parte de las educadoras, una devaluación total del uso del idioma español y de las habilidades de estos padres para criar a sus hijos (Benhart y Freire, 1996). Cualquier propuesta genuina de cambio tendría que comenzar al nivel de las guarderías infantiles, ya que éste es el primer contacto de los padres con el sistema educativo (Lee & Seiderman, 1998). Si esto se hace en las primeras etapas del desarrollo del niño y de su contacto con los sistemas oficiales, los grupos de padres podrían desarrollar un sistema de apoyo y conocimiento que les permita mantenerse unidos y mejor capacitados para aportar más efectivamente en la formación académica de sus hijos.

Los padres pueden sólo hacer el trabajo que a ellos corresponde. El aspecto de comprensión y reconocimiento de otros grupos étnicos y la práctica de los educadores necesita ser dirigido hacia otras alternativas de trabajo cuyas responsabilidades le corresponden al sistema educacional.

¿Cómo se explica la falta de interacción bidireccional en la cual padres y maestros se reúnan y escuchen las preocupaciones de unos y otros? El sistema educacional asume un modelo clásico de interacción: Los maestros convocan a reuniones y conferencias entre padres y maestros, con una agenda unidireccional. Este proceso transforma a los

padres en agentes impotentes (Dehli, 1994; MacLure & Walker, 1998). Generalmente las escuelas esperan que los padres acudan a ellas y reciban lo que se les ofrece, en un formato técnico, generalmente incomprensible para los padres.

Alternativamente, se espera que los padres hagan una presentación elaborada, respaldada por una evaluación profesional de los servicios adicionales requeridos por sus hijos en la escuela. Esto solamente es posible para los padres de clase media/alta familiarizados con el sistema y que cuentan con los medios económicos que les permiten documentar las dificultades que sus niños experimentan, en forma privada si ello es necesario.

Los padres no podrán efectuar cambios significativos hasta que las estructuras se conviertan en una interacción de dos vías donde ambos lados puedan hablar y escucharse, y donde los padres tengan alguna certeza de que sus preocupaciones y las sugerencias de como enfrentar las dificultades serán implementadas o por lo menos exploradas. Esta situación anómala se debe a obstáculos institucionales, más bien que a acciones individuales mal intencionadas.

La escasez o ineficacia de mecanismos y/o recursos, para involucrar a los padres e invitar sus aportes, van probablemente a derivar en la clase de dificultades que hemos detallado, en una población compuesta por diversos grupos étnicos. Un componente clave para establecer un modelo de colaboración, más allá de lo que los padres mismos pudieran hacer, requiere que los maestros desarrollen una mayor comprensión de la cultura y de las preocupaciones de las familias y que obtengan experiencia práctica trabajando directamente con los padres (Corson, Bernhard & González-Mena, In press; Moll, Amanti, Neff & González, 1992). también es fundamental que los educadores entiendan el valor del idioma nativo, no solamente como puente en la adquisición de un segundo idioma, sino como la base del desarrollo global del niño (Freire M., Benhard J., 1997).

En nuestra opinión, los maestros y la escuela pueden jugar un rol de importancia fundamental en crear o facilitar nuevas disposiciones conducentes a una experiencia escolar positiva para las familias inmigrantes y sus niños en el intento de integrarse a la nueva sociedad. Como resultado de este trabajo, presentamos aquí una experiencia y una forma alternativa de participación de los padres dentro del nuevo sistema escolar. Este tipo de participación reconoce el contexto histórico-cultural particular en el cual las familias viven su realidad actual y se basa en la comprensión básica de los padres del proceso institucional dominante, en un área fundamental e ineludible: la educación. En resumen, este proyecto trata de facilitar un proceso a través del cual los padres se transformen en agentes activos y conscientes de una pedagogía liberadora que pudiera llevar a transformaciones substanciales, no solamente en estos padres y sus hijos, sino también en las generaciones futuras.

Referencias

Apple, M. (1992). The text and cultural politics. *Educational Researcher, 21*(7), 4-11.

Barrera, M. (1997). A theory of racial inequality. In A. Darder, R. Torres and H. Gutiérrez (Eds.), *Latinos and education: A critical reader* (pp. 3-44). New York: Routledge.

