DOCUMENT RESUME

ED 483 038 EA 032 593

AUTHOR	Walberg, Herbert J.
TITLE	Improving Educational Productivity.
SPONS AGENCY	Institute of Education Sciences (ED), Washington, DC.
REPORT NO	No-1
PUB DATE	2003-00-00
NOTE	63p.
AVAILABLE FROM	For full text: http://www.temple.edu/lss/pdf/publications/pubs2003-1.pdf.
PUB TYPE	Reports - Research (143)
EDRS PRICE	EDRS Price MF01/PC03 Plus Postage.
DESCRIPTORS	*Educational Improvement; Educational Quality

ABSTRACT

The purpose of this report is to synthesize (a) meta-analyses of control-group research and (b) econometric analyses of large-scale surveys, both of which reveal the causes of achievement. (Author)

Improving Educational Productivity

by

Herbert J. Walberg

University of Illinois at Chicago, Emeritus

2003
Publication Series No. 1

This research was supported in part by the Institute of Education Sciences (IES) of the U.S. Department of Education (ED) through a grant to the Mid-Atlantic Regional Educational Laboratory, the Laboratory for Student Success (LSS). The opinions expressed do not necessarily reflect the position of the supporting agencies, and no official endorsement should be inferred.

Improving Educational Productivity

H. J. Walberg

Abstract

The purpose of this report is to synthesize (a) meta-analyses (statistical analyses of results of many studies) of control-group research and (b) econometric analyses of large-scale surveys, both of which reveal the causes of achievement. Learning is fundamentally a psychological process; student motivation, instruction, and other psychological factors are the well-established, consistent, and proximal causes of learning. Three factors require close attention—quantity and quality of instruction, because educators can alter these factors, and the home environment, because it influences the large amounts of time students spend outside school, about 92% during the first 18 years of life.

Largely because of exposure to language and academic encouragement during the first 6 years of life, children begin school with vastly different abilities, differences that magnify during the school years. Effective, efficient classroom teaching methods (identified here) can diminish gaps between abilities and raise all students' achievement; yet ineffective and costly methods remain prevalent in U.S. schools.

At the school, district, and state levels, similarly effective policies can be identified, including accountability, incentives, external examinations, and small schools and small districts. Some faddish current programs such as brain-based learning and widespread education theories have little evidentiary support. Some widespread and much touted programs such as Reading Recovery and Success for All are expensive yet do no better than other programs, as research independent of the promulgators shows. Two very costly federal programs, Title I and special education, have long had poor records of promoting achievement.

Legislators, school boards, and education leaders have been slow to adopt productive reforms. Accountability and competition, however, provide incentives for education leaders to adopt practices, programs, and policies that are more effective and efficient.

Improving Educational Productivity

During the past quarter century, scholarship grew rapidly on the question of what makes a difference in students' academic learning. The 1966 Coleman Report, the largest U.S. education survey ever undertaken, precipitated such scholarship, since James Coleman and his colleagues' report to Congress (Coleman et al., 1966) found little or no effect of school resources such as per-student spending and class size on how much students learn. Reanalysis of the Coleman Report and additional studies led the eminent economist and former president of the American Association for the Advancement of Science, Kenneth E. Boulding, to write a paper in 1972 titled, "The Schooling Industry as a Possibly Pathological Section of the American Economy." He noted that schools doubled their share of the American economy from about 3 to 6% in real prices from 1930 to 1970 and were "notoriously unprogressive when it comes to productivity" (p. 135).

A few years before the Coleman Report, Benjamin Bloom (1963) published perhaps the first synthesis of many psychological studies of the stability of human characteristics including learning. He estimated that adult academic ability is fairly predictable by the age of 4 and largely predictable by the age of 8, which suggested the importance of early parental and other extramural influences rather than subsequent schooling in determining achievement.

Even so, policymakers increasingly believe that K–12 educators should be held accountable during the school years for the "value added" that they uniquely contribute to learning, a part of the "human capital" of knowledge and skills that largely determines individual and national prosperity and well-being. Though this view is traceable to Smith's 1776 *Wealth of Nations*, the widely influential report, *A Nation at Risk* (1983), gave it late 20th-century prominence.

The Productivity Predicament

The learning productivity problem is better known and even more acute today. Vastly increased spending and many school reforms resulted in stagnant achievement during the past quarter century, even though children's measured intelligence or capacity for learning increased steadily (Walberg, 1998b). Unlike most sectors of the American economy that steadily increase their productivity over time, schools become less rather than more efficient, a serious matter given the size of the education sector and the central and increasing importance of learning in the American economy and society. School productivity or the relation of achievement to costs was 65% higher in 1970–71 than in 1998–99 (Hoxby, 2001).[1] Federal expenditures of more than $120 billion, moreover, failed to diminish the "poverty gap" between poor and middle-class students.

American schools, moreover, have higher dropout rates than commonly reported. The 86% graduation rates reported by the federal government notwithstanding, only 74% of U.S. students actually graduate. The discrepancy is accounted for by the federal government counting dropouts passing General Educational Development (GED) tests as graduates, even though economists find that young people with GEDs do no better on the job market than dropouts. Minorities drop out, moreover, in large percentages: Only 56% of Blacks and 54% of Hispanics finish high school, in contrast to 80% of Whites who graduate (Greene, 2002a).

Internationally, the productivity problem is also obvious: Among two dozen affluent countries, the U.S. is near the top in per-student spending, but American students fall further behind others the longer they are in school (Organisation for Economic Co-Operation and Development, 1995–2001). Moreover, unlike a quarter century ago, today smaller percentages of U.S. students than those in other countries remain in school during the late teen years. Thus, American schools compare unfavorably in both quality and quantity, despite high costs.

Ironically, though research identified the learning problem, policymakers and educators ignored early and subsequent research pointing to productive solutions. They followed fad after fad, many of which were implausible and expensive, and none of which had the evidentiary basis required in such fields as agronomy, medicine, public health, and psychology. An additional irony is that much of the older research on what makes a difference was more rigorous but less acceptable to education theorists, perhaps because it suggests traditional, commonsense educational methods.[2]

Purpose and Scope

My purpose is to synthesize (a) meta-analyses (statistical analyses of results of many studies) of control-group research and (b) large-scale surveys that reveal the causes of achievement. These two kinds of research complement one another. Psychologists prefer control-group experiments, particularly those that randomly assign students to educational methods and conditions, and measure achievement before and after to assess progress. Such experiments have causal creditability, because differences in learning are attributable only to treatments and the luck of the draw, just as in the case of medical experiments that randomly assign patients to alternative regimens.

Experiments may be weak in generalizability, however, since they typically use small and possibly atypical samples of students, such as those in a given urban or suburban school. The statistical analysis of many control-group studies, however, can compensate for the weakness of any single study, since the pervasiveness of an effect can be ascertained by statistically analyzing

a variety (usually all) of samples. Such analyses can show whether an educational method works as well for boys as for girls and for urban, suburban, and rural students at various grade levels and in various school subjects.

Analyses of state, national, and international surveys can also reveal the generalizability of findings. Epidemiologists, economists, political scientists, and sociologists conduct such research, which usually encompasses whole populations or random samples. These analyses, however, yield somewhat less certain causal inferences, since they "control" for alternative causes. In achievement research, these usually include prior achievement, socioeconomic status, or poverty, which may be poorly specified and measured. Such analyses of large-scale surveys may also omit plausible causes, since measures of them were left out of the surveys originally designed for purposes other than the analyst's.

In the last decade, however, such data sets and analyses improved remarkably, particularly in measuring learning rather than achievement, that is, in assessing "value-added" gains or learning over, say, the year from a pretest to a posttest. Analyzing achievement at a single point in time may be misleading, since achievement may be attributable to prior causes, such as infant poverty or prior good or bad teaching, rather than to current conditions or methods.[3]

Though economic, sociological, and political factors affect learning, their influence is indirect. Learning is fundamentally a psychological process; student motivation, instruction, and other psychological factors are the well-established, consistent, and proximal causes of learning. Thus, this report starts with psychological factors before analyzing the social conditions that affect learning less directly.

Psychological Causes of Learning

The scarce resources in learning are opportunity and concentration rather than the amount of information available or the processing capacity of the mind. Herbert Simon, the Nobel economist and psychologist, combined these fields to synthesize what might be called the economics of cognitive learning. Summarized in this section, his synthesis sets the stage for understanding what helps students learn.

If a lifetime were devoted to acquisition of information, according to Simon's estimates, about 200 million items could be stored. "Hence, the problem for humans is to allocate their very limited processing capacity among several functions of noticing, storing, and indexing on the input side, and retrieving, reorganizing, and controlling his [sic] effectors [actions] on the output side" (Simon, 1981, p. 167).

Psychological Processing of Information

Aside from external incentives and opportunities, the major constraints on the acquisition of knowledge are the few items of information, perhaps 2 to 9, that can be held in short-term memory, and the time required, 5 to 10 seconds, to store an item in long-term memory. In chess, mathematics, science, writing, and other fields that have been studied, experts differ from novices not only in having more information in permanent memory but also in processing it more efficiently. Among experts, for example, items of information are thoroughly indexed and can be rapidly brought to conscious memory.

Among experts, items are also elaborately linked with one another, which confers two advantages: the ability to recover information by alternative links, even when memory loses parts of the direct indexing, and the capacity for extensive means–ends or trial-and-error searches. Such processes come into play in problem solving, from the elementary insights of novices through the advanced discoveries of eminent experts.

The expert's greatest advantage over novices is "chunking," the representation of related items of verbal, numerical, spatial, and other information as a single condensed symbol. A few seconds of study, for example, enable amateur chess players to remember the positions of only a few pieces, but masters may take in a whole board, because they readily perceive variations of a few standard chunked patterns rather than the individual pieces.

The sizes of chunks of information assimilated from the environment enlarge with experience and practice, because expertise confers knowledge on what information to acquire and how to code it efficiently. Experience and education (or guided experience) enable learners to assimilate ever-larger parts of the environment. For example, nearly a century ago, Edmund Huey (1908) showed that brief pauses in reading (called "fixations") do not vary in duration among novice and skilled readers; rather, as readers improve, they assimilate increasingly larger chunks of text ranging from parts of letters to words and phrases.

Time for Processing

Simon estimates that 50,000 chunks, about the same number as the recognition vocabulary of college-educated readers, may be required for expert mastery of a special field. The highest achievements in various disciplines, however, may require a memory store of 1 million chunks, which may take 70 hours a week for roughly a decade to acquire, notwithstanding such 7- to 9-year exceptions as Mozart and the chess master Bobby Fischer.

Speed may also be traded for accuracy in assimilation and processing. In accelerated reading or skimming, for example, chunks are sampled, and intervening meanings may be

assumed—perhaps mistakenly. The fastest readers may presume to grasp paragraphs and pages with instantaneous fixations, although those who read *War and Peace* in an hour may recall only that it was about old Russia or, worse, World War II.

Meaning and Culture

Assimilation also depends on the correspondence of meanings in the material and the memory store. Psychological experiments, for example, show that, with randomly placed chess pieces in "unchunked" patterns, masters have little advantage over novices in reproducing the board. Unchunked nonsense words of randomly generated letters, similarly, put both slow and fast readers at a vast disadvantage in recalling sentences.

Language mastery, the fundamental and pervasive skill necessary for achievement in school, is determined more by experience than by psychometric intelligence. Most bright American adults, for example, would be reduced to the level of infancy or feeblemindedness in listening to Swahili or Tamil. In principle, nonetheless, a dictionary, a coach, and unlimited time might overcome their limitation. For children, "total immersion" in their mother tongue is almost universally generally effective for listening and speaking. Immersion, moreover, in a noncognate language for as little as a year appears sufficient to bring children and adults to near-native fluency (though pronunciations are difficult for adults). Decisive is the amount and intensity of the experience rather than age or psychometric intelligence (Walberg, Hase, & Rasher, 1978).

Meaning that conforms to expectation and experience promotes speed of mental processing. Just and Carpenter (1980), for example, showed that even single aberrant words and vague transitions can slow processing of ordinary text. They presented readers with the following two pairs of sentences:

1. A. It was a dark and stormy night and the millionaire *died.*
   B. The killer left no clues for the police to trace.
2. A. It was dark and stormy the night the millionaire *was murdered.*
   B. The killer left no clues for the police to trace. [emphasis added]

Analyses of eye movements showed that readers took a half second longer to read the first pair, because the transition is illogical; they had the distracting and time-consuming task of guessing that the millionaire had been murdered. The time taken by careful writers to avoid such lapses saves the time and good will of readers.

So, too, in the case of parents and teachers. To foster learning, they can best provide logical, readily understood explanations suitable to learners as well as the time, opportunity, and incentives for them to learn. These simple, commonsense principles set the stage for

understanding research on the psychological causes within and outside school that foster achievement.

Nine Psychological Causes of Learning

Practice makes perfect, says an old adage. An analysis of time effects on learning suggests the obvious: 88% of 376 study estimates revealed the positive effects of various aspects of study time such as preschool participation, school attendance, amount of attention to lessons, amount of homework, and length of the school year (Walberg, 1998b). The positive effect of time is perhaps most consistent of all causes of learning.

Though the widely read 1983 report, *A Nation at Risk*, called attention to the American school year as the shortest among economically advanced countries, U.S. students still spend about half as much time in total study hours per year as students in countries that top the achievement charts, and until recently with the advent of summer and after-school programs, time remained neglected among school reforms.[4]

In addition to time, intensity also rules: Illogical or unsuitable instruction or student inattentiveness may mean that little is accomplished, notwithstanding much study time. Other psychological conditions also have a causal bearing on learning. The most consistent, powerful, and direct productivity factors or psychological causes of learning are defined in Table 1 (tables and figures appear at the end of this article in the order of their first mention in the text).

This taxonomy of three sets of nine factors derives from an early synthesis of 2,575 study comparisons (Walberg, 1984) that suggested that these factors are the chief psychological causes of academic achievement (and, more broadly, school-related cognitive, affective, and behavioral learning).[5] Subsequent syntheses have shown results consistent with the original findings. Each of the first five factors—prior achievement, development, motivation, and the quantity and quality of instruction—seems necessary for learning in school. Without at least a small amount of each, the student may learn little. Large amounts of instruction and high degrees of ability, for example, may count for little if students are unmotivated or instruction is unsuitable. Each of the first five factors appears necessary but insufficient by itself for effective learning.[6]

Instructional Time

Time is a particularly pervasive constraint, since U.S. students have the shortest school year among countries of the industrialized world[7] and generally do far less homework than students. For this reason, the out-of-school time factors—home environments, peer groups, and exposure to mass media, particularly television—strongly influence learning.

Even so, time in school may be inefficiently employed. It has been wisely said that it is as useless to teach students what they already know as it is to teach them what they are as yet incapable of learning. High-quality instruction can be understood as providing information cues, correctives, and positive reinforcement or encouragement that insure the fruitfulness of engaged time. Careful diagnosis and tutoring can help make instruction suitable for students. Inspired teaching can help students to persevere. Quality of instruction, then, may be considered an efficient enhancement of study time.