Bernhard, J.K., Freire, M. & Pacini-Ketchabaw, V. (In press). A Latin American parentsí group participates in their childrenís schooling: Parent involvement reconsidered. *Canadian Ethnic Studies Journal.*

Bernhard, J.K., & Freire, M. (1996). Latino refugee children in childcare: A study of parents and caregivers. *Canadian Journal of Research in Early Childhood Education*, 5(1), 59-71.

Bourdieu, P. (1986). The forms of capital. In J. C. Richardson (Ed.), *Handbook of theory and research in the sociology of education* (pp. 241-257). New York: Greenwood Press.

Brown, R. S. (1994). A follow-up the Grade 9 Cohort of 1987 secondary students survey participants. Toronto, Canada: Toronto Board of Education.

Cannella, G. S. (1997). *Deconstructing early childhood education: Social justice and revolution.* New York: Lang.

Corson, D. (1998). *Changing education for diversity.* Philadelphia, PA: Open University Press.

Corson, P., Bernhard, J. K., & Gonzalez-Mena, J. (In press). Culturally situated explorations of child development: A home visit project for prospective early childhood teachers. *Journal of Early ChildhoodTeacher Education.*

Dehli, K. (1994). Parent activism and school reform in Toronto. Unpulblished Doc. Department of Sociology in Education, OISE.

Dehli, K. (1987). 'Ethnicí 'parentí 'communityí: The proper channeling of education politics. In J. Young (Ed.), *Breaking the mosaic: Ethnic identities in Canadian schooling* (pp. 224-239). Toronto, ON: Garamond Press.

Dei, G. (1993a). The challenges of anti-racist education in Canada. *Canadian Ethnic Studies, 25* (2), 36-51.

Dei, G. (1993b). Narrative discourses of black/African- Canadian parents and the Canadian public school system. *Canadian Ethnic Studies, 25* (3), 45-65.

Drever, N. (1996). Frustrated learners: An ethnographic study of Spanish-speaking students at an inner-city high school in Toronto. Unpublished master's thesis, Toronto, ON: University of Toronto.

Freire, M. (1993). Mental health, culture, children and youth. In R. Masi, L. Mensah, and K.A. McLeod (Eds.), *Health and cultures: Exploring the relationships,* Vol. II,Toronto.

Freire, M.; Benhard, J.K. (1997). Caring for and Teaching Children Who Speak Other Languages. *Include Me Too. Human Diversity in Early Childhood.*(Chapter 9, pp 160).

Freire, P. (1972). *Pedagogy of the oppressed.* New York, NY: Continium.

Kazemipur, A. Halli, S. S. (1997). Plight of immigrants: The spatial concentration of poverty in Canada. *Canadian Journal of Regional Science, 20* (1,2), 11-28.

Lareau, A. (1989). *Home advantage: Social class and parental intervention in elementary education.* New York: The Falmer Press.

Lee, L., & Seiderman, S. (1998). Families matter: The Parent Services Project. Boston, MA: Harvard Family Research Project.

Looker, D. E. (1994). Active capital: The impact of parents on youthsí educational performance and plans. In L. Erwin & D> MacLennan (Eds.), *Sociology of education in Canada* (pp. 164-187). Toronto, ON: Copp Clark Longman.

MacLure, M. & Walker, B. M. (1998). Disenchanted evenings: The social organization of talk in parent-teacher consultations in UK secondary schools. Paper presented at the American Educational Research Association meeting, San Diego: April.

Miles, R. (1989). *Racism.* London: Routledge.

Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching. Using a qualitative approach to connect homes and classrooms. *Theory into Practice, 31,* 132-141.

Ng, R. (1987). Ethnicity, schooling and the social division of labour: A response to Jackson. In J. Young (Ed.), *Breaking the mosaic: Ethnic identities in Canadian schooling,* (pp. 183-189). Toronto: Garamond Press.

Ng, R. (1993). Racism, sexism, and nation building in Canada. In C. McCarthy and W. Crichlow (Eds.), *Race, identity and representation in education* (pp. 50-59). New York: Routhledge.

Spindler, G., & Spindler, L. (1990). *The American cultural dialogue and its transmission.* London: Falmer.ddd

Statistics Canada (1998). *Daily*. Ottawa: Supply/Services Canada.

Trueba, H. (1993). Cultural diversity and conflict: The role of educational anthropology in healing multicultural America. In P. Phelan & A. L. Davidson (Eds.), *Renegotiating cultural diversity in American schools.* New York: Teachers College Press.