Similarly, the four psychological environments indicated in Table 1 can expand and enhance learning time. Classroom morale is measured by obtaining student ratings of their perceptions of the classroom group. Good morale means that the class members like one another, that they have a clear idea of the classroom goals, and that the lessons are matched to their abilities and interests; in general, morale is the degree to which students are concentrating on learning rather diverting their energies because of unconstructive social climates. Peer groups outside school and stimulating home environments can help by expanding learning time and enhancing its efficiency; students can both learn in these environments and become more able to learn in formal schooling.

The last factor, mass media, particularly television, can displace homework, leisure reading, and other academically stimulating activities; and it may dull the student's keenness for academic work. For instance, some of the average of 20–30 hours a week high-school students spend viewing television might usefully be added to the mere 4 or 5 average weekly hours of homework they report.

Three factors require close attention here—quantity and quality of instruction, because the educators can alter these factors, and the home environment, because it influences the large amounts of time students spend outside school and because it can be affected by outreach programs.

Amount of Instruction

The power of American schools to affect academic learning is limited by the surprisingly small amount of time children and youth spend in school as a percentage of all time in their lives. The calculations in Figure 1 show that U.S. students spend only an estimated 8% of their time in school during the elementary and high-school years. Students spend only 8.2% of their time in school during the first 18 years of life.[8]

Actually, the preschool years, when children are the exclusive responsibility of parents rather than educators, constitute about a third of the child's first 18 years. The time in these

critical years before school when children learn their mother tongue, motivations, habits, and many other things is much longer than the time spent in school in the remaining 12 years. These formative years, moreover, are crucial for children's intellectual development. As Bloom (1963) pointed out, children's adult abilities are substantially predictable before they reach school. Small preschool advantages and disadvantages often result in huge achievement gaps in the later elementary years.

With respect to time, the long American summer vacation appears particularly injurious to students in poverty, because they gain less exposure to academically stimulating language and experience. But in Atlanta, when poor African American students participated in summer school, they tended to make normal middle-class academic progress (Heyns, 1978). Poor Black and White students in Baltimore lost the most in achievement during the summer months (Entwisle & Alexander, 1992). Said Karl Alexander of Johns Hopkins, "That gap between high-SES [socioeconomic-status] students and low-SES students increases steadily over the years. . . . And that mostly reflects the more substantial strides upper-SES kids make during the summer months" (as cited in Viadero, 1994, p. 36).

Home Environment in Early Childhood

In addition to encouraging and supervising homework and reducing television viewing, parents can improve academic conditions in the home. What might be called "the alterable curriculum of the home" is much more predictive of academic learning than is family SES (see Walberg, 1984). This curriculum includes informed parent–child conversations about school and everyday events; encouragement and discussion of leisure reading; monitoring, discussion, and guidance of television viewing and peer activities; deferral of immediate gratification to accomplish long-term goals; expressions of affection and interest in the child's academic and other progress as a person; and perhaps, among such efforts, laughter and caprice.

In case studies of poor inner-city Chicago families, the children who succeeded in school had parents who emphasized and supported their children's academic efforts, encouraged them to read, and interceded on their behalf at school. Many statistical studies show that indexes of such parent behaviors predict children's academic achievement much better than socioeconomic status and poverty (Clark, 1983).

Cooperative efforts by parents and educators to modify alterable academically stimulating conditions in the home have had beneficial effects on learning (Walberg, 1984). In 29 controlled comparisons, 91% of the comparisons favored children in such programs over nonparticipant control groups. Although the average effect was twice that of SES, some programs

had effects 10 times as large, and the programs appear to benefit older as well as younger students.

At-Risk Students

Sizable proportions of young children, especially those in poverty, are behind in language and other skills before they begin school, and they are often placed in bilingual and special-education programs for the developmentally challenged, in which they are segregated from other children and make poor progress. The origins of their achievement problems are partly attributable to ineffective programs, but the origins can also often be traced to specific parental behaviors observed before children begin school that affect children's reading and other language skills, which are keys to achievement in academic subjects.

Sticht and James (1984) have pointed out that children first develop vocabulary and comprehension skills by listening, particularly to their parents before they begin school. As they gain experience with written language between the first and seventh grades, their reading ability gradually rises to the level of their listening ability. Highly skilled listeners in kindergarten make faster reading progress in the later grades, which leads to a growing ability gap between initially skilled and unskilled readers.

Poverty and Minority Gaps

This growing gap between good and poor readers reflects race and social class differences.[9] A chapter in the authoritative *Handbook of Reading Research* (Wigfield & Asher, 1984), concludes:

> The problems of race and socioeconomic status (SES) differences in achievement have been at center stage in educational research for nearly three decades. Research has clearly demonstrated that such differences exist; black children experience more difficulty with reading than white children, and the discrepancy increases across the school years. Similarly, children from lower SES homes perform less well than children from middle-class homes, and here too the difference increases over age. (p. 423)

These differences stem from early childhood experience, especially with respect to parent behaviors that motivate children. Studies show that middle-class parents are more likely to hold high expectations for their children's achievement and to be more often engaged with them in promoting it.

> Lower-SES mothers provide their children with poorer problem-solving strategies, and they tend to take over for their children rather than letting them do the task. . . . That lower-SES parents view school as a distant, rather formidable institution over which they

have little control; engage in less effective teaching strategies; and lack confidence in their children's ability does not bode well for their children's school performance. (p. 429)

Home observations and interviews reveal among parents further SES differences associated with higher achievement in reading, for example "responsivity of the parent, the kinds of discipline techniques used, the organization of the physical environment, parental involvement, and provision of appropriate play materials" (pp. 431–432).

Such parent behaviors cause huge and growing gaps in preparation for school and learning to read between children in poverty and those in middle-class homes, as revealed in preschool children's vocabulary growth recorded during free play (Hart & Risley, 1995). Though vocabulary differences were miniscule at 12 to 14 months of age, by age 3, sharp differences had emerged: Welfare children had vocabularies of about 500 words, middle/lower SES children about 700, and higher SES children had vocabularies of about 1,100 words, more than twice that of welfare children.

Educative Practices of Parents

These SES differences in vocabulary were strongly associated with parent behaviors exhibited in their homes. Higher SES parents spent more minutes per hour interacting with their children and spoke to them more frequently. On average, higher SES parents spoke about 2,000 words per hour to their children; Black welfare parents, only about 500 (Hart & Risley, 1995, p. 68). According to Hart and Risley, by age 4, "an average child in a professional family would have accumulated experience with almost 45 million words, an average child in a working-class family would have accumulated experience with 26 million words, and an average child in a welfare family with 13 million words" (p. 198), as shown in Figure 2.

Higher SES parents, moreover, used "more different words, more multi-clause sentences, more past and future verb tenses, more declaratives, and more questions of all kinds. The professional parents also gave their children more affirmative feedback and responded to them more often each hour they were together" (Hart & Risley, 1995, pp. 123–124). By age 4, professional parents encouraged their children with positive feedback 750,000 times, about 6 times as often as welfare parents did. The welfare parents, on the other hand, had discouraged their children with negative feedback about 275,000 times, about 2.2 times the amount employed by professional parents. Such parenting behaviors predicted about 60% of the variation in vocabulary growth and use of 3-year olds.

As Entwisle and Alexander (1993) point out, such differences are compounded when lower SES children enter school at age 5 or 6. Not only do they lack vocabulary and other skills, but they must accommodate to a middle-class institution:

> The conventions of the school, with its achievement orientation, its expectation that children will stay on task and work independently without close monitoring, its tight schedule of moving from lesson to lesson, its use of "network" English, its insistence on punctuality, and its evaluation of children in terms of what they can do instead of who they are, all can be daunting. (p. 405)

Yet "Many minority and disadvantaged children cross the first-grade threshold lacking competencies and habits of conduct that are required by the school" (p. 405). Further,

> Lower SES children are much more often identified by their kindergarten teachers as being at risk for serious academic or adjustment problems; they are absent more in the first grade; and they receive lower teacher ratings on behaviors related to school adjustment such as interest/participation and attention span/restlessness (the latter two strongly predict later academic progress). (p. 407)

Students who are behind at the beginning of schooling or slow to start usually learn at a slower rate; those who start ahead gain at a faster rate, which results in what has been called cumulative advantage or the "Matthew effects" of the academically rich getting richer after the passage in the chapter of Matthew in the Bible (Walberg & Tsai, 1984). These effects are pervasive in school learning, including the development of reading comprehension and verbal literacy (Stanovich, 1986). Ironically, although improved instructional programs may benefit all students, they may confer greater advantages on those who are initially advantaged. For this reason, the first 6 years of life and the "curriculum of the home" are decisive influences on academic learning.

Success for Students in Poverty

These depressing patterns are hardly inevitable, as suggested above and in subsequent sections. Preschool programs, full-day kindergarten, and school–parent programs can help parents mitigate Matthew effects. In 47 states and the District of Columbia, moreover, effective education policies and teaching practices have enabled more than 4,500 high-poverty and high-minority schools (high meaning over 50%) to perform among the top one third of schools in their states and often to outperform predominantly white schools in advantaged communities. These schools educate about 1,280,000 low-income students, about 564,000 Black students, and about 660,000 Latino students (the groups overlap).

How do these schools do it? Their principals tend to report the following features of their schools:

- extensive use of state/local standards to design curriculum and instruction, assessment of student work, and evaluation of teachers;
- increased instruction time for reading and mathematics;
- substantial investment in professional development for teachers focused on instructional practices to help students meet academic standards;
- comprehensive systems to monitor individual student performance and to provide help to struggling students before they fall behind;
- parental involvement in efforts to get students to meet standards;
- state or district accountability systems with real consequences for adults in the school; and
- use of assessments to help guide instruction and resources, and as a healthy part of everyday teaching and learning.

These findings (Education Trust, 2001) corroborate research syntheses of control-group research and large-scale analyses of surveys discussed in subsequent sections of this report.

Because children in poverty often failed to thrive in the early grades and fell increasingly behind in the later grades, Head Start and other preschool programs have been provided for the last 3 decades. A 1985 meta-analysis of about 300 studies of these programs revealed that their moderate immediate effects on achievement and other cognitive tests faded within 2 to 3 years; that is, program students did better on achievement tests than control-group students at the end of the program, but the difference between the groups diminished to insignificance (White, 1985). Since 1985, the programs have attempted to improve by concentrating on children's academic readiness, and recent reviews have been more encouraging (Currie, 2001; Karoly et al., 1998).

The only long-term study of an academically focused school-related program showed significant long-term effects and cost-effectiveness. The Chicago Child–Parent Centers (CPC) provided academic and family-support services to children, beginning at age 3. The program emphasized the acquisition of language and prematheматical experiences through teacher-directed, whole-class instruction, small-group activities, and field trips. Parental participation in the program was intensive, with coordinating activities in each center's parent resource room.

Compared with matched control-group children, the 989 CPC children in the program showed higher cognitive skills at the beginning and end of kindergarten, and they maintained greater school achievement through the later grades. As reported in the *Journal of the American Medical Association*, a study funded by the National Institutes of Health and the Department of

Education showed that, by age 20, CPC graduates had substantially lower rates of special-education placement and grade retention than the control group, a 29% higher rate of school completion, and a 33% lower rate of juvenile arrest. A cost-benefit analysis showed that at a per-child program cost of $6,730 for 18 months of part-day services, the age-21 benefits per child totaled $47,759 in increased economic well-being and reduced expenditures for remediation (Reynolds, 2000; Reynolds, Temple, Robertson, & Mann, 2001). Very few education studies have either followed children as long or calculated the costs and benefits of the programs.

Several features made for the program's effectiveness. Unlike other early childhood programs that emphasize "developmental appropriateness," self-esteem, and play,[10] the CPC program directly taught academic language and number skills. The staff coordinated preschool activities with continuing kindergarten services in neighborhood schools. Parents were intensely involved in the program and provided academically stimulating experiences for their children at home.[11]

The results extend the range of evidence for the effectiveness of three of the nine productivity factors, namely, the home environment, the quality of instruction (particularly its academic emphasis), and the amount of instruction, since the children were given the advantage of extra academic time before starting school. Even so, both the program and the evaluation are unique. Most programs lack the CPC features, and even the RAND review of recent evaluations found that about half the programs showed no significant effect on achievement.

Quality of Instruction

Because the research on achievement is voluminous, scholars have synthesized it in several ways to make it more useful to policymakers and educators.[12] Rather than describing the findings of each study, they employ meta-analyses of many studies to calculate the consistency and magnitude of the effects of educational conditions and methods so that the most effective can be chosen on the broadest evidentiary basis. Other things being equal, for example, a teaching method proven superior in 90 out of 100 studies is preferable to one that excels only in 60 out of 100 studies.

The preferred calculation, featured here, is the "effect size," which reveals the size of any particular effect averaged across many studies.[13] The costs of the methods are ignored, since data about them are largely unavailable, and since most methods can be incorporated as part of normal school budgets because they simply involve different ways of teaching. Educators, moreover, presumably have had experience with many of the methods or at least should have been exposed to them in the course of professional study and experience. A meta-analysis of 50 years of

research on influences on learning reviews methods that have positive impacts on learning (Wang, Haertel, & Walberg, 1993b).[14]

Nine Categories of Instructional Methods

Table 2 shows the effects of instructional methods divided into nine categories. These can be more broadly grouped: graphic representation, especially in the form of road maps of what is to be learned; goal setting; and feedback to provide direction and redirection. Identifying similarities and differences, summarizing, and generating and testing hypotheses require students to think and express ideas in forms different from presentations. Cooperative learning provides opportunities for students to assimilate and present ideas by explaining various aspects to one another. Homework and practice are indexes of engaged study time. Reinforcement and recognition provide incentives for performance.[15]

The largest collection of estimated effect sizes, which covered 275 methods and conditions (Walberg & Lai, 1999),[16] provides further illustration. Discussed here are several of the largest effects from that collection, including those for traditional methods that have large effects, several newly published effect estimates, and a few selected to show the range of inquiry about instructional quality (see Table 3), on which the following discussion is based.[17]

General Methods

The elements of instruction in Table 3 can be considered the most fundamental psychological variables in learning. Cues present what is to be learned and how to learn it. Engagement is the degree to which learners actively participate. Corrective feedback signals mistakes and furnishes redirection. Reinforcement—one of the largest general effects uncovered—provides encouragement and information that learning is correct.

Mastery learning combines the elements of instruction and requires mastery of learning units before students proceed to the next unit of instruction. In particular, it allows some students as much as 5 times more instructional time and additional cues, corrective feedback, and reinforcement. Computer-assisted instruction can provide these elements to each student individually. Though beneficial to students in general, even college students, it appears particularly effective in developing skills among handicapped students and those in the early grades.

Direct instruction can be viewed as traditional or conventional whole-group teaching done well. Specifically, it consists of phases: (a) daily review, homework check, and, if necessary, reteaching; (b) rapid presentation of new content and skills in small steps; (c) guided student practice with close monitoring by teachers; (d) corrective feedback and instructional

reinforcement; (e) independent practice in seatwork and homework with high, more than 90 percent, success rates; and (f) weekly and monthly reviews. Comprehension instruction is similar and consists of three phases: (a) modeling, in which the teacher exhibits the desired behavior; (b) guided practice, where the students perform with help from the teachers; and (c) application, in which the student performs independently.[18]

Most of the other general methods in Table 3 can be broadly summarized under the rubrics of the instructional elements. Goal setting, adjunct questions, explanatory graphics, and frequent testing provide cues, reinforcement, and corrective feedback. Homework, especially with comments and grades, provides engagement.