Acerca de los Autores

Judith K. Bernhard
Ryerson Polytechnic University
jbernhar@acs.ryerson.ca

Dr. Judith Bernhard is a native of Chile, a lecturer, author, educator, and parent. She is currently an associate professor at the School of Early Childhood Education at Ryerson University, Canada. Over the last 20 years, she has been involved in a number of initiatives including a national study on diversity (Paths to Equity: Cultural, linguistic, and racial diversity in Canadian Early Childhood Education) and the establishment of the Early Childhood Diversity Network. Dr. Bernhard teaches child development courses and has a particular interest on infants, toddlers, immigrant and refugee children. Her focus is on pluralistic models that honour the role of culture in determining optimal caregiving. She has written extensively on the topics of human development and the situation of Latin American children and their families in Canada.

Veronica Pacini-Ketchabaw
Ontario Institute for Studies in Education
University of Toronto

Veronica Pacini-Ketchabaw is a native of Argentina. She is currently a doctoral candidate at the Ontario Institute for the Studies in Education of the University of Toronto and an instructor at the School of Early Childhood Education at Ryerson Polytechnic University in Canada. She is involved in a national study of multi-age approaches to early childhood education in contexts of cultural and linguistic diversity. she has worked with the Latin American community in Toronto in a variety of research projects. She is interested in how social relations of knowledge relate to educational issues confronting culturally and linguistically diverse populations.

Marlinda Freire
Hospital for Sick Children, University of Toronto

Marlinda Freire, M.D., F.R.C.P.(C) Chief Psychiatrist, Toronto District School Board. Assistant Professor, Faculty of Medicine,

Department of Psychiatry, University of Toronto. Staff Psychiatrist, Hospital for Sick Children. Research Associate. Joint Centre of Excellence For Research on Immigration and Settlement, Federal Government of Canada. Research Associate. Centre for Refugee Studies, York University. Associate Fellow on Latin America and the Caribbean, York University.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http: epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@ asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@ asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@ unh.edu .

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto @ servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
aacosta@cucea.udg.mx

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Andrew Coulson
a_coulson@msn.com

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Richard M. Jaeger
University of North Carolina--Greensboro

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
Arizona State University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
Ann Leavenworth Center for Accelerated Learning

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois--UC

Robert Stonehill
U.S. Department of Education

Robert T. Stout
Arizona State University

David D. Williams
Brigham Young University

archives | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 174 times since December 8, 2000

Education Policy Analysis Archives

Volume 8 Number 53 **December 8, 2000** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2000, the **EDUCATION POLICY ANALYSIS ARCHIVES.**
Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

The Use of Logic in Educational Research and Policy Making

Rick Garlikov
Birmingham, Alabama (U.S.A.)

Abstract
While educational research is an empirical enterprise, there is significant place in it for logical reasoning and anecdotal evidence. An analysis of the article by Scott C. Bauer, "Should Achievement Tests be Used to Judge School Quality?" (*Education Policy Analysis Archives, 8*(46). Available: http://epaa.asu.edu/epaa/v8n46.html) is used to illustrate this point.

I want to use the following to help demonstrate the importance of logic, philosophy (particularly conceptual analysis), and insights based on anecdotal evidence, for educational research and policy making.

In "Should Achievement Tests be Used to Judge School Quality?" (*EPAA*, Vol. 8, Number 46) Scott C. Bauer stated the following:

> At the 1998 Annual Meeting of the Mid-South Educational Research Association, W. James Popham raised the following question: Is it appropriate to use norm-referenced tests to evaluate instructional quality? Specifically, he challenged participants to consider whether norm-referenced tests measure knowledge that is taught and learned in schools. Popham then invited researchers to participate with him in a study to answer the question: Should student scores on standardized achievement tests be used to evaluate instructional quality in local schools?
>
> In a subsequent paper, Popham (1999) laid out the basic argument that frames this study. While standardized achievement tests are useful tools to provide evidence about a specific students' mastery of knowledge and skills in certain content domains, "Employing standardized achievement tests to ascertain educational quality is like measuring temperature with a tablespoon" (p. 10). There are several difficulties with using aggregate measures from norm-referenced tests to judge the performance of a school. [Two of these are described, which I omit here.]
>
> [Third,] scores on standardized achievement tests may not be attributable to the instructional quality of a school. Student performance may be caused by any number of factors, including what's taught in schools, a student's native intelligence, and out-of-school learning opportunities that are heavily influenced by a students' home environment. Popham terms this last issue the problem of "confounded causality."
>
> Here we report the results of one of several local studies designed to provide empirical evidence to answer the question of whether student scores on standardized achievement tests represent reasonable measures of instructional quality.

This last sentence is only true if the term "reasonable" is understood to mean something like "credible to people who think about the issue in certain ways" or "credible to reasonable people who think about the issue in certain ways." It has to be understood in a way not dissimilar from the legal principle of considering "what a reasonable person would have believed or done in a similar situation" in order to assess the guilt or innocence of a defendant. This is because the study only actually surveys what people **believe** in regard to whether students who gave correct answers to individual standardized test questions were more likely to have been taught the information necessary to answer those test items in school or were

more likely to have learned it elsewhere. The study did **not measure** whether students did learn the information in school or whether they learned it elsewhere, but whether teachers and parents **thought** students learned the information in school or learned it elsewhere.

Consider the following paragraph in Bauer's article:

> The notion that aggregate scores on standardized tests should serve as an indicator of school quality relies on an assumption of causality. The underlying logic is that the scores are predominantly caused by something the school does or has some control over. For this assumption to hold, at a minimum we must be willing to believe that student performance on standardized tests is related to school quality, that the tests measure the skills and abilities stressed in school programs, and that there are no antecedent factors that might otherwise explain aggregate student performance on the tests. If the data presented here are credible, the soundness of this assumption must be questioned. On average about half of the items on the rated test suffer from "confounded causality" on at least one of these criteria.

There is an ambiguity in the word "should", as he uses it, in the first sentence—the two meanings being (1) "should" in the political sense of whether policy ought to rely on standardized test scores to judge schools because people accept or believe that test items show direct causal correlations between the quality of school instruction and student test scores and thus, by extension, accept test scores of a measure of the efficacy of what is taught and learned in schools, (2) whether test items actually show direct causal correlations between school instruction and student test scores and thus serve as an actual measure of what is taught and learned in schools.

In the second sense it is not true that "For this assumption to hold [i.e., the assumption that scores are predominantly caused by something the school does or has some control over], at a minimum we must be willing to believe that student performance on standardized tests is related to school quality...." For the assumption to hold, what is necessary is that student performance on standardized test scores actually is related to school quality. Our beliefs about the accuracy of that statement have nothing to do with whether the assumption holds or not. We can believe it all we want, or disbelieve it all we want, and neither that belief nor that disbelief will make it true or false.

The proper conclusion is not that nearly half the items rated suffered from confounded causality, but that teachers and parents believed nearly half the items suffered from confounded causality.

The test for seeing **how much, if anything** of what is measured on standardized tests is actually taught in schools would require a very different kind of study—one which attempts either to find out precisely where students learned the information which they used to answer test items correctly, or at a minimum to find out whether

students knew the information before it was taught in school or not, using some sort of pre-test/instruction/post-test differentiation methodology.

However, this latter would still only account for students learning the information prior to instruction. It would not account for students' learning the information during or after instruction, though **not because of** the instruction (alone). For example, it is a fairly common phenomenon for teachers to "teach" a principle that students do not understand, and that a parent or someone else then explains to the student in a way that the student comprehends it. Now it may be that the parent would not have done this without the teacher's introduction, but it is still then a joint teaching effort, not a result only of school instruction alone. And I suspect there is some evidence that in school districts where there is not such parent- or mentor-child interaction about school work, students do not learn it as well nor test as well. I also suspect that success on achievement tests, and academic or "grading" success in school in general comes in large part from parent or mentor interaction with school-initiated subject matter. The same argument could be given with regard to students' learning on their own—through reflection or additional study from other sources—material that was introduced in the classroom but that was not learned **in** the classroom nor **from** what the teacher (or textbook) said or did.

The point, however, is that **where and when** students have learned something **is** a **social science** kind of question, as is the question of where and when **what proportions** of students learn a particular item in school or elsewhere. And it is not dependent upon where or when parents or teachers or anyone **thinks** students have learned something—unless the parent or teacher knows for sure. (The problem for the social scientists, however, in this latter case is ascertaining whether the parent does know for sure or not, because even if the parent is correct and does know, it is difficult for someone else to know the parent's claim is correct, particularly if the researcher or other third party was not present during the process.)