Special Methods

Some instructional methods, though they exemplify general principles discussed in previous sections, apply only to particular skills. Consider reading, perhaps the most important skill learned before and during schooling. Phonemic awareness, repeated oral reading, and phonics provide beginning readers with mastery of sound-and-letter correspondences they may not have learned at home, in preschool, or in kindergarten (National Reading Panel, 2000).

Writing may be best taught by writing practice, that is, having students express in their own words what they have inquired about. They can also learn by applying questions and criteria such as clarity and concision to their own and others' writing and then making improvements. Combining their own sentences with those of others adds to their skill in employing appropriate sentence variety.

Grouping allows increases in instructional suitability. Accelerating gifted students allows them to learn at a faster pace without detracting from other students' learning. Tutoring tailors instructional elements to each student. Mainstreaming "handicapped" students into regular classes rather than segregating them in all-day or "pull-out" programs avoids stereotyping and stigmatizing them and helps them make normal progress. What they usually need is more and better, not special, instruction.

The last set of results in Table 3 shows that teachers themselves benefit from instructional elements, particularly feedback on their classroom practices, whether on new methods of teaching or on those that should be in their repertoire. New learning of difficult teaching skills may require specific practice with cues, reinforcement, corrective feedback, and engagement until they reach mastery, just as in the case of students.[19]

Student Effects

Although psychologists and sociologists have long studied the correlations or coincidence of student backgrounds and achievement, even substantial and consistent correlations are weak indicators of causality, since they lack experimental or statistical controls.[20] Even so, they are worth considering, just as consistent correlations of cigarette smoking and lung cancer hardly prove but do suggest consideration of causality, particularly if they corroborate other evidence, say, causality in experimental studies of mammals exposed to varying degrees of cigarette smoke. In addition, some student characteristics such as motivation can be indirectly altered by incentives, as indicated by both experimental and multivariate studies.

Prior Knowledge

Table 4 expresses correlations as effect sizes comparable to those in Tables 2 and 3. Students' prior knowledge has a huge predictive and possibly causal effect, perhaps since knowledgeable students can increase their learning from a bigger base (see previous discussion of chunking and Matthew effects). Previous success may also motivate students.

Motivation

Motivation itself is closely associated with how much students learn. Multivariate analysis of surveys and control-group studies of reinforcement corroborate its causal influence. Perhaps the most exciting demonstration of motivational effects is the Dallas O'Donnell Foundation Advanced Placement Incentive Program. The Foundation paid both teachers and students $100 for each Advanced Placement examination passed. In the nine participating Dallas public schools, sharply increasing numbers of boys and girls of all major ethnic groups took and passed the AP exams. The number rose more than twelvefold from 41 the year before the program began to 521 when it ended in 1994–95. After terminating, the program continued to have carryover effects: In the 1996–97 school year, 2 years after the program ended, 442 students passed, about 11 times more than the number in the year before the program began (Walberg, 1998a).[21] This massive effect sharply contradicts the prevalent idea in education that learning must be intrinsically motivated for its own sake.

Home Environment

The effect of the home environment can be taken very seriously for several reasons. Control-group studies corroborate many correlational findings. The home effect is far larger than apparent socioeconomic effects. Something can be done about home environments: School–parent programs can help parents academically stimulate their children by reading to them, taking

them to libraries, guiding and discussing leisure television viewing, cooperating with home visitors and teachers, and similar practices.

Grouping

Grouping students reflects common sense. If students with similar levels of knowledge and skills are grouped together, teachers can avoid teaching them what they already know and what they are yet incapable of learning; with instruction more suited to them, students should find learning more efficient and pleasant. What forms can such grouping take and what are the achievement and other effects?

Developing Prerequisites

As discussed in a previous section, a rigorous long-term study suggests that children at risk of school failure because of poverty appear to benefit from high-quality, academically focused preschools that prepare them for learning in kindergarten and subsequent grades. Closer in preparation to middle-class children, such better-prepared children may continue to benefit as late as early adulthood. Many other studies, however, show no effects or quick fade-out of early gains.

Grade Retention

By itself, retaining students in grade appears ineffective. On the other hand, "socially promoting" unqualified students may give them and their classmates little reason to study. This policy, common in big cities, probably devalues the high-school diplomas of qualified graduates in the eyes of employers and others.

As discussed below, Chicago's Summer Bridge program for failing children threatened grade retention and provided intensive academic summer school. Though some students failed, the program showed outstanding effects; it was not only effective but also highly cost-effective (Betts & Costrell, 2001). Thus, preschool and summer bridge programs tend to homogenize student achievement, that is, bring laggards up to others' achievement levels, which probably contributes to more effective learning for both groups.

Classroom Grouping

Widely used in elementary schools, homogeneous achievement grouping within classes has small, positive effects (about .25 on average). In the later elementary grades, Matthew effects (of the rich getting richer) have typically caused wide variations in student achievement; a sixth grade may have third- and ninth-grade readers. Probably for this reason, the "Joplin plan" of

bringing like-ability students from different classes and grade levels into homogeneous groups has larger effects (about .35) than within-class grouping. As identified by ability or achievement tests, highly able students benefit from "enrichment" of their studies, that is, the provision of greater depth of regular grade-level content (.40). "Accelerated" homogeneous high-ability classes that allow students to study advanced-grade material benefit them greatly (.90; Kulik, in press).

Tracking

By high school, student achievement levels differ more widely, and most American high schools practice tracking; about 86% of high schools, for example, track mathematics classes. Some scholars urge "detracking" (Oakes & Lipton, in press), that is, heterogeneously grouping all high-school classes, but surveys "show solid support for tracking among parents, teachers, and students" (Loveless, 1998, p. 1). Research on detracking is insufficiently rigorous to support the policy.

School Effects

School-level research is less rigorous than studies of individual children and classes. Why? If, in a particular school, half the teachers practiced ineffective methods and half practiced effective methods, the net result would be an average teaching effect, which would conceal important effects within the school. Many school-level studies, moreover, have inadequately measured and controlled for prior achievement and other productivity factors with strong records of affecting learning. Even so, for the sake of completeness, the possible school-level influences are worth considering, particularly those corroborated by control-group research and statistically controlled analyses of student and classroom effects.

Curriculum Alignment

Table 5 shows a strong influence of opportunity to learn, which refers to the extent that education goals, curriculum, instruction, and testing are "aligned." Most centrally, this means that what is tested overlaps with what is taught.[22] Aside from the Australia, Canada, Germany, and the U.S., most nations have national education systems, which allow such alignment across schools in each country. As discussed in a subsequent section, many individual states such as North Carolina and Texas are aligning their systems of education, so that if education goals are X, Y, and Z, curricula, teaching, and testing are geared not toward M, N, and O but toward X, Y, and Z.

Goal Setting

Psychological studies support the idea of setting national, state, and local achievement goals. Laboratory control-group research and field studies in a wide variety of organizations confirm the effects of setting goals on task performance. Nearly all studies showed that setting specific, challenging goals led to higher performance than setting easy goals, "do your best" goals, or no goals. "Goals," it has been concluded,

> affect performance by directing attention, mobilizing effort, increasing persistence, and motivating strategy development. Goal setting is most likely to improve task performance when the goals are specific and sufficiently challenging . . . feedback is provided . . . the experimenter or manager is supportive, and assigned goals are accepted by the individual. (Lock, Shaw, Saari, & Latham, 1981, p. 125)

Other School Effects

Table 5 shows that school-level instructional time, student monitoring, and parental involvement influences are positive and coincide with classroom- and student-level research. The school-level effects are smaller, however, perhaps because, as noted above, they average important differences among classrooms and students within schools and because they may be unreliably reported on questionnaires rather than observed. Perhaps because they are vague and difficult to measure, school climate, administrator leadership, and staff cooperation are the weakest apparent school-level influences.

Policy Effects

Adam Smith pointed out in the *Wealth of Nations* that human capital in part determines prosperity and the quality of life of nations and individuals. The Chicago School of Economics confirmed this now commonly held view. Modern "information economies" require ever-increasing knowledge and skills; individuals who possess them are likely to thrive. For this reason, economists have increasingly joined psychologists in asking what best promotes knowledge and skills. To find out, they have, in the last decade or two, conducted policy analyses of state, national, and multinational achievement surveys.[23]

Even more than psychologists, economists assume that people tend or try to act rationally; they seek to employ scarce means that best advance the realization of their values (which may include altruistic and unpriced values).[24] Thus, information or, more precisely, perceptions of present and future benefits and costs greatly matter. New information and changed incentives can therefore change behavior. This working assumption helps explain not only traditional economic phenomena but also many policy issues.[25]

State Reforms

A Nation at Risk and subsequent reports showed Americans the importance of achievement for national and individual prosperity and welfare. The congressionally commissioned National Assessment of Educational Progress, however, has shown little achievement change since then, which has led to increasingly substantial reforms. Some, discussed below, have shown positive learning effects. Validated by large-scale achievement surveys, economic principles help explain why some states made such substantial reforms and estimated the achievement impact.

It might be asked, for example, which states developed the best achievement standards (and why). A general economic answer is those states that had the greatest incentive. The specific answer is those states that had the poorest achievement. Within the last 2 decades, the National Assessment of Educational Progress began reporting the first valid state achievement comparisons (based on random samples of students within each participating state). Within the past decade, states with the poorest achievement, typically deep South states such as Alabama and Mississippi, developed clear, specific content standards and implemented assessments aligned with the standards (Betts & Costrell, 2001). Since citizens, businesspeople, and legislators in such states recognized their deficiencies as well as the dependency of growth and welfare on educated young people, they had the greatest incentives to develop good standards and accountability—and did so.

Student Incentives

Similarly, student incentives, particularly high standards, promote learning. The threat of grade retention, for example, can serve as an incentive for greater effort, although intensive remediation seems necessary. An example is Chicago's Summer Bridge program, which gave parents and students the choice of grade retention or passing an intensive, focused summer course. Depending on the grade level and subject, grade-equivalent increases in reading and mathematics scores over the short summer session ranged from one half to a full year. The gains were extraordinarily effective, time-efficient, and cost-effective, and they were sustained in subsequent school years. The program, moreover, most benefited the initially lowest achieving students (Betts & Costrell, 2001, pp. 31–45; see also Bishop, 1996, on external examinations as discussed subsequently).

Tough grading standards and required homework also benefit learning. Requiring high-quality work for a given assigned grade generally raises achievement, particularly for high-achieving students who might not otherwise be sufficiently challenged. Corroborating control-

group research, analyses of national surveys show that the effects of the amount of homework teachers require each week are

> very strong, indicating that math homework is a more important determinant of gains in achievement than any of the standard measures of school quality, such as teacher education and experience or class size. (Betts & Costrell, 2001, pp. 33–34)

A previous section (on motivation) described the Dallas O'Donnell Foundation's $100 payment of high-school teachers and students for each college-accredited Advanced Placement examination passed, which raised the numbers of passing students twelvefold. Elementary-school children can benefit from encouragement, praise, feedback about accomplishments, and other nonmonetary reinforcement. Large-scale studies show that precise measurement of accomplishments and prescription of subsequent efforts multiplies such reinforcement effects. The commercial Accelerated Reader (AR) program, for example, assumes the well-documented but often unheeded idea that the more children read, the better they read. AR developers further assume that reading material that appeals to children and that appropriately challenges them promotes their reading ability most efficiently.

The AR developers have categorized some 30,000 children's books according to their reading difficulty and children's interests, such as arts, sports, and history. Since teachers lack personal knowledge or ready objective information to guide students' reading of books commonly held in school and neighborhood libraries, the AR computer program suggests to teachers the books best suited in difficulty and genre for each student. After a student has read a chosen book, the AR computer program assesses the student's comprehension and awards points for the degree of mastery and the difficulty of the book. Over a year, teachers and students can trace progress on a point system based on the volume and difficulty of books read. Statistical analyses of as many as 600,000 student records show Accelerated Reader's excellent results in promoting reading mastery (Topping & Paul, 1999).

External Examinations

The Cornell economist John Bishop intensively studied effects of curriculum-based external examination effects on learning. He analyzed surveys of the examination effects on learning of the (U.S.) Advanced Placement program, the New York State Regents, and U.S. state and Canadian provincial systems. He also analyzed examination effects on learning in the United States in comparison with effects in Asian and European nations. The examinations have the common elements of being externally composed and geared toward agreed-upon subject matter students are to learn within a nation, state, or province. Often given at the end of related courses,

they have substantial positive effects on learning (for a summary, see Bishop, 1996). Made publicly available, the examinations allow citizens, policymakers, educators, parents, and students to assess and compare achievement standings and progress.

The largest and most sophisticated international comparative analysis of national achievement yet conducted corroborates Bishop's and related findings (Woessmann, 2001).[26] Using data from 39 countries that participated in the Third International Mathematics and Science Study, a Kiel (Germany) Institute of World Economics study found that nations where students learned most employed external, curriculum-based examinations, and policymakers closely monitored the results.

How and why should such examinations yield striking effects? Though there are variations in their design, the examinations cover uniform subject matter in humanities, sciences, and other courses. Since the exams are graded by educators other than the students' own teachers, students have little incentive to challenge their teachers about course content and standards. Rather students and teachers work together toward the common goal of meeting examination standards. Because the exams and courses are uniform, teachers can concentrate not on what to teach but how to teach, and the students' subsequent teachers can depend on what students have been taught.

Accountability

In 1989, the National Governors' Association "Education Summit," with then President George Bush and business leaders, gave impetus to business-style accountability for schools. "Systemic reform," as recommended by summiteers, means aligning the chief parts of school systems with one another, specifically fitting state tests and curricula with state goals or standards and making exam results widely known.

Like the accountability of business boards and executives, school accountability requires simultaneous centralization and decentralization, centralization of standards at the state level and decentralization of operational responsibilities to the district or school level. State policymakers set goals and measure progress, but, unlike in the past, encourage local school districts and schools to plan and execute effective practices. State officials can set high targets for achievement or value-added learning gains and maintain more objectivity in evaluating the results than when they determine both goals and means, and without this division of labor, local districts might set easy-to-reach, unmeasurable, or obfuscated goals.

As discussed below, leading authorities on accountability contributed to a conference and book on the subject to assess the last decade's progress. As the editor pointed out, concern for

achievement is bipartisan, and surveys show that the public strongly supports objective testing, higher standards, and greater specificity about what students should learn (Ravitch, 2001, p. 4). Large-scale research on school accountability shows strong public recognition of the need for accountability and corroborates the expected positive learning effect.

Need for accountability. Reminiscent of the authors of the *Nation at Risk* report, representatives of the Business Roundtable and the National Alliance for Business have argued that standards-based reform and accountability are keys to the nation's future economic performance. Jobs requiring literacy skills, for example, will grow faster than all other occupations. More jobs also require skilled labor, particularly that involving computers.

Yet, as previously pointed out, U.S. high-school students lag behind those in other countries in essential subjects. An estimated 78% of our nation's institutions of higher learning offer remedial courses for first-year students unready for college work. It appears that about half of American firms provide training to make up for inadequate schooling, perhaps a considerable fraction of the estimated annual $55 billion budget for employee training. A U.S. Department of Labor study showed that illiteracy cost eight southern states $57.6 billion in lost productivity, substandard work, unrealized taxes, unemployment claims, and social problems (Goldberg & Traiman, 2001).

Effect of accountability. A decade ago, few states specified what students should know and be able to do, but 49 states now do so, and the number of states with adequate academic standards has increased. The more sustained and comprehensive the accountability system, moreover, the better states' learning progress appears. A study commissioned by the National Educational Goals Panel revealed the reasons that North Carolina and Texas made the largest gains on the National Assessment of Educational Progress:

- grade-by-grade standards with aligned curricula and textbooks,
- expectations that all students would meet the standards,
- statewide assessments linked to the standards,
- accountability for results with rewards and sanctions for performance,
- deregulation and increased flexibility in ways the standards can be met, and
- computerized feedback systems and achievement data for continuous improvement.[27]

Policy analysts have begun rating the states for both standards and accountability, which to be most effective, must presumably go together. Good standards are rigorous, clear, written in plain English, communicate what is expected of students, and can be assessed. Good

accountability systems are aligned with the standards and include school report cards, ratings of schools, rewards for successful schools, authority to reconstitute failing schools (for example, by replacing the staff), and the actual exercise of such legislated consequences. Only five states—Alabama, California, North Carolina, South Carolina, and Texas—have solid standards and strong accountability systems (Finn & Kanstoroom, 2001).

Employing standard economic principles, legislators and state school boards also are designing increasingly refined accountability systems and tying incentives to test results (Betts & Costrell, 2001). For example, states increasingly "disaggregate" test scores to be sure that various groups are well served. Texas, for instance, reports separate results for boys and girls, and for Whites, Blacks, and Hispanics. Similarly, the National Assessment of Educational Progress reports percentages of students that meet Advanced, Proficient, Basic, and Below Basic standards, which encourages improvement at all levels rather than on only a single standard that is too easy for some students, schools, and districts and too challenging for others. By a large margin, the U.S. Congress recently passed legislation that will extend features of the North Carolina and Texas accountability-reform principles to all 50 states.

Cost of tests and accountability. Though some educators have protested the costs of accountability systems, their costs are surprisingly small and represent a miniscule percentage of school budgets. The payments to commercial firms for standardized testing, standard setting, and accountability in year 2000 was $234 million, which was less than a tenth of a percent of K–12 school costs and amounted to $5.81 per American student. For the 25 states with available information, the total costs per student run between $1.79 and $34.02, higher on average than commercial costs alone, since some states develop their own tests, develop their own standards, and run their own accountability systems. Even so, the total costs are tiny, despite the public's, parents', and legislators' strong interest in accountability.

These costs, moreover, will undoubtedly decline in the longer run since they were estimated in the midst of states' development of accountability systems; after development and initial revision, much of the activity can be routinized at much lower costs. Even now, raising teacher compensation by 10% would cost 12.399% more than current accountability systems. Reducing class size by 10% (or about 2 students per class) would cost 8.81% more than current accountability in the nation (Hoxby, 2002).

Small Schools and Small Districts

In the half century through 1990, the number of U.S. school students rose from 25.4 million to 41.2 million. The number of districts, however, declined from 119,900 to 15,400, and

the number of schools declined from 247,100 to 81,700. During the period, accordingly, the average number of students per school rose from an average of 103 to 504, and the number of students per district rose from 214 to 2,683. The distribution of both schools and districts is positively skewed; there are a few huge ones and many relatively smaller ones concentrated in rural areas, particularly near the Canadian border.

The massive increases in school and district size took place despite research showing that large organizations tend to become departmentalized, impersonal, bureaucratic, inefficient, and lacking in individual and institutional accountability. Their goals tend to become diffuse, and they tend to be more subject to needs of their employees and special interests than to their clients.

Psychological effects. Perhaps for analogous reasons, the first large-scale study showed similar inefficiency of large districts and large schools in 38 states (Walberg & Walberg, 1994). The study showed no effect of per-student spending but significant effects of each state's average district and school size. Why? Consider Montana: Usually at the top of state achievement surveys, its many districts have as few as 100 to 200 students, so school board members may be able to speak insightfully about many of the individual faculty and students in their single school. In New York City, board members might be stumped to name more than 50 of the roughly 900 schools of the 1.1 million students. If something goes wrong in a Montana school, a parent might ask a school board member at a grocery store to look into it. Can this be imagined New York City?

Teachers in the tiny Montana district, to continue the example, would be likely to know not only the students but also their siblings and other relatives. Parents, teachers, and school board members can readily communicate. Being small, neither the district nor the school would multiply programs and courses excessively, but they would stick to fundamental subjects in a core curriculum taken by most students, such as English, mathematics and science, civics, history, and geography, foreign language, and art and music. Such a curriculum has been shown to be conducive to high achievement and advantaged university admission. In the 1990s, several dozen statistically controlled studies showed the achievement advantages of small schools, which tend to be concentrated in small districts (for a comprehensive review, see Lee, Bryk, & Smith, 1993).

Economic effects. Citizens in smaller districts, moreover, are likely to be best informed about local conditions and their own desires; they are rationally motivated to avoid inefficient spending and ineffective programs for the children in their communities. Small districts confer local control conducive to achievement and other school outcomes. Such local control gives all community residents—not just parents—an incentive to monitor local public schools and ensure

effectiveness and efficiency, that is, better outcomes and value for money (Borland & Howsen, 1992; Hoxby, 1994).[28] "Capitalization," rising property values to homeowners, results from better school outcomes and value, as realtors often stress and as a few dozen economic studies show (Yinger, Bloom, Borch-Supan, & Ladd, 1988; all but one of 28 studies in their detailed review showed this "capitalization" effect).

Thus, the psychological and economic advantages of small schools and small districts make them more effective and efficient. Of course, after much painful district consolidation and huge capital investments in large school buildings, the clock cannot easily be turned back. But it can be recommended that districts think twice about further consolidation and building ever-larger schools. More radically, legislators have been considering the breakup of large districts such as Los Angeles and New York into completely freestanding units with separate boards and superintendents. Citizens in parts of Los Angeles are pressuring legislators to allow secession.

Big urban districts such as Chicago and New York fostered "schools within schools" that are attempts to recover the intimacy, accountability, effectiveness, and efficiency of smaller schools of yesteryear, though it remains to be seen if such values can be recaptured in big buildings. In any case, special forms of accountability seem necessary in large districts to insure effective, efficient schools that are satisfying to parents.

Choice

No scholarly analyst argues that private schools or voucher programs chosen by parents deter learning. The question is whether or not schools of choice promote learning both among their own and other students in the community. Six scholarly reviews conducted at Harvard University, the RAND Corporation, the Urban Institute, and by New York scholars reach the same overall conclusion about the positive achievement effect of choice.

Private Vouchers

Paul Peterson of Harvard's Kennedy School synthesized his extensive findings from (true experimental) evaluations of private voucher programs[29] in Milwaukee, Cleveland, New York, and Washington as follows:

> According to the test score results, African American students from low-income families who switch from public to a private school do considerably better after two years than students who do not receive a voucher opportunity. However, students from other ethnic backgrounds seem to learn after two years as much but no more in private schools than their public school counterparts. (Peterson, 2001, pp. 274–275)

Referring to Peterson's experiments, the RAND review concluded:

> Small-scale, experimental privately funded voucher programs targeted to low-income students suggest a possible (but as yet uncertain) modest achievement benefit for African-American students after one to two years in voucher schools (as compared with local public schools). (Gill, Timpane, Ross, & Brewer, 2001, pp. xiv–xv)[30]

The Urban Institute summarized private voucher research as follows:

> The results of this research also showed that attending a private school was beneficial, but only for African American students. On average African Americans who received vouchers scored .17 standard deviations higher on the combined test scores than African Americans in the control group. After two years they scored .33 standard deviations higher than their counterparts in the control group. (Goldhaber, 2001, p. 64)

These gains, if sustained, would be very substantial. They would eliminate the often-observed (one standard deviation) Black–White gap in 6 years.

Why should private schools apparently benefit African American students and not others? Because they favor choice more than other groups, more African American families have elected to participate. Since their numbers are greater, a given effect is statistically more likely to be significant. For reasons explained in a previous section, it is also possible that big city school systems where African Americans are concentrated treat them bureaucratically and indifferently, whereas private schools must please or lose their "customers."

General Effects of Choice

In a more comprehensive review, two political scientists have considered choice more broadly including choice within public systems. They point out, as earlier reviews do not, that a

> combination of evidence is important in a domain in which economists, political scientists, sociologists, educational scholars, and others often read work only in their own disciplines. Moreover, while other researchers have reviewed various pieces of the choice literature, most are focused on only one aspect or type of choice. Here a broader analysis is sought. (Teske & Schneider, 2001, p. 609)

They conclude:

> While not all of these studies conclude that choice enhances performance, it is significant to note that the best ones do, and that [we] did not find any study that documents significantly lower performance in choice schools. (p. 619)

Why is choice popular and reasonable?

> Consensus results show that parents are more satisfied with choice, that they report using academic preferences to make choices, and that they tend to be more involved with their child's education as a consequence of choice. (p. 609)

Two economists have analyzed the competitive effects of choice on education outcomes revealed by over 35 studies. Their review concerned—not charters or vouchers—but naturally occurring traditional competition within geographical areas such as cities and metropolitan areas. The studies typically analyzed the parentages of students enrolled in private schools and the degree of district monopoly, for example, the presence of many small districts as opposed to one district within a county. They conclude:

> A sizable majority of these studies report beneficial effects of competition across all outcomes, with many reporting statistically significant coefficients." The positive benefits included increased academic test scores, graduation rates, efficiency (outcomes per unit of per-student spending), teacher salaries, housing prices, and adult wages. (Belfield & Levin, 2001, p. 1).

In a large-scale study too recent to be reviewed by scholars cited above, Jay Greene has developed a *2001 Education Freedom Index* by weighing the amount of (a) charter-school choice, (b) subsidized private-school choice, (c) home-schooling choice, and (d) public-school choice in an overall index for each state participating in the National Assessment of Educational Progress. On the index, Arizona had the greatest amount of overall choice; Hawaii, with only one school board for the whole state, the least. Greene found that, controlled for median household income, per-pupil spending, and the percentage of ethnic minorities in each state, achievement test scores and (value-added) score gains were significantly associated with the amount of total weighted choice in the state (Greene, 2002b).[31]

Productivity Deterrents

Many prevalent and incipient education policies and practices are ineffective or inefficient or both, making them unproductive. Some take time away from what works consistently and well, and some are costly, disruptive, distracting, and have unanticipated harmful consequences. Their prevalence helps explain why American students fall behind, despite high and substantially rising expenditures. For the sake of increasing educational productivity, it is worth considering them, so that they may be generally avoided or mitigated. Discussed in this section, they range from applications of pseudoscientific psychology to categorical federal education policies. First discussed, however, are theories that underlie unproductive practices.

Slack Professional Standards and Unvalidated Theories

Some influential education theorists and educators oppose accountability, standards, testing, and the evidence-based learning principles discussed above, most of which comport with what the legislators, public, parents, and students themselves expect from schools. A Public Agenda national survey of high-school students, for example, showed that three fourths believe stiffer examinations and graduation requirements would make students pay more attention to their studies. Three fourths said schools should promote only students who master the material. Almost two thirds reported they could do much better in school if they tried. Nearly 80% said students would learn more if schools made sure they were on time and did their homework. More than 70% said schools should require after-school classes for those earning Ds and Fs (Bradley, 1997; Johnson & Farkas, 1997).

In these respects, many educators differ sharply from students and the public. Interviews with a national representative sample of elementary- and secondary-school educators and students revealed the following percentages agreeing with the degree of academic challenge in their schools (Harris Interactive, 2001):

Statement	Principals	Teachers	Students
School has high academic standards	71	60	38
Classes are challenging	67	48	23
Teachers have high expectations of students	56	39	25

The apparent slackness of many practicing educators may derive from views prevalent in schools of education they have attended. A 1997 Public Agenda survey of education professors showed that 64% thought schools should avoid competition. More favored giving grades for team efforts than favored grading individual accomplishments. Only 12% thought it essential for teachers to expect students to be neat, on time, and polite, compared to 88% of the public. Only about a fifth agreed with the public that teachers should stress correct spelling, grammar, and punctuation. Only 37% thought it essential for teachers to learn how to maintain an orderly classroom.

Teacher educators also differ from employers and other professions on measuring standards or even employing them at all. Employers use standardized multiple-choice examinations for hiring. So do selective colleges and graduate and professional schools for admission decisions. Such examinations are required in law, medicine, and other fields for licensing, because they are objective, efficient, and reliable. In the case of teachers, it would seem

that knowledge of the subject matter is prerequisite to teaching it. Indeed, indicators of academic mastery, including objective examination results and completion of rigorous courses, appear influential on their students' achievement, at least in technical fields such as mathematics. Yet 78% of teacher educators wanted less reliance on objective examinations (Farkas & Johnson, 1997).

Nearly two thirds of teacher educators admitted that their programs often fail to prepare candidates for teaching in the real world, and only 4% reported that their programs typically dismiss students found unsuitable for teaching. Thus, even starting with their undergraduate education, many prospective educators are exposed to disparaging views of standards, incentives, and individual accomplishments. These views have often led to

- the notion that "authentic learning" only arises from "intrinsic motivation" in which student preferences rather than curriculum and course requirements dominate the choice of what and how to learn,
- an indifference or hostility to specifying objectives and measuring results,
- a holding that children cannot learn until the "teachable moment" or until the "developmentally appropriate" time,
- a devaluing of knowledge (since "you can always look it up"),
- an insistence that students should discover or "construct" their own understanding rather than being taught, and
- the idea that comprehension must be "socially constructed" in peer groups rather than individually acquired.

These views may be characterized as constructivism rather than instructivism, in which the teacher imparts knowledge and skills, and learner-centered rather than learning-centered, in which the teacher employs well-defined goals, definite subject matter, and explicit assessment of student progress.[32]

An example of how these views take form in classrooms is "discovery learning," in which students are to discover or rediscover scientific or other principles instead of being directly taught. Discovery takes precious time from comprehension and practice of principles. According to two eminent cognitive psychologists and a Nobel laureate in economics, the evidentiary basis of such discovery theory consists largely of proponents who cite one another's values and opinions. In their opposition to direct teaching, education theorists criticize the practice of knowledge and skills as

> "drill and kill," as if this pejorative slogan provided empirical evaluation. . . . Nothing flies more in the face of the last twenty years of research than the assertion that practice is

> bad. All evidence, from the laboratory and from extensive case studies of professionals, indicates that real competence only comes with extensive practice. By denying the critical role of practice, one is denying children the very thing they need to achieve competence. (Anderson, Reder, & Simon, 1998, p. 241)

Although current education theory is ill-informed about scientific findings, it often draws faddishly on "pop" psychology.

Brain-Based Learning and Other Exotic Techniques

K–12 and other education sectors have been subject to many pseudoscientific fads for which there is usually initial enthusiasm but paltry or analogous evidence. At the request of the U.S. Army, for example, the National Academy of Sciences evaluated exotic techniques and "shortcuts" for learning and enhancing performance described in "pop" psychology and claimed to be successfully employed for "super-learning." Little or no evidence, however, was found for the efficacy of learning during sleep, mental practice of motor skills, "integration" of left and right brain hemispheres, biofeedback, and such parapsychological techniques as extrasensory perception, mental telepathy, and mind-over-matter exercises. Nor could creditable evidence be found for "neurolinguistic programming," in which instructors identify students' mode of learning and mimic the students' behavior as they teach (Druckman & Swets, 1994). Even so, "brain-based learning" is gathering momentum in education circles.

Widespread, Unsubstantiated Programs

School-board members and most educators lack education and experience in accountability, evaluation, and methods of psychometrics and statistics that would enable them to choose effective, efficient programs and weed out others. Though these tasks should be central to leaders aiming to measure, evaluate, and improve learning, they are neglected. Consequently, popular programs are often chosen by fad and reputation rather than by a careful review of evidence of their results and costs. In this section, two widespread programs illustrate such choices.

Reading Recovery. Begun in 1976 in New Zealand, Reading Recovery was implemented in 40 states within 8 years. Reading Recovery was founded on assumption that reading is essential for school success and on the value that children should be "recovered" as they begin to fail. The program also assumed that if children repeat their errors less often, they will have less to relearn and will increase their self-esteem as readers. They would presumably be less susceptible

to the Matthew effects of falling further and further behind their classmates. Who would take issue with these assumptions and values? But does the program work?

Program advocates reported research that appeared to support the effectiveness of Reading Recovery. But an independent review in the leading reading research journal (Shanahan & Barr, 1995) pointed out that children who were repeatedly absent or who were not making progress and were transferred out of the program were untested. This yields a fallacious "selection effect," because the apparent program success is attributable to an unknown degree simply to purging unsuccessful learners out of the Reading Recovery program but not out of the control groups. The apparent recovery effect, moreover, diminished after first grade; this result contradicted the claim of avoiding the Matthew effects.

Further, because Reading Recovery teachers tutor a single student pulled out of regular classes for long periods, Reading Recovery students lose time in regular instruction. The annual per-student cost for the program tutoring alone, moreover, is at least three quarters that of a full program for other students in all subjects all day for the school year. As discussed in a previous section, phonics, phonological awareness, and repeated oral reading instruction have substantial effects and can be conducted cheaply, routinely, and effectively with a whole class.

Success for All. The Success for All (SFA) program also makes a big claim in its title and is aimed at teaching reading to early-grade children in poverty. Educational administrators and several scholars have praised SFA; Congress issued legislation and about $150 million of initial funding favoring its adoption; and, in school-finance litigation, a New Jersey judge ordered its use in failing schools. The number of SFA schools rapidly expanded, and the program seemed just the breakthrough needed.

Yet the first independent review of the evidence showed both obvious and subtle potential conflicts of interests among government funders, foundations, SFA program developers, and evaluators (Greenberg & Walberg, 1998). Nearly every study by SFA developers and their associates showed positive effects, but independent studies showed no effects. Contrary to SFA claims, average SFA third graders were not up to grade level in Baltimore, where the program originated; by fifth grade, they were 2.5 years behind. A later review of subsequently accumulated evidence revealed the same research flaws and program failures as well as obvious conflicts of interests among program developers and evaluators (Pogrow, 2000).

In SFA self-evaluations, the compared groups clearly differed, most obviously in that SFA schools were given substantially more funds, materials, and services than control schools. SFA, moreover, requires that 80% of the teachers vote secretly to adopt SFA. Such schools are

unusual in attaining such agreement and no doubt in determination. Such schools may have inspired leaders or zealous teachers. Even so, independent evaluations showed SFA does the same or less well, even with the advantages of staff consensus and extra resources.

SFA self-evaluators also biased their research in other ways. They concentrated on reading tests to compare the SFA program, which heavily concentrates time and energy on reading, with ordinary school programs aimed at improving achievement in all the standard subjects. A more comprehensive evaluation would employ tests in all subjects to determine whether SFA sacrifices broad achievement in mathematics and other subjects for narrow reading skills. SFA self-evaluators also employed unusual tests that favor SFA and that were administered to individual children and subjectively scored by SFA's staff. A more objective evaluation would employ standard tests administered by educators entirely independent of SFA.

Federal Categorical Programs

Because the programs are aimed at particular groups, the programs require categorizing and often segregating students (a) in poverty, (b) in need of special-education services, and (c) in need of special English language instruction. The federal government provides only about 6% of K–12 school funding, but accompanying federal dollars are rules, regulations, and program requirements, which distract school boards and educators from their primary objective of increasing learning. They also impose large costs for nonteaching staff, which amounted in the mid-1990s to 25.1% of current expenditures for primary and secondary schools in the U.S., in contrast to the 13.4% average of other advanced countries (Organisation for Economic Co-Operation and Development, 1995, p. 103).

In what has been called the "colonization of local and state agencies," federal rules require up to three-quarters of state education department staff for federal compliance rather than productivity improvement. Staff preoccupation with bureaucratic means rather than educational ends frustrates state boards and superintendents aiming to improve achievement. Especially frustrated have been reform-driven superintendents in large cities where the programs are concentrated and account for a large part of their staff and budgets (Hill, 2000; see in particular pp. 25–26).

Despite burdensome costs and administrative problems, categorical programs, as discussed in this section, have done little good. Neither their principles nor operational practices appeared to contribute to the learning of needy students. Their failings may go a long way toward explaining the problems of poor and minority students in cities where they and federal programs are concentrated.

The Title I program for students in poverty. The federal government has spent about $125 billion on Title I and now allocates about $8 billion annually. The program was to have reduced the gap between middle-class students, often Whites in suburbs on the one hand, and on the other, poor students, often African Americans and Hispanics in cities and rural areas. Congressionally mandated and independent studies show that the Title I program, even after 3 decades, has not diminished, much less eliminated, the poverty gap. As Figure 3 indicates, even recent reforms of Title 1 have little to show: The poverty gap has remained essentially the same. Why?

A synoptic evaluation (Farkas & Hall, 2000)[33] of Title I points to the lack of evidenced-based practices. Recommended for preschoolers are two of the large effects in Table 3, phonological awareness and phonics, as well as parent programs for academically stimulating their children and skills-based summer programs between kindergarten and first grade. Recommended for 1st through 12th grade are careful, continuous assessment and remediation; schoolwide behavioral management to avoid wasting instructional time on discipline problems; and research-based teaching methods—all consonant with the idea of extending the quantity and quality of instruction as discussed in previous sections.

Special education. The other huge federal categorical program, special education, is comparable to Title I in federal spending, ineffectiveness, and inefficiency. It includes about a tenth of American children and currently costs $7.4 billion in federal money and an (imprecisely) estimated $35 billion to $60 billion, counting state and local contributions. By some estimates, 40% of the new money flowing to K–12 education in recent years went into special education, partly because of administrative complications and expense, special testing, and smaller classes, but not counting big litigation costs over placement of students in special programs (Finn, Rotherhan, & Hokanson, 2001, p. v; see particularly the Horn & Tynan chapter, 2001b; for effect size estimates, see the Lyon et al. chapter, 2001, especially p. 272). About 90% of special students were classified in such categories of mild disorders as "learning disabled," "language impairment," "mild retardation," "emotional disturbance," and "behavioral disorder" rather than the long-established, scientifically creditable categories of blindness and deafness, which can be reliably diagnosed.

Perhaps because it brought more funds into districts, the numbers of mildly disabled students classified, particularly "learning-disabled students,"[34] rose sharply. A National Academy of Sciences Panel (Heller, Holtzman, & Messick, 1982), however, found the classification and placement of students in special education ineffective and discriminatory against minorities. The

Panel recommended that students should be segregated from others only if (a) they can be reliably classified, and (b) segregated placement shows superior learning results.

Much research, however, shows that the classification systems for placing "mildly handicapped" students in special programs are deeply flawed; specialists in the field have come to little consensus about taxonomic, psychometric, and diagnostic procedures. By various classification systems, for example, as many as 80% of all American K–12 students can be classified as "learning disabled" (Reschly, 1987; Ysseldyke, 1987). Once students are placed, moreover, they are unlikely to return to regular classes without a planned and explicit exit strategy, which is rarely offered. As recommended by eminent psychologists, a better solution is to avoid psychological testing and identify children in need of assistance only by their level of achievement (Sternberg & Grigorenko, 2001). This concentrates educators' attention on what they need to know rather than on ill-founded psychological deficits. It enables children to stay with their peers and avoids psychological stigma that reduces their own and others' expectations of their potential.

Special-education programs, moreover, often lead to discrimination, unfairness, and litigation. Special educators, for example, often demeaningly classify big-city poor and minority children as "mildly retarded" and may expect less of them. But middle-class parents have learned to "game the system." Their children are more often labeled "learning disabled," a less stigmatic term. In affluent Greenwich, Connecticut, educators classify about one in three high-school students as learning disabled. Such students may get special tutors, note takers, laptop computers, and extra time on tests, including those for college admission. A national poll revealed that 65% of parents said that the extra attention paid to disabled students came at the expense of their own children (Horn & Tynan, 2001a, 2001b), a view more likely to be associated with alienation and withdrawal than with parent involvement.

Aside from classification and fairness problems, evidence shows the unproductive effects of special education. At twice the cost of regular education, special-education placement has an estimated effect of about .07, a small fraction of the effect of methods that can be routinely applied in regular classrooms, such as graded homework, which is about 10 times larger, and frequent testing, which is 7 times larger (see Table 3). Since reading mastery is often the core difficulty of learning-disabled and other mildly disabled students, the policies and practices recommended for children in poverty, such as early intervention and effective methods of teaching reading (see previous section), are likely to be much more effective and cost-effective than special-education placement (Lyon et al. 2001, especially pp. 272, 276–278). Syntheses of special-education research, moreover, suggest that methods that are successful in regular

education like direct instruction and diagnostic-prescriptive methods are also successful with special-education students (Gersten, Schiller, & Vaughn, 2000; Kavale & Forness, 1985; Swanson, 1999). If so, why classify? Why segregate? Who benefits?

The Present Teaching Force

Maintaining certification as the criterion for employment and reemployment and graduate credits and experience as the basis of compensation may mean that unproductive teachers are paid just as much as their colleagues who best promote learning. These policies offer no incentives for improvement. Why should even the best teachers work hard and long when their compensation will be the same as the worst performers? Why not put their energies and talents into moonlighting, travel, or their families?

A national survey of 853 public school superintendents and 909 principals corroborates such concerns (Farkas, Johnson, Duffett, Foleno, & Foley, 2001). Large majorities of superintendents (76%) and principals (67%) said they need more autonomy to reward outstanding teachers. Almost the same percentages said they need more autonomy to remove ineffective teachers. Nearly all superintendents (96%) and principals (95%) said making it much easier to remove bad teachers—even those with tenure—would be somewhat or very effective.

Teacher selection and compensation. As explained in the opening section, U.S. students fall behind those in other countries the longer they are in school, despite the nearly highest per-student spending in this country. Summaries of research by economists, moreover, find little or no consistent influence of spending on learning in studies of U.S. samples (Hanushek, 1986). Though noneconomists in schools of education have disputed these findings (Greenwald, Hedges, & Laine, 1996), many policymakers might agree that what money buys is more than the total amount spent. The two main determinants of spending are teacher compensation and class sizes, both discussed here. Do they buy higher achievement?

Public-school teachers' salaries have long been chiefly determined by whether they are certified, their years of teaching experience, and their degree level, commonly a bachelor's or master's. Despite thousands of doctoral dissertations in education written each year, little solid evidence shows these salary determinants promote student learning. In fact, studies by labor economists suggest that verbal ability, knowledge of the subject matter, and graduation from a selective college are at least as important as the usual salary determinants.

To investigate the contribution of the contending measurable teacher attributes to student learning, the following equation could be estimated:

Student achievement = Student input + teacher experience + teacher education + teacher verbal ability + teacher pedagogical knowledge + teacher subject-matter knowledge + teacher certification

In this equation, student input is indexed by previous achievement and demographic characteristics such as poverty, verbal ability is indexed by verbal tests or college selectivity or reputation, knowledge is measured by tests or course completion in the subject matter such as science, and a weight is estimated for each factor.

No study, however, comes close to this equation. The consequence of flawed studies is misleading implications for teacher certification, hiring, retention, and compensation. For example, simply showing that the students whose teachers have a master's degree achieve better may reflect not the learning advantage of a master's but the fact that teachers who are more experienced are more likely to have master's degrees and vice versa. Similarly, failing to take previous student achievement and demographics into consideration may mean that an apparent connection between experience and achievement is attributable to teachers transferring to middle-class schools that achieve well in any case. Estimating the equation above would test these and other causal possibilities.

A limited standard of proof calls for including prior achievement and student demographics in testing the possible influences of the other factors one or two at a time. A recent search uncovered only 18 such studies, nearly all by economists. These studies suggested that college selectivity, verbal test scores, and, only for high-school students in mathematics, subject-matter knowledge contribute to student learning (Wayne & Youngs, 2001).

Examining studies that control only for student input trades a larger pool of studies for research rigor. Since teacher effects and costs are so critical, even less certain evidence is worth considering. Such research corroborates the importance of verbal facility and college selectivity, but suggests that only 3% of teachers' contribution to student learning is attributable to teacher experience and graduate degree attained. Few studies show significant positive effects of experience and education, and some studies show significant negative effects. Certified teachers apparently perform no better those who are uncertified (Goldhaber, 2002).

Even though teaching comprises about half the total schooling costs, research provides no support for traditional and current policies of certifying, selecting, and compensating teachers. Arbitrarily excluding candidates on weakly predictive or nonpredictive criteria is arbitrary; in an apparently tight labor market, this longstanding policy unduly excludes large numbers of younger and older people who are as likely to teach as well as other candidates and the present labor force. In fact, Teach for America demonstrates that very recent graduates of elite colleges,

knowledgeable of the subjects they teach, but with no experience and little pedagogical training, are highly regarded by their principals and that they also induce greater achievement than other teachers (Kopp, 2001; Raymond & Fletcher, 2002).

Class size and student/teacher ratio. Student–teacher ratios fell from about 27 in 1955 to 17 in 1997 (U.S. Department of Education, 1997, Table 64), which accounts for much of the substantial rise in per-student expenditures. More teachers, however, do not necessarily make for smaller classes, since they may perform administrative and special duties, especially in large cities with substantial federal programs, which may require much bureaucracy. In any case, as pointed out in the opening section, student achievement remained stagnant, despite the sizable investment in more teachers per student.

The first meta-analysis of education research on class-size effects on achievement suggested a small beneficial effect of class-size reductions. The biggest apparent effects were in reductions below class sizes of 10; classes between 15 and 35 students differed very little in achievement. Few studies had been made of classes between 8 and 15, because classes in this range were rare and prohibitive in cost. In any case, the overall effect of class-size reduction appeared to be much smaller than the use of effective teaching methods (Glass & Smith, 1979). Even a small effect was disputed. Large-scale studies, mostly by economists, showed no consistent effect of class-size reductions (Betts, 1995; Hanushek, 1999a).[35]

A much-noted Tennessee experiment seemed to show an effect of reduced class sizes (Finn & Achilles, 1990), even though a single study may not outweigh the inconsistent results of many other studies. Even at face value, moreover, the Tennessee study showed a very small effect, limited to kindergarten. Continuing exposure to smaller classes in first through third grade showed no advantage, and returning students to normal-sized classes in fourth through sixth grades showed no harm. So reduced class size apparently only benefited kindergartners, and changes in class sizes did not affect achievement in the six later grades (Hanushek, 1999b).

In addition, the apparent effect was not of class-size reduction alone but also of accompanying monetary incentives for increased student achievement. Thus, the apparent small transient effect may be attributable to smaller class sizes, monetary incentives, or a combination of these factors.

A more recent large-scale natural experiment on all Connecticut elementary schools overcomes limitations of the Tennessee research. It is perhaps the most comprehensive study ever made of the class-size question, because it measured the effects of natural changes in class sizes from 10 to 30 students over 2 decades. It showed no class-size reduction overall, nor any at the

upper or lower range of class-size reduction, nor in the earlier or later grades, nor for disadvantaged or middle class students (Hoxby, 1998).

What would happen if a state concentrated resources on reducing class sizes? California policymakers did just this at a cost of about $5 billion per year from 1996 through 2001. About two thirds of California school districts took money from libraries, art, music, and maintenance to reduce class sizes in the first three grades. After 3 years, evaluators could infer no achievement effect of class-size reduction. As they concluded, "There is no clear relationship between changes in the amount of exposure to CSR [class-size reduction] and changes in the average level of achievement. Increased exposure is not associated with greater gains in achievement" (Stecher & Bohrnstedt, 2001, p. 90).[36]

In view of definitively inconsistent research and California's experience, further class-size reductions seem unpromising. Such reductions, moreover, have been exceedingly costly. They are even more costly today, since student/teacher ratios have already been cut massively in recent decades. Reducing class size, for example, by a single student from 15 to 14 incurs more than twice the teaching costs of a single-student reduction from 35 to 34, even aside from the costs of new classrooms.

Conclusion

Syntheses of experimental and quasi-experimental classroom studies of instructional methods and large-scale econometric studies reveal policies and practices that work well and cost relatively little. Other policies and practices, even though prevalent in American schools, are costly, but little evidence suggests their efficacy. Though more research would yield better estimates and resolve some uncertainties, the present body of knowledge about effects and costs suggests how American schools can be made more productive.

Notes

1. Professor Hoxby (2001) concludes as follows: "Consider the simplest productivity calculation, achievement per dollar, without any attempt to control for student characteristics. Such a calculation (which I describe in detail below) suggests that average public school productivity was about 65 percent higher in 1970–71 than in 1998–99. [If we] were simply to restore school productivity to its 1970–71 level, then the average student in the United States would be scoring at an advanced level where fewer than ten percent of students now score" (p. 2).

2. These theories are discussed in a subsequent section.

3. For several reasons, economists have often led this work. They have generations of experience in inferring causality from nonexperimental data. Acutely conscious of priced and nonpriced costs and benefits, they have a long history of influence on legislators, jurists, and executives. Perhaps they may be more influential on policy since their professional incentive is to answer the policy question, what works best given a set of goals, rather than confusing goals and means or advocating values.

4. A recent synthesis of many estimates from comparative studies showed that Asian students who typically top the international achievement charts spend far more time studying than American students, since they have more school days during the year, usually attend private after-school tutoring schools, and do lots of homework for both schools. On average, Chinese students spend about double the American study time, and Korean about 73% more. Many Asian immigrant families continue extensive study, which undoubtedly accounts in part for their academic success. See Paik, Wang, and Walberg (2002). Fortunately, in recent years American educators have more often employed preschools, all-day kindergarten, extended-day schooling, and summer school, especially for students at risk of failure.

5. Subsequent larger collections of research syntheses are discussed elsewhere, including Fraser, Walberg, Welch, and Hattie (1987). The learning influences reported in Table 1 are correlations, which are roughly proportional to multivariate regression weights. In 23 studies of about 250,000 students in 13 countries, 89% of 341 mulitivariate regression weights controlling for the these and other factors were in the expected positive direction. See Paschal and Starhia (1989).

6. It may be useful for economists to think of these five factors as having Cobb-Douglas-like relations with achievement, namely, multiplicativity and diminishing returns. These imply that each factor is necessary, but each addition to its amount or intensity results in smaller and smaller additional amounts of learning. There may also be inflection points. Too much study, for example, may be counterproductive; too much motivation may become anxiety.

7. Despite the well-known 1983 warning of the National Commission for Excellence in Education's report, *A Nation at Risk*, the U.S. school year remains at about 180 days in contrast to 190 to 220 in most of Western Europe and up to 260 days, including Saturdays, in the Orient.

8. They spend only about 13% of their *waking* hours in school, assuming 9 hours per day of sleep, which is, of course, longer in the early years and shorter in late adolescence.

9. Though in the U.S. there are more poor Whites than poor African Americans and poor Hispanics, the rates of poverty are higher among the minority groups, which has led to increased policy research in the last few decades. Far more research, however, has been carried out on African American than on Hispanic students.

10. In contrast to such learning-centered approaches that directly teach children language and numbers are such learner-centered "Theoretical Principles of Child Development and Learning": "Children learn best when their physical needs are met and they feel psychologically safe and secure. Children construct knowledge. Children learn through social interaction with adults and other children. Children learn through play. Children's interests and 'need to know' motivate learning. Human development and learning are characterized by individual variation" (Bredekamp, Knuth, Kunesh, & Shulman, 1992, p. 1).

11. Unvalidated education theory dominates many preschool and kindergarten programs, which is probably the major reason why dozens of evaluations have shown only transient effects; see subsequent section on this subject.

12. The largest synthesis thus far drew upon three sources: (a) 134 meta-analyses of 7,827 studies of about 10 million students; (b) content analysis of 86 reviews (leading authorities' summaries of and commentaries on research), 44 research handbook chapters, 18 other chapters, 20 government and commissioned reports, and 11 journal articles containing 3,700 references; and (c) ratings of the learning effectiveness of educational conditions and methods by 34 authorities. See Wang, Haertel, and Walberg (1993a).

13. Technically but simply, an effect size or "effect" may be thought of and actually often calculated as the difference between an experimental and control group means divided by the standard deviation of the control group. Its size may be estimated for various educational conditions and for student characteristics such as grade level and ethnic group, to test the consistency or robustness of the effect. For the same purpose, separate estimates may be for studies with varying degrees of experimental rigor. As might be expected, factors with big effects are usually robust.

14. Other publications for educators that contain many references and suggestions for implementing the best methods are Cawelti (2000); Maryland Department of Education (1990);

and uncopyrighted booklets in the series edited by the present author on teaching, tutoring, and other topics. These booklets are distributed by the United Nations Educational, Scientific, and Cultural Organization to 189 countries and also freely downloadable and republishable gratis on the Internet at www.ibe.org.

15. The effects reported are based on research in which investigators generally insured implementation. In practice, the various methods need to be reasonably well implemented to insure similar effects. On the other hand, long-term, well-managed implementation might result in larger effects.

16. These estimates are based on control-group rather than correlational research and mostly include experimental studies that randomly assign students or classes to groups or quasi-experiments that roughly equate the groups by measuring gains from a pretest, taken before enacting the method, to a posttest afterwards, or by employing statistical methods, such as covariance or regression, of determining the "net effect" or "value-added learning gain" after adjusting for initial achievement and other variables.

17. To my knowledge, the effects are the best but hardly infallible estimates. They depend, for example, on the quality of the underlying research; still, many studies by many scholars in many circumstances and pointing in the same direction compel more creditability than any single study, no matter how well planned and executed. The ways of synthesizing research have improved during the past quarter century, but the newer ways usually yield similar results. So older studies are not necessarily excluded here, since they may be the only ones available.

18. An interesting variant is "reciprocal teaching," in which paired students take turns leading dialogues on pertinent features of a text. By assuming the planning and executive control ordinarily exercised by teachers, students learn planning, structuring, and self-management similarly to the way tutors learn from teaching, and they learn why it is said that if you want to learn something well, teach it. Comprehension instruction encourages students to measure independently their progress toward explicit goals—a big lesson in life.

19. If both teacher- and self-instruction are considered, perhaps these elements apply to much of human learning, including such diverse fields as sports, ballet, chess, music, foreign languages, and the professions.

20. Most creditability can be given to the nine factors in Table 1, because the underlying research controlled for the complete set of factors in large national and international surveys and synthesized control-group research.

21. Many education theorists deny the role of incentives and hold that true or superior learning only takes place when it is intrinsically valuable to the student. But there is little

evidence that students are unaffected by long- and short-term external incentives. Even if they were unaffected, they need preparation for adult life, and most adult work, with the notable exceptions of that in bureaucracies and public schools, employs merit pay, that is, rewards results.

22. Opportunity to learn results comport with common sense. Students taught Japanese would undoubtedly obtain better reading and listening test scores than students not taught Japanese.

23. The studies typically employ huge international, national, and state samples and estimation of learning effects statistically controlled for spending, demographics, prior achievement, and other variables.

24. If people seem irrational in overindulging in smoking, drinking, drugs, carousing, or, in the case of students, sports, watching television, and mall walking instead of studying, observers might ask if their perceptions and calculations of benefits and costs are identical to those of the people they observe.

25. In his popular economics textbook, the Harvard professor N. Gregory Mankiw lists 10 fundamental principles of economics (1998, pp. 3–16), including "people face tradeoffs," "the cost of something is what you give up to get it," and "people respond to incentives." Inspired by the Nobel laureate Gary Becker, economists have ingeniously explained and demonstrated such principles in the fields of crime, immigration, fertility, marriage, government, addictions, social influence, and public health. See Tommasi and Ierulli (1995).

26. Employing delegation of means or division of labor, highly achieving nations allowed teachers considerable discretion over instructional methods.

27. The authors also attributed the gains in the two states to the intensity and stability of business support for the reforms but not to per-pupil spending, pupil/teacher ratios, proportion of teachers with advanced degrees, and average of teacher experience. See Grissmer and Flanagan, 1998, 2001.

28. Hoxby also points out that smaller districts can often accommodate families with varying preferences. A parent's voice and vote count far more in a district with an enrollment of 200 than in one of several thousand students on issues that may be important to them. If, for example, parents dislike emphasis on soccer over football, they can raise the issue and campaign for their views with likeminded others. Since, in many cases, relatively few people vote in school board elections, a few dozen votes of a resolute minority may be decisive.

29. Private vouchers are scholarships given by wealthy individuals and firms so that students, most often poor, minority students in big cities, can go to parochial and independent schools of their parents' choice. Since the programs are usually highly oversubscribed, students

are admitted by lottery, which makes for a randomized experiment with a "treated" and a control group that can be compared after a time.

30. In addition, the authors point out that vouchers promoted racial integration and that charter schools generally have racial-ethnic compositions similar to local public schools.

31. This study might be faulted for using states rather than schools or students as units of analysis, but it is consistent with other evidence and covers the whole of the U.S.

32. These views may be traced to the French Romanticism of Jean-Jacques Rousseau and to German Hegelianism, especially as passed on to educators by John Dewey. At an extreme today, university humanities departments' postmodernism and social constructivism promote the idea that knowledge is socially constructed by groups, and that one group's knowledge, even that of mathematicians and scientists, is as good as another's. Taking this view seriously undermines the authority of the teacher, science, and the canon of received knowledge, since student and savage views are just as authentic and valid as the views of those who have devoted their lives to acquiring and sharing special knowledge. Though such views would seem absurd, they are common among education theorists and those responsible for educating teachers; see survey results discussed above in this section. The philosophical underpinnings of such views are described in Hirsch (1996), Ravitch (2001), and Stone (1996).

33. Other problems with Title I include (a) measuring poverty, (b) the possible conflict of interest between educators who seek additional funds and families and students who may not wish to be identified as poor, and (c) concentrating Title I services on only poor children while trying to avoid the administrative problems and possible stereotyping harms of segregating them.

34. On the basis of extensive research syntheses, in their book two prominent University of California authorities in this field titled two chapters, "Learning Disability: A Pseudo-Science" and "Learning Disability: A Victim of Its Own History." See Kavale and Forness (1985).

35. In addition, Asian classes, which have as many as 60 students, usually rank at the top of international achievement surveys.

36. In an interview, the first author, Stecher, a senior social scientist at RAND Corporation, said, "It would be nice if we could give an unequivocal answer to the achievement question. Then people could decide if the benefits were worth the costs. Unfortunately we can't." (as cited in RAND Corporation, 2002). The authors have yet to reach a final conclusion and will continue to study the initiative, but their latest finding corroborates the pervasive elusiveness of the class-size effect.

Table 1

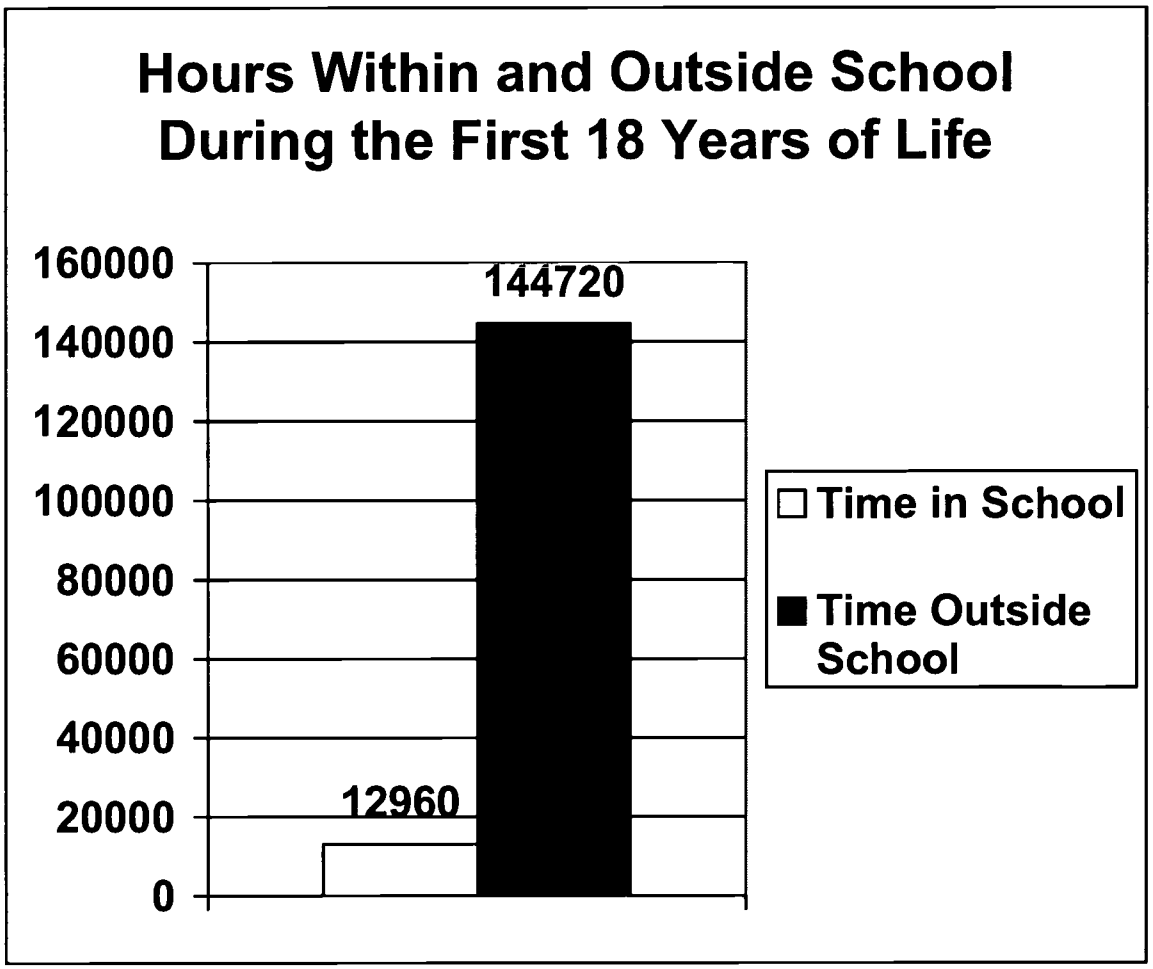
Nine Educational Productivity Factors

Factor	Proportional Learning Influence
A. Student Aptitude	
1. Ability or preferably prior achievement as measured by the standardized tests	.92
2. Development as indexed by chronological age or stage of maturation	.51
3. Motivation or self-concept as indicated by personality tests or the student's willingness to persevere intensively on learning tasks	.18
B. Instruction	
4. Amount of time students engage in learning	.47
5. Quality of the instructional experience, including method (psychological) and curricular (content) aspects	.18
C. Psychological Environments	
6. Morale or student perception of classroom social group	.47
7. Home environment or "curriculum of the home"	.36
8. Peer group outside school	.20
9. Minimal leisure-time mass media exposure, particularly television	.20

From "Synthesis of Educational Productivity Research," by B. J. Fraser, H. J. Walberg, W. W. Welch, and J. A. Hattie, 1987, *International Journal of Educational Research, 11* [entire issue].

Note. Estimates are calculated from data reported on p. 220. The indexes in the table are on the same scale as the effect sizes in Table 3 but are not necessarily pure, one-way causal effects.

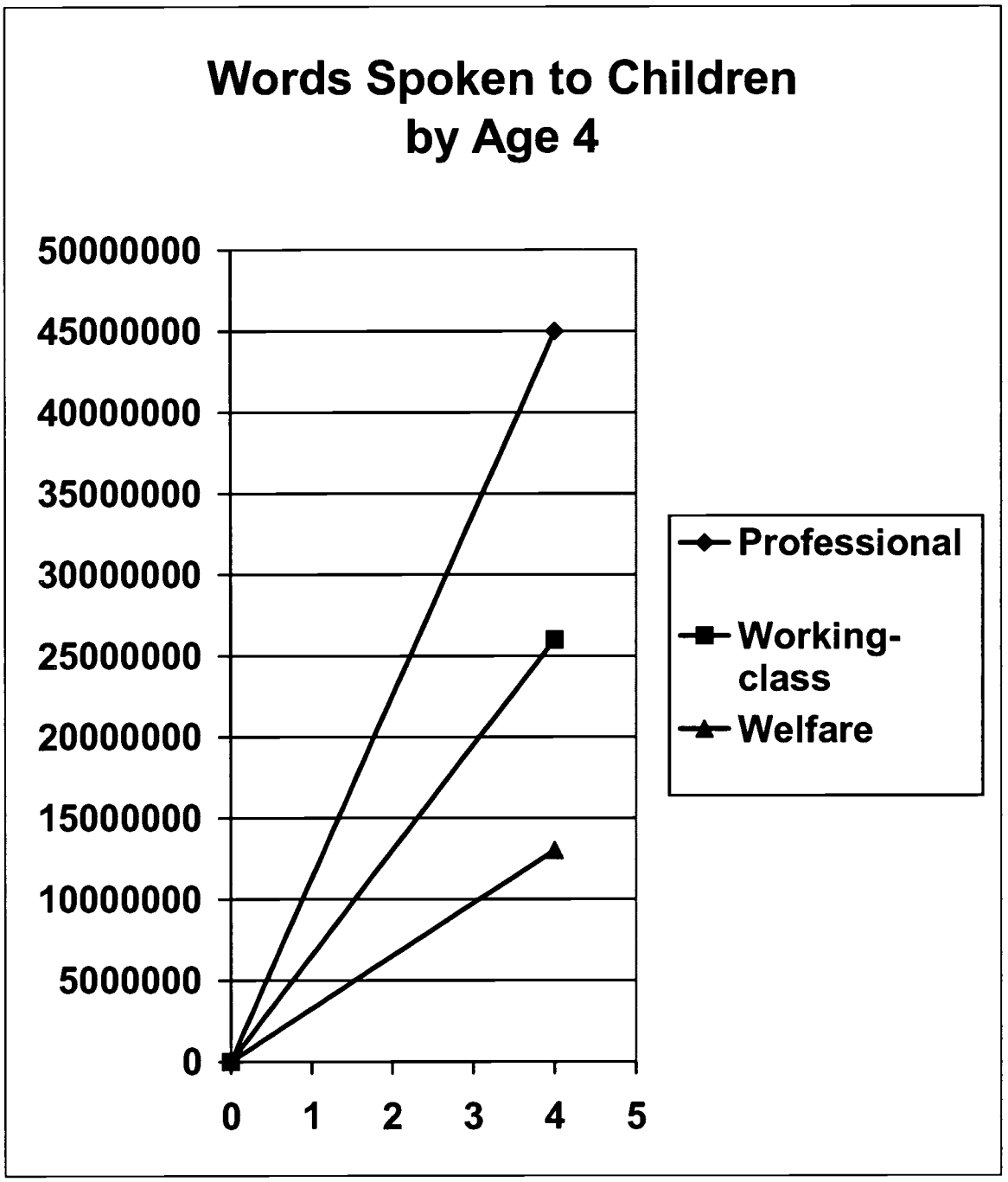
Figure 1



Basis of Calculating School Time in the First 18 Years of Life

Time or activity	Years	Days per year	Hours per day	Total hours	Percentage of time
Hours, birth through age 18	18	365	24	157,680	100.0
Sleep	18	365	9	59,130	37.5
School	12	180	6	12,960	8.2
Preschool years	6	365	24	52,560	33.3
Nonschool hours (total minus sleep, school, and preschool hours)				144,720	91.8

Figure 2



From *Meaningful Differences in the Everyday Experience of Young American Children*, by B. Hart and T. R. Risley, 1995, Baltimore: Paul H. Brooks, p. 200.

Table 2

Instructional Strategy Effects

Category	Average Effect
1. Identifying similarities and differences	1.61
2. Summarizing and note taking	1.00
3. Reinforcing effort and providing recognition	.80
4. Homework and practice	.77
5. Nonlinguistic representations (e.g., maps and other graphics)	.75
6. Cooperative learning	.73
7. Setting goals and providing feedback	.61
8. Generating and testing hypotheses	.61
9. Activating prior knowledge	.59

From *A New Era of School Reform: Going Where Research Takes Us*, by R. J. Marzano, 2000, Aurora, CO: Mid-Continent Research for Education and Learning, p. 63.

Note: The effects in this and other tables are generally ordered from largest to smallest as indicated by the effect sizes.

Table 3

Selected Effects of Quality of Instruction

General Methods	
Elements of Instruction	
Cues	1.25
Reinforcement	1.17
Corrective Feedback	.94
Engagement	.88
Mastery Learning	.73
Computer-Assisted Instruction	
For Early Elementary Students	1.05
For Handicapped Students	.66
Teaching	
Direct Instruction	.71
Comprehension Instruction	.55
Teaching Techniques	
Homework With Teacher Comments	.83
Graded Homework	.78
Frequent Testing	.49
Pretests	.48
Adjunct Questions	.40
Goal Setting	.40
Assigned Homework	.28
Explanatory Graphics	.75

Special Methods	
Reading Teaching	
Adaptive Speed Training	.95
Phonemic Awareness	.86
Repeated Oral Reading	.48
Phonics	.44
Writing Teaching	
Inquiry	.57
Scales	.36
Sentence Combining	.35
Early Education Programs	
Preschool	.22–.50
Full-Day vs. Half-Day Kindergarten	.48
Grouping	
Acceleration of Gifted Students	.88
Tutoring	.40
Staff Development	
Feedback	.70
Staff Development for Reading Teaching	.61
Microteaching	.55

From "Meta-Analytic Effects for Policy," by H. J. Walberg and J.-S. Lai, 1999, in *Handbook of Educational Policy*, ed. G. J. Cizek, San Diego, CA: Academic Press, pp. 419–453.

Table 4

Student and Family Influences

Variable	Effect
Prior Knowledge	1.43
Motivation	.73
Family Background	
Home Environment	1.42
Parental Income	.67
Parental Occupation	.42
Parental Education	.38

From *A New Era of School Reform: Going Where Research Takes Us*, by R. J. Marzano, 2000, Aurora, CO: Mid-Continent Research for Education and Learning, pp. 69, 70, 73.

Note. The indexes in the table are on the same scale as the effect sizes in Table 3 but are not necessarily pure, one-way causal effects.

Table 5

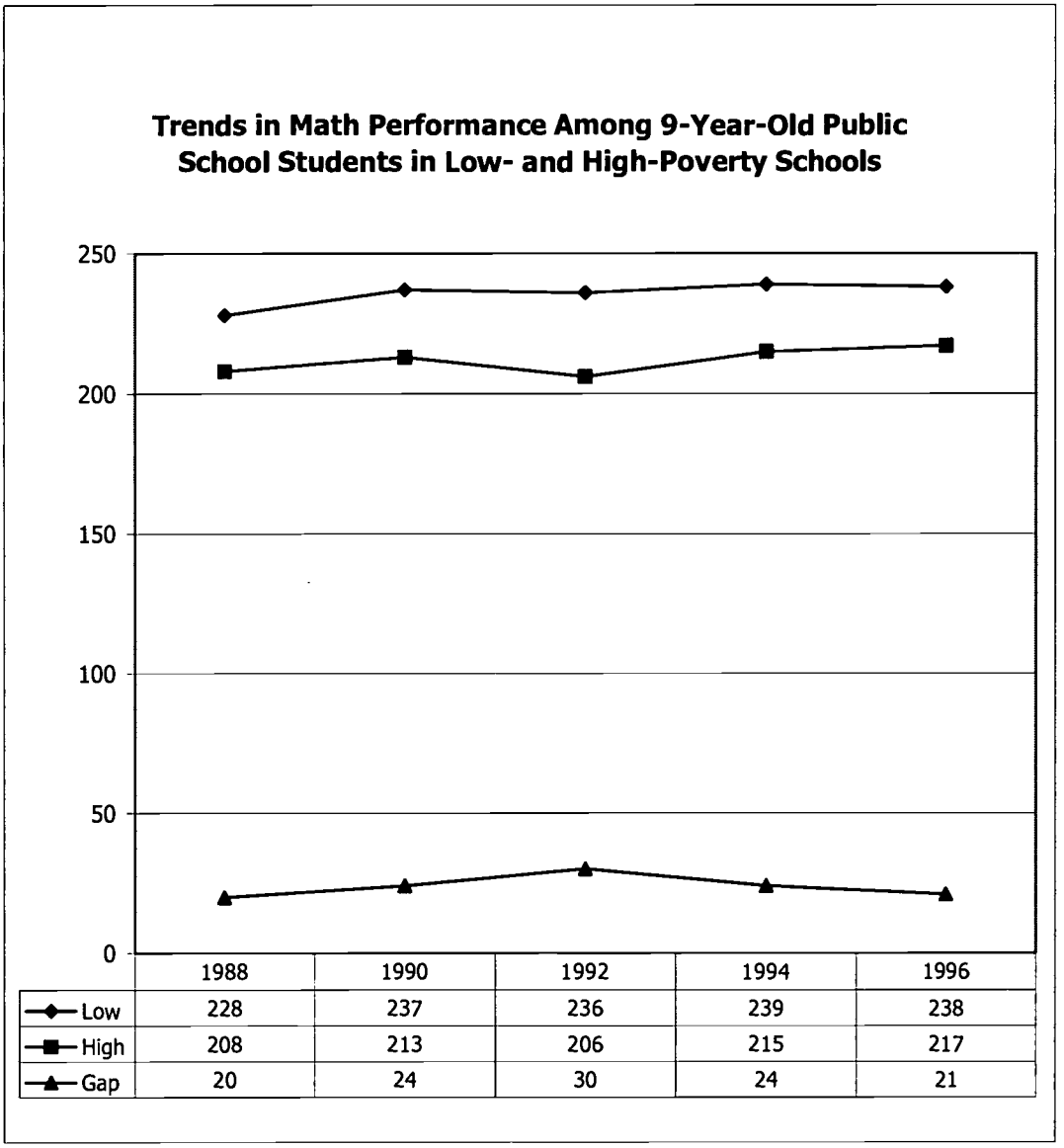
School-Level Possible Influences

Variable	Effect
Opportunity to Learn	.88
Time	.39
Monitoring	.30
Pressure to Achieve	.27
Parental Involvement	.26
School Climate	.22
Leadership	.10
Cooperation	.06

From *A New Era of School Reform: Going Where Research Takes Us*, by Robert J. Marzano, 2000, Aurora, CO: Mid-Continent Research for Education and Learning, p. 56.

Note. The indexes in the table are measured on the same scale as the effect sizes in Tables 3 but are not necessarily pure, one-way causal effects.

Figure 3



	1988	1990	1992	1994	1996
Low	228	237	236	239	238
High	208	213	206	215	217
Gap	20	24	30	24	21

From *Prospects: Final Report on Student Outcomes* by M. J. Puma, N. Karweit, C. Price, A. Ricciuti, W. Thompson, and M. Vaden-Kiernan, 1997, Cambridge, MA: Abt Associates [prepared for the U.S. Department of Education, Office of the Undersecretary], p. 6.

Note. The scale ranges from 0 to 500; high-poverty schools had 76–100% students eligible for free lunch, low-poverty 0–25%.

References

Anderson, J. R., Reder, L. M., & Simon, H. A. (1998). Radical constructivism and cognitive psychology. In D. Ravitch (Ed.), *Brookings papers on education policy, 1998* (pp. 227–255). Washington, DC: Brookings Institution.

Belfield, C. R., & Levin, H. M. (2001). *The effects of competition on educational outcomes: A review of U.S. Evidence.* New York: Columbia University, Teachers College, National Center for the Study of Privatization in Education.

Betts, J. R. (1995). Is there a link between school inputs and earnings? In G. Burtless (Ed.), *Does money matter? The link between schools, student achievement, and adult success* (pp. 141–191). Washington, DC: Brookings Institution.

Betts, J. R., & Costrell, R. M. (2001). Incentives and equity under standards-based reform. In D. Ravitch (Ed.), *Brookings papers on education policy, 2001* (pp. 9–74). Washington, DC: Brookings Institution.

Bishop, J. H. (1996). The impact of curriculum-based external examinations on school priorities and student learning. *International Journal of Educational Research, 23,* 653–752.

Bloom, B. S. (1963). *Change and stability in human characteristics.* New York: John Wiley.

Borland, M., & Howsen, R. (1992). Student academic achievement and the degree of market concentration in education. *Economics of Education Review, 11*(1), 31–39.

Boulding, K. E. (1972). The schooling industry as a possibly pathological section of the American economy. *Review of Educational Research,* 129–143.

Bradley, A. (1997, February 12). Survey reveals teens yearn for high standards. *Education Week,* pp. 1, 38–39.

Bredekamp, S., Knuth, R. A., Kunesh, L. G., & Shulman, D. D. (1992). *What does research say about early childhood education?* Oak Brook, IL: NCREL.

Cawelti, G. (2000). What is our knowledge base for improving student achievement? *Iowa Educational Leadership, 28,* 3–8.

Clark, R. (1983). *Family life and school achievement: Why black children succeed or fail.* Chicago: University of Chicago Press, 1983.

Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F. D., et al. (1966). *Equality of educational opportunity.* Washington, DC: U.S. Department of Health, Education, and Welfare; and U.S. Government Printing Office.

Currie, J. (2001). Early childhood programs. *Journal of Economic Perspectives 15,* 213–238.

Druckman, D., & Swets, J. A. (Eds.). (1994). *Enhancing human performance: Issues, theories, techniques.* Washington, DC: National Academy Press.

Education Trust. (2001). *Dispelling the myth revisited* (Washington, DC: Education Trust.

Entwisle, D. R., & Alexander, K. L. (1992). Summer setback. *American Sociological Review, 57,* 72–84.

Entwisle, D. R., & Alexander, K. L. (1993). Entry into school: The beginning school transition and educational stratification in the United States. *Annual Review of Sociology, 19,* 401–423.

Farkas, G., & Hall. L. S. (2000). Can Title I attain its goal? In D. Ravitch (Ed.), *Brookings papers on education policy, 2000* (pp. 59–123). Washington, DC: Brookings Institution.

Farkas S., & Johnson, J. (1997). *Different drummers: How teachers of teachers view public education.* New York: Public Agenda.

Farkas, S., Johnson, J., Duffett, A., Foleno, T., & Foley, P. (2001). *Trying to stay ahead of the game: Superintendents and principals talk about school leadership.* New York: Public Agenda.

Finn, C. E., & Kanstoroom, M. (2001). State academic standards. In D. Ravitch (Ed.), *Brookings papers on education policy, 2001* (pp. 131–180). Washington, DC: Brookings Institution.

Finn, C. E., Jr., Rotherham, A. J., & Hokanson, C. R., Jr. (Eds.). (2001). *Rethinking special education for a new century.* Washington, DC: Thomas B. Fordham Foundation; and Progressive Policy Institute.

Finn, J. D., & Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal, 27*(3), 557–577.

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. (1987). Synthesis of educational productivity research. *International Journal of Educational Research, 11,* 73–145.

Gersten, R., Schiller, E. P., & Vaughn, S. (Eds.). (2000). *Contemporary special education research: Syntheses of the knowledge base on critical instructional issues.* Mahwah, NJ: Erlbaum.

Gill, B. P., Timpane, P. M., Ross, K. E., & Brewer, D. J. (2001). *Rhetoric versus reality: What we know and what we need to know about vouchers and charter schools.* Santa Monica, CA: RAND Corporation.

Glass G. V., & Smith, M. L. (1979). Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis, 1*(1), 2–16.

Goldberg, M., & Traiman, S. L. (2001). Why business backs education standards. In D. Ravitch (Ed.), *Brookings papers on education policy, 2001* (pp. 81–90). Washington, DC: Brookings Institution.

Goldhaber, D. (2001). The interface between public and private schooling. In. D. H. Monk, H. J. Walberg, & M. C. Wang (Eds.), *Improving educational productivity* (pp. 47–76). Greenwich, CT: Information Age Publishing.

Goldhaber, D. (2002). The mystery of good teaching. *Education Next, 2*(1), 50–55.

Greenberg, R. C., & Walberg, H. J. (1998). The Diogenes effect: Why program evaluations fail. In K. Wong (Ed.), *Advances in educational policy: Perspectives on the social functions of schools* (pp. 167–178). Greenwich, CT: JAI Press.

Greene, J. P. (2002a). *High school graduation rates in the United States.* New York: Manhattan Institute.

Greene, J. P. (2002b). *2001 Education freedom index.* New York: Manhattan Institute.

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

Grissmer, D., & Flanagan, A. (1998). *Exploring rapid achievement gains in North Carolina and Texas.* Washington, DC: National Educational Goals Panel.

Grissmer, D., & Flanagan, A. (2001). Searching for indirect evidence for the effects of statewide reforms. In D. Ravitch (Ed.), *Brookings papers on education policy, 2001* (pp. 181–229). Washington, DC: Brookings Institution.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24,* 557–577.

Hanushek, E. A. (1999a). The evidence on class size, In S. E. Mayer & P. Peterson (Eds.), *Earning and learning: How schools matter* (pp. 131–168). Washington, DC: Brookings Institution.

Hanushek, E. A. (1999b). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis, 21,* 143–164.

Harris Interactive. (2001). *The MetLife survey of the American teacher 2001: Key elements of quality schools.* New York: Author.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore: Paul H. Brooks.

Heller, K. W., Holtzman, W., & Messick, S. (1982). *Placing children in special education: A strategy for equity.* Washington, DC: National Academy of Science Press.

Heyns, B. (1978). *Summer learning and the effects of schooling.* New York: Academic Press.

Hill, P. T. (2000). The federal role in education. In D. Ravitch (Ed.), *Brookings papers on education policy, 2000* (pp. 11–58). Washington, DC: Brookings Institution.

Hirsch, E. D. (1996). *The schools we need and why we don't have them.* New York: Doubleday.

Horn, W., & Tynan, D. (2001a). Revamping special education. *Public Interest, 144,* 36–54.

Horn, W., & Tynan, D. (2001b). Time to make special education "special" again. In C. E. Finn, Jr., A. J. Rotherham, and C. R. Hokanson, Jr. (Eds.), *Rethinking special education for a new century* (pp. 23–51). Washington, DC: Thomas B. Fordham Foundation; and Progressive Policy Institute.

Hoxby, C. M. (1994). *Does competition among public schools benefit students and taxpayers? Evidence from natural variation in school districting* (Working Paper No. 4979). Cambridge, MA: National Bureau of Economic Research.

Hoxby, C. M. (1998). *The effects of class size and composition on student achievement.* Cambridge, MA: National of Bureau Economic Research.

Hoxby, C. M. (2001). *School choice and school productivity (or could school choice be a tide that lifts all boats?).* Washington, DC: National Bureau of Economic Research Conference on the Economics of School Choice.

Hoxby, C. M. (2002). The cost of accountability. In W. M. Evers & H. J. Walberg (Eds.), *School accountability* (pp. 47–74). Stanford, CA: Hoover Institution Press.

Huey, E. B. (1908). *The psychology and pedagogy of reading.* Bristol, England: Thoemmes Press.

Johnson, J., & Farkas, S. (1997). *Getting by: What American teenagers really think about their schools.* New York: Public Agenda.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading. *Psychological Review, 87,* 329–354.

Karoly, L. A., Greenwood, P. W., Everingham, Houbé, J., Kilburn, M. R., Rydell, C. P., et al. (1998). *Investing in our children: What we know and don't know about the costs and benefits of early childhood interventions.* Santa Monica, CA: RAND Corporation.

Kavale, K. A., & Forness, S. R. (1985). *The science of learning disabilities.* Boston: Little, Brown.

Kopp, W. (2001). Ten years of teach for America. In M. C. Wang & H. J. Walberg (Eds.), *Tomorrow's teachers* (pp. 221–234). Richmond, CA: McCutchan.

Kulik, J. A. (in press). Tracking, de-tracking, and skill grouping. In A. J. Reynolds, M. C. Wang, & H. J. Walberg (Eds.), *Can unlike students learn together?*

Lee, V. E., Bryk, A., & Smith, J. B. (1993). The organization of effective secondary schools. *Review of Research in Education, 19,* 171–268.

Lock, E. A., Shaw, K. N., L. Saari, M., & Latham, G. P. (1981). Goal setting and task performance. *Psychological Bulletin, 90,* 125–152.

Loveless, T. (1998). *The tracking and ability grouping debate.* Washington, DC: The Fordham Foundation.

Lyon, G. R., Fletcher, J. M., Shaywitz, S. E., Shaywitz, B. A., Torgesen, J. K., Wood, F. B., et al. (2001). Rethinking learning disabilities. In C. E. Finn, Jr., A. J. Rotherham, & Charles R. Hokanson, Jr. (Eds.), *Rethinking special education for a new century* (pp. 259–287). Washington, DC: Thomas B. Fordham Foundation; and Progressive Policy Institute.

Mankiw, N. G. (1998). *Principles of economics.* Fort Worth, TX: Harcourt Brace Dryden.

Maryland Department of Education. (1990). *Better thinking and learning.* Baltimore, MD: Maryland Department of Education, Division of Instruction.

Marzano, R. J. (2000). *A new era of school reform: going where research takes us.* Aurora, CO: Mid-Continent Research for Education and Learning.

National Commission for Excellence in Education. (1983). *A nation at risk: The imperative for school reform.* Washington, DC: U.S. Government Printing Office.

National Reading Panel. (2000). *The national reading panel report: Teaching children to read* Washington, DC: U.S. Department of Health and Human Services, Public Health Service, National Institute of Health, National Institute of Child Health and Human Development.

Oakes, J., & Lipton, M. (in press). Can unlike children learn together? In A. J. Reynolds, M. C. Wang, & H. J. Walberg (Eds.), *Can unlike students learn together?*

Organisation for Economic Co-Operation and Development. (1995). *Education at a glance: OECD indicators.* Paris: Author.

Organisation for Economic Co-Operation and Development. (1995–2001). *Education at a glance* [Annual editions]. Paris: Author.

Paik, S., Wang, D., & Walberg, H. J. (2002). Timely learning improvements. *Educational Horizons, 80*(2), 69–71.

Paschal, R. A. & Starhia, W. E. (1989). *Educational productivity studies: A quantitative synthesis.* Paper presented at the annual meeting of the American Educational Research Association.

Peterson, P. E. (2001). Choice in American education. In T. Moe (Ed.), *A primer on America's schools* (pp. 249–284). Stanford, CA: Hoover Institution Press.

Pogrow, S. (2000). Success for All does not produce success for students. *Phi Delta Kappan, 82*(1), 67–80.

Puma, M. J., Karweit, N. Price, C., Ricciuti, A., Thompson, W., & Vaden-Kiernan, M. (1997). *Prospects: Final report on student outcomes* [prepared for the U.S. Department of Education, Office of the Undersecretary]. Cambridge, MA: Abt Associates.

RAND Corporation. (2002, February 4). *Press release.* Santa Monica, CA: RAND Corporation.

Ravitch, D. (2001). Introduction. In D. Ravitch (Ed.), *Brookings papers on education policy, 2001* (pp. 1–8). Washington, DC: Brookings Institution.

Raymond, M., & Fletcher, S. (2002). Teach for America: The first evidence on classroom performance. *Education Next, 2*(1), 62–68.

Reschly, D. J. (1987). Learning characteristics of mildly handicapped students: Implications for classification, placement, and programming. In M. C. Wang, M. C. Reynolds, & H. J. Walberg (Eds.), *Handbook of special education: Research and practice: Vol. 1. Learner characteristics and adaptive education* (pp. 35–58). Oxford, England: Pergamon Press.

Reynolds, A. J. (2000). *Success in early intervention: The Chicago Child–Parent Centers.* Lincoln, NE: University of Nebraska Press.

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. L. (2001). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools. *Journal of the American Medical Association, 285,* 2339–2346.

Shanahan, T., & Barr, R. (1995). Reading Recovery: An independent evaluation of the effects of an early instructional intervention for at-risk learners. *Reading Research Quarterly, 40,* 958–996.

Simon, H. A. (1981). *Sciences of the artificial.* Cambridge, MA: MIT Press.

Stanovich, K. E. (1986). Matthew effects in reading. *Reading Research Quarterly, 21,* 360–406.

Stecher, B. M., & Bohrnstedt, G. W. (2001). *Class size reduction in California: Findings from 1999–00 and 2000–01.* Sacramento, CA: California Department of Education.

Sternberg, R. J., & Grigorenko, E. L. (2001). Learning disabilities, schooling, and society. *Phi Delta Kappan, 83*(4), 335–338.

Sticht, T. G., & H. James, J. H. (1984). Listening and reading. In P. D. Pearson, R. Barr, M. Kamil, & P. L. Mosenthal (Eds.), *Handbook of Reading Research* (Vol. 1, pp. 293–317). New York: Longman.

Stone, J. E. (1996). Developmentalism: An obscure but pervasive restriction on educational improvement. *Education Policy Analysis Archives, 4*(8). Retrieved from http://olam.edu.asu.edu/epaa/v4n8.html

Swanson, H. L. (1999). *Interventions for students with learning disabilities.* New York: Guilford Press.

Teske, P., & Schneider, M. (2001). What research can tell policymakers about school choice. *Journal of Policy Analysis and Management, 20,* 609–631.

Tommasi, M., & Ierulli, K. (Eds.). (1995). *The economics of human behavior* Cambridge, England: Cambridge University Press.

Topping, K. J., & Paul, T. D. (1999). Computer-assisted assessment of practice at reading: A large-scale survey using accelerated reader data [Entire issue]. *Reading and Writing Quarterly, 15*(3).

U. S. Department of Education. (1997). *Digest of education statistics* Washington, DC: Author.

Viadero, D. (1994, July 13). Focus on: Research. School's out for the summer. *Education Week*, pp. 35–37.

Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership, 41*(8), 19–27.

Walberg, H. J. (1998a, November 4). Incentivized school standards work. *Education Week*, p. 48.

Walberg, H. J. (1998b). Uncompetitive American schools: Causes and cures. In D. Ravitch (Ed.), *Brookings papers on education policy, 1998* (pp. 173–206). Washington, DC: Brookings Institution.

Walberg, H. J., Hase, K., & Rasher, S. P. (1978). English mastery as a diminishing function of time. *TESOL Quarterly, 12*(4), 427–437.

Walberg, H. J., & Lai, J.-S. (1999). Meta-analytic effects for policy. In G. J. Cizek (Ed.), *Handbook of Educational Policy* (pp. 419–453). San Diego, CA.: Academic Press.

Walberg, H. J., & Tsai, S.-L. (1984). Matthew effects in education. *American Educational Research Journal, 20,* 359–374.

Walberg, H. J., & Walberg H. J., III. (1994). Losing local control. *Educational Researcher 23*(5), 23–29.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993a). Toward a knowledge base for school learning. *Review of Educational Research, 63,* 249–294.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993b). What helps students learn? *Educational Leadership, 50,* 74–77.

Wayne, A. J., & Youngs, P. (2001, November). *Teacher characteristics and student achievement gains: A review.* Paper presented at the annual meeting of the Association for Public Policy Analysis and Management, Washington, DC.

White, K. R. (1985). Efficacy of early intervention. *Journal of Special Education, 19,* 401–416.

Wigfield A., & Asher, S. R. (1984). Social and motivational influences on reading. In P. D. Pearson, R. Barr, M. Kamil, & P. L. Mosenthal (Eds.), *Handbook of Reading Research* (Vol. 1, pp. 423–452). New York: Longman.

Woessmann, L. (2001). Why students in some countries do better: International evidence on the importance of education policy. *Education Matters, 1*(2), 67–74.

Yinger, J., Bloom, H. S., Borch-Supan, A., & Ladd, H. F. (1988). *Property taxes and housing values*. Boston: Academic Press.

Ysseldyke, J. E. (1987). Classification of handicapped students. In M. C. Wang, M. C. Reynolds, & H. H. Walberg (Eds.), *Handbook of special education: Research and practice: Vol. 1. Learner characteristics and adaptive education* (253–271). Oxford, England: Pergamon Press.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").