But now consider Popham's (or Bauer's, I can't tell which) claim: "Finally, scores on standardized achievement tests may not be attributable to the instructional quality of a school. Student performance may be caused by any number of factors, including what's taught in schools, a student's native intelligence, and out-of-school learning opportunities that are heavily influenced by a students' home environment."

If that is true, as it certainly seems to be since students do learn things, or figure out things, on their own or from others outside of school – things which sometimes are tested on standardized tests—**that** is alone sufficient to show that test scores **cannot** be reasonably attributable to instructional quality in schools alone. For if there are possible and reasonably likely other "confounding" or contributing causes of student success on standardized tests, then logic alone demands that test scores cannot **legitimately** be used to assess the quality of school instruction. Surveys about parent or teacher

beliefs regarding this matter are unnecessary and logically irrelevant.

But that does not make this survey nor this paper unimportant. There are two things involved that are important. The first is that something may be politically popular even if it is not legitimate. So a survey of whether people think that standardized test scores reflect the quality of instruction in schools may be important to know for determining public policies (and news reporting policies) about using and/or reporting such assessments. If it turned out that the public did not have as much confidence in or concern about this form of assessment as legislators and newspapers seem to think they have, it might be politically feasible to get rid of these tests in a way that reasoning alone will not permit, because what is thought important to report in the news and what is thought necessary to legislate are often more dependent on what is believed to be desired by the public than on what reason might show is desirable or what evidence might show is false about public perceptions.

Second, this survey is interesting and useful as a teaching tool for the public, and in that regard is very important. For what Bauer has done is to show that people who look at individual test items are not confident about the significance of individual test item scores, and that **therefore they cannot be confident about the meaning or significance of aggregate scores**, and that, by extension, no one can be. It is one thing for someone to believe tests are significant without looking at and reflecting on the individual questions and the significance of each of them; it is quite another to believe that tests scores have significant meaning after examining the individual test questions and their likely significance. The survey was a way of getting people to do such an examination and to show them, and others, what happened when they did. For many people that is more convincing than logic alone, even if it should not logically be necessary.

I point out the above using the Bauer study because that study is not unique in educational research in regard to trying to demonstrate what is essentially a logical matter by use of empirical research. Further, it is not unique in educational research for researchers to draw logically unwarranted or unjustified conclusions from perfectly good data that they have collected. The point is that while logic and philosophy or conceptual analysis alone are often insufficient to provide knowledge about educational phenomena, they are both necessary in order to understand the significance of such data.

Moreover, they often show what data to seek. When Popham, or anyone, first realized that there logically could be confounded causality in regard to students' answering standardized test items correctly, that realization alone showed there was a problem that needed to be studied empirically in order to determine whether the logical possibility was the actual or likely or even systematic or overwhelming occurrence. But all too often in educational research and in educational policy-making, it is "empirical" research that is held to be all that is important, not logic nor anecdotal evidence nor insight based on anecdotal evidence. That seems to me to be a mistake

because while logic and apparent single occurrences alone do not show what is happening systematically or statistically, they point out matters that either need to be studied empirically or they point to conceptual problems that may have to be addressed before empirical studies can be done. In some cases they also point out the actual futility of relying on a practice or policy that intuitively seems to be effective and that may even be traditional—such as determining the efficacy of schools by comparing (standardized) test scores. There are far more logical and conceptual matters involved in education and in educational research than is commonly believed or accepted. And I think it is a grave mistake to think that empirical studies alone are the proper or necessary way to do educational research and the only proper means to guide educational policy.

Reference

Bauer, S.C. (2000). Should Achievement Tests be Used to Judge School Quality? *Education Policy Analysis Archives, 8*(46). Available: http://epaa.asu.edu/epaa/v8n46.html

About the Author

Rick Garlikov
Rick Garlikov is a philosopher and photographer who resides in Birmingham, Alabama. He holds a graduate degree in Philosophy from the University of Michigan. He is the author of *The Meaning of Love*, of *Making the Most of Your University Courses: What to Expect Academically at College*, and of *Teaching About Thinking; Thinking About Teaching: Why Teaching "Facts" Is Not Enough*, an online book of essays about teaching for reasoning and understanding. Rick conducts introductory philosophy instruction via e-mail and offers a free "homework help" service that tries to help students (or parents) understand in greater depth the material with which they are having difficulty, so that they can then work through their assignments (or teach their children) on their own. These materials and services are available at http://www.garlikov.com.

Copyright 2000 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Greg Camilli
Rutgers University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher
Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher
Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu