DOCUMENT RESUME

ED 370 210 EA 025 845

AUTHOR Lee, Valerie E.; Smith, Julia B.
TITLE Effects of High School Restructuring and Size on Gains in Achievement and Engagement for Early Secondary School Students.
INSTITUTION Center on Organization and Restructuring of Schools, Madison, WI.; Wisconsin Center for Education Research, Madison.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE 1 Apr 94
CONTRACT R117Q00005-94
NOTE 64p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Academic Achievement; Educational Change; High Schools; *High School Students; Program Evaluation; Public Schools; *School Effectiveness; *School Restructuring; *Student Development
IDENTIFIERS *National Education Longitudinal Study 1988

ABSTRACT

School restructuring continues to be a common approach to improving education. Despite restructuring's continued and growing support, there is little research to support its effectiveness. The theoretical contrast exposed in school restructuring is between bureaucratic and organic organizational forms. A study assessed the effect of restructuring on students during their early high school years. Data were used from the first two waves of the National Educational Longitudinal Study of 1988 with a nationally representative sample of 11,794 high school sophomores in 820 secondary schools. Restructuring effects were evaluated on gains in students' engagement and achievement in mathematics, reading, social studies, and science between grades 8 and 10, as well as the social distribution of the gains. Schools were categorized as restructured, moderate, or traditional based on 30 structural practices measures. Restructured high schools and unrestructured schools were contrasted with traditionally reformed schools. High school size was an important structural feature. Results showed that students' achievement and engagement were significantly higher in restructured schools and lower in unrestructured schools. Achievement and engagement gains were also more equitably distributed in restructured schools. Smaller schools also had higher and more equitable engagement and achievement. (Contains 72 references.) (JPT)

EFFECTS OF HIGH SCHOOL RESTRUCTURING AND SIZE ON GAINS IN ACHIEVEMENT AND ENGAGEMENT FOR EARLY SECONDARY SCHOOL STUDENTS

Valerie E. Lee
University of Michigan

Julia B. Smith
University of Rochester

Revised April 1, 1994

FINAL DELIVERABLE TO OERI

This paper was prepared for the Center on Organization and Restructuring of Schools, supported by the U.S. Department of Education, Office of Educational Research and Improvement (Grant No. R117Q00005-94) and by the Wisconsin Center for Education Research, School of Education, University of Wisconsin-Madison. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the supporting agencies. We are very grateful for useful suggestions from several people: Steven G. Heeringa, Manager of Sampling, Survey Research Center, Institute for Social Research, University of Michigan, for advice on design weights in stratified samples with longitudinal designs; Anthony Bryk for advice on analyzing data with gain score outcomes; David Cohen, Fred Newmann, Andrew Porter, and Brian Rowan for useful suggestions for revision of a previous draft. We accept full responsibility for any errors in analysis or interpretation which remain.

Effects of High School Restructuring and Size on Gains in Achievement and Engagement for Early Secondary School Students

Abstract

This study assesses the impact of attending restructured secondary schools on students in their early years of high school. Employing data from the first two waves of the National Educational Longitudinal Study of 1988, the study focused on a nationally representative sample of 11,794 high school sophomores in 820 secondary schools. Restructuring effects were evaluated on gains in students' engagement and achievement (in mathematics, reading, social studies, and science) between 8th and 10th grade, as well as the social distribution of those gains. The concept of restructuring was tapped with a set of 30 structural practices which we classified as *restructured*, *moderate*, or *traditional* according to their prevalence in these high schools. Restructured high schools (those engaging in several restructuring reforms) and unrestructured high schools (with none of the 30 practices in place) were contrasted with traditionally reformed schools. High school size was considered as an important structural feature. Results were strong and consistent: students' achievement gains and engagement were significantly higher in restructured schools and lower in unrestructured schools. The distribution of gains in achievement and engagement was also more equitable in restructured schools. School size also had strong effects, with higher and more equitable engagement and achievement more prevalent in smaller high schools.

Effects of High School Restructuring and Size on Gains in Achievement and Engagement for Early Secondary School Students

The movement aimed at restructuring American education continues to be popular in the lexicon of contemporary school reforms. The sustained attraction of this clarion call for fundamental change reflects our nation's continuing dissatisfaction with its schools. Despite its growing popularity among school people and educational policy specialists, the reform movement embodied in the term "restructuring" rests on thin and inconsistent theory. The recommendations for reform included under this heading have been drawn from diffuse sources, which include the rather narrow research base on effective teaching practices, production models formulated in industrial settings, or popular perscriptions to empower disenfranchised interest groups. Rather than trying to develop a coherent theoretical base for the movement from those disparate sources, in this paper we attempt to locate the movement within a theoretical contrast that has become reinvigorated in the last half-decade to describe differences in secondary schools -- the contrast of bureaucratic with organic organizational forms. We hope that the effort to provide a theoretical grounding for the movement is useful, and that the particular grounding we provide is generally appropriate. We admit, however, that the theoretical "shoe" doesn't always fit the policy "foot."

The "restructuring focus" we take aims at identifying secondary schools by the organizational practices they follow. The practices that fit our definition of restructuring capture two related ideas: (1) they represent a movement from the bureaucratic organizational form of American schooling toward the organic form, and (2) they also represent a departure from conventional practice. Even a cursory overview of the history of American education reveals that the organization of secondary education has been subjected to reform almost from its inception. As Tyack (1974) points out, the separation of secondary schooling from the traditional village school into a more formal and systemized -- and ultimately bureaucratic -- organization was itself a significant educational reform in the late nineteenth and early twentieth century. Over the last decade or so, many important social critics (e.g., Boyer 1983; Sizer 1984) describe the secondary school as so fundamentally linked to the success of our economic and political future that its failure jeopardizes the future of the nation

itself. Its multiple ailments have prompted the constant and strident calls for reform, revision, and restructuring of high schools. Almost no one argues that the high school, as an organization, is "working." Particularly poignant was the description of the crisis in A Nation at Risk, one that mobilized the nation to consider serious educational reform:

> ...while we can take justifiable pride in what our schools and colleges have historically accomplished and contributed to the United States and the well-being of its people, the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a nation and a people (National Commission on Excellence in Education, 1983:5).

Although calls for reforms to meet existing goals have been constant, the goals of American education have themselves undergone substantial revision over the 120-year institutional life of public secondary schooling. A major thrust of the progressive movement was the establishment of national "web" of large and comprehensive high schools, schools which were meant to exemplify such typically American ideals as efficiency, differentiation, specialization, depersonalization, and standardization -- in effect, highly-managed, smoothly-functioning, and well-oiled machines whose goals were the production of human capital. Few educational reform efforts "succeeded" as well as the development of the comprehensive high school. Within the last few decades, however, the very organizational dimensions which undergird the American high school have been questioned. They have increasingly been described as inappropriate responses to the human concerns of organizational members (e.g., Newmann and Oliver, 1967; Tyack, 1974). Bigger schools may not be better schools, in terms of either economic efficiency (Fox, 1981; Guthrie, 1979) or academic development (Godding and Wagner 1985; Haller et al. 1990; Monk 1987). Moreover, a standard of equity in the distribution of educational outcomes has joined the classic standard of excellence in assessing the effectiveness of schools (Coleman et al. 1966; Edmonds 1984; Purkey and Smith 1983). As the nation rethinks and revises what goals are appropriate for a democratic educational system, it must also revise the means for accomplishing those goals. Because secondary schools are vital agents in the educational system of a democracy, they come under special scrutiny.

Background

Two Theories of School Organization

The structure of an organization refers to the type, character, and number of relationships between different members around its technical core of work (Perrow 1967; Simon 1976). Because our study aims to establish a link between organizational form and student achievement, we find it useful to invoke an organizational contrast which focuses on the "core technology" of schools. The two organizational forms have very different conceptions of the core technology and, thus, make very different assumptions about knowledge, learning, and teaching. Although one form (the bureaucratic) assumes a technology that is routine, clear, and stable, the alternative form (the organic) assumes a non-routine core technology (Burns and Stalker 1961; Rowan 1990). But how do these assumptions get played out in the structure of schools?

The bureaucratic form. Organizations with a bureaucratic form characteristically include specialized and differentiated work roles, a top-down hierarchy of decisionmaking, and a formalization of goals and expectations into affectively neutral rules and codes of behavior (Lee, Bryk, and Smith 1993; Rowan 1990). In such organizations, the routine technical activity -- instruction -- would break down knowledge into a curriculum composed of discrete and fixed subjects. Teaching would consist of imparting specialized knowledge; instruction would be organized into a standardized and sequenced pattern within subjects. Learning would be assessed by measuring mastery of subject matter, and learners would be sorted into specialized instructional treatments to maximize the matching of subject matter and learner. The organization of instruction with practices such as departmentalization and tracking would make sense under this specialization model. Such reforms as increased academic standards, tightened graduation requirements, and teacher-proof curriculum materials also fit the bureaucratic model.

The organic form. The other end of this continuum is occupied by organizations with an organic or communitarian form (Lee, Bryk, and Smith 1993; Rowan 1990). Here, tasks are less certain and conditions more changeable and unpredictable. Such organizations typically emphasize shared responsibility for work, shared commitment to a common set of goals, lateral communication and power in decision-making, and greater personal-

ization and individual discretion framing expectations and behavior. This organizational form, typical of small high schools in the early 20th century (especially in rural areas), would conceive knowledge as multi-dimensional and interdisciplinary. Teaching would be responsive to students' opinions, talents, and tastes. Learning would be built more around concrete "problems" than abstract "subjects," and assessment would be more flexible and less standardized. Organizational responses to a more organic definition of the school's core technology might include independent study, interdisciplinary teaching, flexible scheduling, cooperative learning, and mixed-ability classes.

Moving away from the dominant form. These theories of teaching and learning are in fact well established in American education, and have undergirded historical and theoretical discussions about the proper direction of school reform for at least a century (Cuban 1984, 1990). However, one theory dominates in secondary schools. The reforms of the progressive era moved schools in a bureaucratic direction and its "perfect" product, the comprehensive high school. Although some reforms have tried to make the traditional system "work better" through tinkering, until quite recently it has been assumed that the bureaucratic structure itself was solid. Despite a history of theoretical discussions involving both organizational forms, a bureaucratic form still constitutes the "tradition" against which current structural reform efforts are targeted, at least for secondary schools.

However, more recent calls for reform have begun to recognize the need for a profound change in the structure of schools. For example, Sarason (1990) states: "Change will not occur unless there is an alteration of power relationships among those in the system and within the classroom" (pg. xiv). Recent calls for the "restructuring" of American high schools fit this category. We suggest fitting the restructuring movement within the theories of teaching and learning we have described. Calls for restruc- turing schools seem to suggest shifting schools away from the bureaucratic and toward the organic organizational model.

Complicating the Theories of Restructuring

Organizational inconsistency. Although the theories describing the bureaucratic and organic forms of high school organization may seem distinct in the abstract, unfortunately the complex organization of the

American high school somewhat blurs the clarity of the theory. Characterizing schools as "loosely coupled systems" reveals the source of this confusion. Weick (1976) argues that large and small decisions are constantly being made in schools -- at different levels and in different situations, formally and informally, with loose monitoring of and coordination between the myriad daily activities that surround the work of the school.

In loose systems, different goals, methods, and activities routinely reside side-by-side within a single organizational framework. This situation renders both the casual and the systematic observer unable to characterize the organization as consistently either bureaucratic or organic, or somewhere in between. The comprehensive high school surely includes several smaller organizations within it. Thus, some parts of high schools are bureaucratically organized, others are organic. An uneven communication system acts to protect the technical core from being inconvenienced by such inconsistency (or perhaps even having it recognized -- Rowan 1990; Shedd and Bacharach 1991).

"Action-reaction". Within the typical high school organization, efforts at reform may also appear ambiguous, even contradictory, depending on the observer's position in or out of the system. While each form seems theoretically distinct, it is not uncommon within a single high school to find reforms underway aimed at moving the school toward a more organic organization, and simultaneously other efforts targeted at strengthening the formal hierarchy. Indeed, in their theory of formal organizations, Blau and Scott (1962) argue that efforts to shift an organization away from traditional (i.e., bureaucratic) power relationships would induce countervailing efforts to strengthen those ties -- a natural reaction to anything more than superficial change in the power structure. Under this "action-reaction" theory, schools would not be expected to easily replace one organizational form for another. Rather, we might predict an attempt to "restructure" the organization with new dynamics in existing relationships, possibly in combination with other efforts to strengthen those relationships. High school reform, under such theoretical complexity, is not a simple zero-sum process of replacement. Rather, it involves an interactive balancing between new methods and existing thinking about schools and learning.

The difficulty of real reform. This expansion of the theory suggests that the modern comprehensive high school is essentially a conservative organization in which fundamental change is quite difficult to effect. Thus, it could be expected that reforms aimed at nudging the organization away from a bureaucratic toward an organic form would be difficult to initiate, would not take hold easily, would be resisted by many, and would be firmly in place in few high schools. One part of the definition of school restructuring with which we began this paper, substantial departure from conventional practice, captures the essence of this idea. Thus, even now, when secondary school organizational reform is on the front burner of educational policy, we would expect the form of the reform effort to be related to the likelihood of its adoption in any one high school.

The reform "bandwagon" is moving fast and the spirit to reform is both common and desirable. However, reforms with an organic form are probably less common than those which do not seem to disturb the heart of an essentially bureaucratic organzation. The few studies which have examined the degree of implementation of restructuring (e.g., Berends and King, 1994; CORS 1992; Brown 1993) have concluded that the fundamentally restructured school is very rare.

Effects of Restructuring

There is no absence of recent writings about the school restructuring movement. Many have been helpful in conceptualizing the issues (see, for example, Conley 1993; Elmore 1990; Murphy and Hallinger 1993). To our knowledge, however, only one study -- in a single school district (Jefferson County, Kentucky) -- has evaluated the *effects* of school restructuring on student outcomes. Jefferson County schools have been engaged in a serious restructuring effort for over a decade. The study focused on 42 of the 157 schools in the district, which were categorized into three groups based on their commitment to restructuring (Kyle 1993). Study schools were matched across the three groups by demographic composition, mobility rates, and level (elementary, middle, secondary). Group I schools had a sustained commitment (3-5 years) to restructuring. Group II schools, in the exploration stage of restructuring, had only short-term experience with a wide range of reforms. Group III schools, generally satisfied with their current practices, were uncommitted to restructuring even in theory.

The study compared the school groups on three outcomes: basic improvement (achievement on standardized tests), annual improvement rate (between 1988 and 1991), and positive involvement (attendance, dropout and suspension rates, and parent involvement). The pattern of effects was consistent across outcomes. Compared to either Group II or Group III schools, Group I schools showed higher basic improvement, and more positive involvement. Equally interesting, Group II schools -- where reform was neither stable nor consistent, but where a multitude of reforms had been tried -- showed less favorable outcomes than schools in Group III. Elementary, middle, and high schools all followed the same pattern of effects. Despite some understandable design weaknesses (small sample size, lack of statistical testing of differences, and possible selection bias), the willingness of Jefferson County to take a hard-nosed look at the effects of its commitment to reform and publicize the results is to be commended. The results here are quite encouraging -- in this district, school restructuring appears to have positive effects for students.

School Size as a Reform Issue

Although school size is a potential organizational correlate of restructuring, this structural feature of schools is seldom seen as a policy issue *per se*. The inconsistency of research findings on the effects of high school size results from problems in the research: inconsistent definitions, weak methodology, and (primarily) the lack of a clear focus on what, precisely, might be affected by a change in school size and on the process through which those effects might work. Here we restrict our focus to the influence of school size on learning. Much existing research on this topic has concluded that the effects on students can only be indirect. That is, size could influence the economic, academic, or social organization of secondary schools. These organizational characteristics, in turn, could have consequences for students.

Resource strength. Larger school size may evoke economies of scale, since such schools have more resources available for teachers' salaries, instructional materials, or staff development. However, as unit size increases, an administrative layer is often added which consumes some of these resources (Bidwell and Kasarda 1975; Guthrie 1979). Moreover, the relationship between size and resource base is not linear (Friedkin and Neccochea 1988); rather it is contingent upon the community's socioeconomic

status. The higher incidence of "exceptional problems" (e.g., delinquency, drug abuse, learning disability) in larger schools in low-income areas offsets the economies of scale derived from a larger resource base. Although the theoretical link between increasing school size and increased resources available to the unit is logical, research has not supported this benefit. Instead, large units have added demands on the resource base -- a thicker administrative layer, some disenfranchisement of the community served, potential inequity in the distribution of those resources. Even with greater aggregate financial resources in large schools, there is no evidence that these increased resources increase student learning.

Curriculum specialization. Another important issue is the relationship between organizational size and program specialization. Although bigger schools may result in what some call quality ("specialization of personnel and more effective use of particular kinds of capital equipment," according to Chambers [1981:31]), increasing specialization of personnel and services is not necessarily seen as a characteristic of "good" schools (Bryk, Lee, and Holland, 1993; Grant 1988; Lightfoot 1984). The link between size and specialization, although reasonable, would only be beneficial under the bureaucratic model. A classic argument supporting the comprehensive high school has been that larger schools have larger numbers of students with similar needs, which allows for the creation of specialized programs to address those needs (Conant 1959; Sher and Tompkins 1977). In contrast, smaller schools must focus their resources on core programs, with the consequence that marginal students (at either end of a distribution of ability and interests) are either excluded from programs or absorbed into more general programs which may not meet their needs (Haller et al. 1990; Monk 1987; Powell, Farrar, and Cohen 1985).

Larger schools certainly offer more courses, but the added courses tend not to meet the needs of the most students (Monk 1987). Instead, large schools specialize course offerings, creating more courses for the most able students (Haller et al. 1990). This results in more homogeneity within classrooms and greater ability differences between them. The consequences are not necessarily an improvement in student learning. Findings from the large body of research on tracking are consistent: the typical curriculum in American high schools, which offers a wide array of courses requiring very different levels of effort and commitment from students and extensive individual choice among courses, results in students' academic experiences

and outcomes which are considerably differentiated by social background and ability. Lee and Bryk (1989a) found that school size, while unrelated to school average achievement, was associated with greater social differentiation of achievement. Lee and Smith (1993) supported this conclusion in their study of middle-grade schools, finding a negative effect of size on both achievement and engagement for 8th graders. These findings build on more general empirical studies linking differences in students' academic experiences to social stratification in academic outcomes (c.f. Alexander, Cook, and McDill 1978; Garet and DeLaney, 1988; Heyns 1974; Lee and Bryk 1988; Oakes, 1985, 1990; Rosenbaum 1976). Thus, while larger schools typically offer a more diverse set of academic offerings to meet specialized student needs, the consequences of this curriculum expansion are actually harmful in social terms.

Social relations. Learning may also be influenced indirectly by school size, through its impact on personal interactions. The consequences of working in large schools, for either students or teachers, are especially important for engagement or isolation, commitment or alienation, social relations between members (Barker and Gump, 1964; Rutter, Maughan, Mortimore, Ouston, and Smith 1979; Seeman 1975). In bigger schools, contact between individuals becomes more formal and more rational (Anderson 1982; Bridges and Hallinan 1978). Goals become more diverse and possibly more conflicting as size increases (Forsyth and Hoy 1978; Fuller, Wood, Rapoport, and Dornbusch 1982; March and Olsen 1976; Rutter et al. 1979). Management is more complex, making more difficult the monitoring of daily activities and intervening as problems arise (Newmann 1981; Gottfredson and Daiger 1979; Wehlage, Stone, and Kleibard 1980).

Research that contrasts bureaucratically and organically organized schools shows negative effects of school size -- either direct or indirect -- on both mean values and the social distribution of educational outcomes (Bryk and Driscoll 1988; Bryk, Lee and Holland 1993). Indirectly, these effects operate through school organization by influencing either the specialization of roles, the structure of social relations, differentiation of the curriculum, or all three. Prominent on their list of suggested school reforms for "expanding opportunities for mobility and enrichment," the Carnegie Foundation recommends "breaking up large schools into smaller units" (Carnegie Foundation for the Advancement of Teaching, 1992:79). While the empirical evidence about the effects of school size is not

completely consistent, the general conclusion supports a movement toward smaller high schools.

Past Difficulties With Organizational Research in Schools

As suggested, the move to reform high schools has a long history, although the "school restructuring" reform movement is recent. Interest in the influence of school organization on students has been common among sociologists of education for over half a century (certainly, since the publication of Willard Waller's 1932 classic, The Sociology of Teaching). Quantitative studies on this topic (particularly at the secondary level) were uncommon until rather recently, however, for several reasons. First is the limited variability in structural form. Almost from the beginning, the organization of the American high school has been characterized by a strongly bureaucratic form, described quite vividly several decades ago by Conant (1959) and more recently by Powell, Farrar, and Cohen (1985).

Until a few years ago most empirical research on this topic misconceptualized the questions and methods for investigating them. "School effects" questions are by nature multilevel; thus, efforts to quantify how schools affect students require methods which capitalize on the nested nature of the questions and data. Until such methods were available, researchers analyzed data either at the level where the major proportion of variation in outcomes occurred (i.e., between students) -- thereby assuming students in the same school are independent of each other in how school factors influence them, or at the level where the "treatment" was administered (i.e., between schools) -- assuming no variation among students in school effects. Statistical and substantive problems plague both approaches.

Another weakness concerns the timing of data collection. Recent studies employing longitudinal data from *High School and Beyond* (HS&B), even with appropriate statistical methods, were restricted to studying school effects during the latter years in high school, since data were collected at the end of the sophomore and senior years. These studies thus ignored potentially powerful effects at the *beginning* of high school. It seems possible that organizational effects might not be constant over the high school years. Rather, they could accrue early and level off thereafter.[1] The study described here hopefully overcomes these difficulties, for the reasons spelled out below.

Data and Statistical Needs

Quantitative research which evaluates how variation between schools in organizational structure affects students has stringent data requirements. Large samples are needed at two nested levels, students and schools. To capture how school structure affects student *learning*, we also need measures of students' academic status at two or more time points, preferably: (1) at their entry point into the school and (2) after a lengthy expose to the organizations to be evaluated. In addition, since schools' structural variation may have differential effects on learning in different subject matters, we need measures of student learning across the curriculum. As structural variation may also have profound effects on non-cognitive outcomes, it is advisable to examine the dependent measures to be considered beyond student achievement.

As suggested, even with the ideal data structure just described -- substantial samples of students and schools, achievement measured across the curriculum and at more than a single time point, and non-cognitive outcomes also measured longitudinally -- such a study requires appropriate statistical software to analyze nested data. Fortunately, several such multilevel programs are now commerically available. Their growing use in teaching and research, and a burgeoning body of published studies using these methods, allows researchers to investigate organizational questions like the ones posed here. We use one of these programs, Hierarchical Linear Models (HLM), in this study.

Research Hypotheses

This study aims to evaluate how some of the reforms which are part of the movement to alter the organizational form of American secondary schools -- currently called "school restructuring" -- are taking hold in the nation's high schools, and how these organizational changes are affecting students. Within the evaluation framework, we pose several hypotheses which we test with current and nationally representative data.

Hypothesis 1. The first hypothesis focuses on the prevalence of particular organizational practices aimed at changing the structure of high schools. Specifically, we suggest that more fundamental reforms, those that aim to move schools toward the organic form, are less common than reforms that aim to change only the bureaucratic form.

Hypothesis 2. Our subsequent hypotheses are aimed at evaluating the effects of attending schools which have (or do not have) in place different types of reform on students' experiences during their first two years of high school. The second and third hypotheses focus on the nature of organizational practices. Our second hypothesis suggests that students who attend schools with educational practices characterized by the organic form are positively affected, in comparison to those who attend schools that have adopted practices structuring their students' experiences in more traditional or bureaucratic ways. We suggest positive effects for students in both cognitive and non-cognitive domains. The outcomes explored include learning in several areas of the curriculum, as well as engagement with academic life. We also hypothesize positive effects on the social distribution of these outcomes -- toward more social equity in outcomes.

Hypothesis 3. The third hypothesis involves a variation on the "restructuring theme" posed in the second hypothesis. Here we suggest that students attending schools which have structural practices in place that are consistent with the bureaucratic rather than the organic organizational form are favored, compared to students attending schools that engage in none of the practices we identify as structural reforms. Effects are hypothesized on the same outcome set as described in Hypothesis 2.

Hypothesis 4. Another hypothesis concerns *school size*. We suggest that students attending smaller high schools are favored by that experience, above and beyond the other organizational practices considered under the "restructuring" or "traditional" definitions. The positive effects of attending small schools on students, we hypothesize, occur on both cognitive and non-cognitive outcomes. In addition to the averages outcome effects, we also suggest that attending smaller high schools engenders a more socially equitable distribution of learning and engagement across students of differing social backgrounds.

Method

Sample and Data

Sampling design. The sample for this study was drawn from the first and second waves of the *National Educational Longitudinal Study of 1988* (NELS), a general-purpose study of the educational status and progress of a large sample of students and schools sponsored by the National Center for

Education Statistics (NCES). About 25 eighth-grade students in each of about 1,000 American middle-grade schools were sampled in 1988 -- about 22,000 students in all (Ingels, Abraham, Rasinski, Karr, Spencer, and Frankel 1989). In 1990, the same students were traced to the high school they were attending, based on a locator questionnaire completed by 8th graders, in which they were asked to indicate the two high schools they were most likely to attend (Ingels, Scott, Lindmark, Frankel, and Meyers 1992). Despite obvious difficulties in locating students, response rates were reasonably high.[2]

Data filters. This study includes only the NELS sophomores who fit the following data filters: (1) students had to have full cognitive test-score data from both the base-year and first follow-up; (2) there must be data from their high schools and their teachers; (3) students had to be enrolled in public, Catholic, and elite private secondary schools (i.e., students in other private schools were dropped); and (4) they must have been attending high schools with at least 5 NELS-sampled students in them.[3] These data filters are quite similar to those applied on our study of school restructuring using base-year NELS data (Lee and Smith, 1993). Application of these filters resulted in a sample of 11,794 sophomores in 820 high schools, averaging 14.4 students per school. The large majority (650) were public high schools, with smaller number of Catholic (68) and independent (47) secondary schools.

Design weights. The base-year NELS sample included oversampling of certain types of students and schools (especially private schools and schools with high concentrations of Hispanic and Asian students). This oversampling was sustained, of course, in the first follow-up. This sampling procedure necessitated the use of design weights in all analyses of NELS follow-up data. Although NCES data generally come with design weights for students and schools, the NELS follow-up data tapes provided design weights only for *students*, but not for schools. This is because the high schools were not, themselves, selected in the NELS sampling frame. Rather, they were selected by NELS students. Thus, the schools' representativeness among the population of U.S. high schools is unknown. This lack of school-level design weights presented us with a serious dilemma. Our research questions suggested the need to employ hierarchical statistical methods in this study. As the questions and method both focus on variability between schools, we needed school weights. We thus found it necessary

to construct a set of "pseudo-design weights" ourselves for the high schools attended by NELS sophomores.[4] The HLM analyses in this study employ these school weights, which were adjusted to a mean of one on our sample to preserve appropriate significance testing in multivariate analyses. Only the descriptive statistics for students (Table 2) employ the NELS student design weights supplied by NCES, also adjusted for appropriate statistical testing.

Measures

School restructuring measures. The logic underlying our definition of the construct of school restructuring rests on two criteria: the definition of organic organization in the literature review above, and also on a somewhat more practical definition of restructuring as "substantial departure from conventional practice." Each of these criteria contributed to the final determination of how we defined and used this construct in the subsequent analyses.

Using data from the NELS first follow-up school questionnaire, typically completed by the school principal, we identified a large set of individual practices which describe a school's structure (30 items). This list captures important elements of how American secondary schools define their efforts toward reform. Many of the items tap practices identical to those used in earlier work on this topic (Lee and Smith 1993). The items invited dichotomous administrator responses, indicating whether each practice was currently in place in the secondary school (coded 1=yes, 0=no).

Classification of structural practices. The 30 practices can be broken into groups according to their adherence to or departure from a bureaucratic structural form. Practices which reinforce a top-down power orientation from administrators to teachers and students include: strong departmentalization, emphasis on teacher expertise and specialization, emphasis on formal instructional requirements for students, and recognition programs for teachers which operate within the worker-reward paradigm. The natural extention of this type of structure to the external community focuses on parents as recipients of information provided by teachers and administrators.

Practices which shift schools away from the bureaucratic toward a more organic form can be identified on three domains: (1) those which aim to

reorganize instruction -- mixed-ability classes; cooperative learning focus; independent study in different curriculum areas; flexible time for classes; (2) altering authority and expertise in the school -- interdisciplinary teaching teams; students evaluating teachers; staff solving school problems; and (3) building a more communally organized school -- using parent volunteers; students keeping the same homeroom for several years; common planning time for teachers; schools-within-schools. There is a conceptual clarity about these practices taken as a group. Together they represent movement toward a more organic form of school organization in the areas of instruction, authority, and community.

Testing Hypothesis 1: Substantial departure from conventional practice. The logic laid out above, describing the prevalence of the comprehensive high school model on the current American educational scene, suggests that organic reform practices would be less common than reforms which fit the traditional bureaucatic model. To test this hypothesis, we computed the probability of each practice being in place in the 820 high schools in our sample (i.e., the proportion of average NELS high schools reported engaging in the practice). These item probabilities ranged from .09 to .69. The average school engaged in 12 of the 30 practices. The individual structural practices, ranked by the probability with which they occur in schools, are displayed in Table 1, along with their individual NELS variable names.[5]

Insert Table 1 about here

The results displayed in Table 1 show that the practices that fit our conceptual definition of restructuring are the least common reforms in American secondary schools. This grouping of practices and the probabilities of their occurrence provides some confirmation for Hypothesis 1 -- that organic reform efforts are less common than bureaucratic reforms in high schools. We thus grouped the structural practices that represent organic reforms and also represent a substantial departure from conventional practice (in terms of low frequencies) under the label "restructured." Practices that adhere to the more common and bureaucratic organizational form were grouped under the "traditional" or "moderate" labels. The division between traditional and moderate practices was probabilistic rather than conceptual -- demarcated by a gap in the probabilities that the

reforms were in place in the sample schools (i.e. a drop from probabilities of .56 to .46 from Table 1). Neither category of reforms represents a shift away from traditional forms of instruction, authority, or community.

Classification of schools. Although Table 1 describes reform practices, the design of our study requires that we describe schools in terms of the reforms they currently practice -- particularly those that capture the idea of restructuring as we defined it. In investigating schools which engaged in these practices (grouped as they are in Table 1), we found that the categories were not independent. In their efforts toward reform, schools typically adopt several practices simultaneously. We found, for example, that besides the relatively low probability of schools engaging in any single restructuring practice, those who did were quite likely (a) to adopt more than one restructuring practice, and also (b) to adopt several traditional or moderate practices. On the other hand, we were suprised that more than a trivial number of schools reported that they did not have in place *any* of the 30 practices in Table 1. Based on these findings, we classified schools into three categories, as follows:

- o Unrestructured schools. Twelve percent (97) of the 820 sample schools reported engaging in none of the 30 practices listed in Table 1;
- o Moderately or traditionally reformed schools. These schools (346, or 42 percent) reported that they engaged in one or several moderate or traditional practices, but did not choose to engage in a meaningful number of practices we have classified as restructured (see below);
- o Restructured schools. Schools in this category (377, or 46 percent) reported that they engaged in at least 3 of the restructuring practices on this list, as well as several practices we have listed as traditional or moderate.

Any classification of schools which includes almost half of the sample can hardly be called unusual. However, the categorization of schools in this fashion held together conceptually and statistically.[6] Concentrating on the 12 practices categorized in Table 1 as "restructured," we investigated the proportions of schools engaging in several of them simultaneously. As the number of practices increases, the proportions drop steeply, suggesting that multiple restructuring reforms are difficult to sustain together. One fifth of the sampled schools (20.2 percent) reported engaging in more than 4 practices on this list, and only 6 percent of schools reported that more than 6 (or half the practices) were in place. None of the schools in this sample engaged in more than 9 of the 12 practices. We do not suggest

that adopting many reforms of this type, simultaneously, is necessarily good for students, only that it is rare.

Readers may wonder, "Why three reforms?" The logic we followed in arriving at this figure is explained by the procedure we followed. After having divided the practices listed in Table 1 into the three groups conceptually, we considered whether or not to rank the restructuring practices in a "better/worse" order. As we had neither theory nor evidence to support such a ranking, we abandoned that strategy. We then wondered whether certain reforms might be adopted in common with others. Would the instructional reforms, the authority reforms, or the communal reforms be chosen by schools in groups based on that logic? We found that they were not. We then investigated the *numbers* of reforms that schools might engage in simultaneously. It was clear that schools that practiced restructuring reforms also had several traditional reforms in place, but did the restructuring reforms get adopted in groups? We found that they did. We settled on three restructuring reforms as a reasonable cut point, because it was at this point that the probability of adopting any single reform of this type alone was exceeded by adopting it in concert with others. We revisit the logic of these decisions later in the paper. However, we should preface that discussion by suggesting that a "Choose any three from this list" reform strategy is not supported by the results of this study.

This categorization scheme was used to present descriptive information on students and schools (displayed in Tables 2 and 3). The grouping of schools was also used to construct the major contrasts we used to evaluate the effects of school restructuring. For the multivariate analyses, we created two restructuring dummy-variable contrasts: (1) no restructuring (compared to traditional reforms) and (2) restructured (also compared to traditionally reformed schools).

School size. As explained earlier in the paper, a movement away from large comprehensive high schools is consistent with other efforts to restructure schools along other dimensions. Thus, in addition to the restructuring contrasts we have also investigated the effects of high school size. Because the variable measuring total school enrollment was negatively skewed (i.e., there were a considerable number of small high schools in this sample), we used a logarithmic transformation of the variable in our multivariate analyses. Our reason for focusing on school enrollment effects is that we consider small school size a facilitating

factor for school restructuring. Simply stated, the organic form of school organization is more common, and probably easier to implement, in smaller schools.

Outcomes. **Cognitive outcomes.** We consider five dependent variables in this study: 8th-to-10th grade gains in achievement in four subject areas (mathematics, reading, history, and science) and engagement with school. The achievement measures -- change scores drawn from standardized multiple choice tests -- are quite different from the authentic forms of student achievement that Newmann (1991) advocates as the ideal outcomes on which the effects of school restructuring should be evaluated. As cognitive outcomes, we used gain scores taken directly from the NELS file, constructed as follows. Both the base-year and first-follow-up test scores in each subject were scaled with Item Response Theory (IRT) methods, to adjust for relative item difficulty and other psychometric properties of multiple choice tests (Ingels et al. 1992:125). These difference scores between each student's IRT-adjusted test score at 10th and 8th grade thus have the advantage of representing growth measures that form an *interval scale* for all levels of student ability.[7] In science and social studies, the test items at the base-year and first follow-up were identical. In mathematics and reading, the follow-up tests were "tailored" to students' ability.[8]

There is some controversy about using gain (or difference) scores to measure change within persons between two time points. Although it is intuitively appealing, the statistical properties of gain scores have been questioned. Their validity is said to be suspect, primarily because gain scores are frequently correlated with initial status. In addition, their reliabilities are often modest, primarily because they represent the difference between two measures that are, themselves, less than perfectly reliable. When analyzing two-time point data that measures changes in individuals, the researcher typically pursues one of two alternative strategies. The first is to employ the difference between post-test and pre-test scores as an outcome (i.e., the gain score), and and then perhaps to also include a measure of initial status as a covariate. The second decision (more common in this type of analysis) is to employ the post-test as an outcome, and control for the pre-test and other potentially confounding factors (the covariance model).

After much reading on the topic, many consultations with colleagues, and a considerable amount of exploratory computing, we decided to follow

the advice of John Willett and employ gain scores as outcomes. Willett claims that "the difference score has been inappropriately criticized" and that "the difference score is not the outcast that many critics have claimed" convinced us of the usefulness of gain scores as outcomes, especially in this instance (Willett, 1994:3). The fact that both the pre-tests (i.e., the 8th grade scores) and the post-tests (the scores from 10th grade) have been subject to IRT scaling (and thus the gains were IRT-scaled) convinced us that one common difficulty -- correlation between gains and initial status -- should be eliminated or at least attenuated.[9] As others have found, we discovered that reliabilities of the NELS gain scores were not high, which we discuss below.

Engagement. We were also interested in how attending restructured schools affects students' engagement. Our standardized measure of engagement was created as a factor-weighted sum of eight z-score variables shown to be associated with principal-components factor analysis. These components share several qualities: they measure students' behaviors and attitudes about their current high school classes, they reflect the frequency of these attitudes and behaviors, and they are associated with the same subject areas measured by the NELS tests. These items capture two related constructs: *working hard in school* and *feeling challenged in school*. The composite had high reliability (alpha = .84). Because the 10th grade items measuring engagement were different from those collected in the 8th grade, we structured our analyses of the engagement outcome as a covariance model.

Control measures. All multivariate analyses include two sets of control variables, on students and on schools. Student-level controls include several demographic measures: social class, or SES (standardized at mean (M)=0, SD=1); minority status (Hispanic or Black=1, non-minority=0); and gender (female=1, male=0). Despite our use of gain scores, we also included two controls for initial status (ability and engagement) at 8th grade in analyses for all five outcomes. For the gain score in each subject area, the ability control was constructed as a standardized composite of 8th grade IRT scores in the three other areas. For example, the control for analyses of mathematics gain was constructed as a z-score sum of 8th grade scores in reading, history, and science. A control for 8th grade engagement was included for all outcomes -- both for 10th grade engagement and for the 4 gain-score outcomes. In addition, the covariance analysis

for engagement included a control for general ability at 8th grade (a standardized test composite of reading and mathematics).

School-level controls. Analyses of the five outcomes (engagement and the four cognitive gain scores) included a common set of controls for the demographic and structural characteristics of schools. School demographic controls include average school SES, minority concentration, school sector (two dummy variables contrasting Catholic or independent governance structure to public schools). A reasonable potential alternative explanation for differences in academic gain across the three types of schools focuses on the academic behaviors of students. Thus, we included two controls which characterize the academic character of these high schools. The first, measuring its academic emphasis, was constructed from student reports of average number of mathematics and science courses they took during their 9th and 10th grade years. These student reports were summed and aggregated. The second, a measure of coursetaking differentiation in schools, was constructed as the aggregated standard deviation of the academic emphasis variable just described. School size was also in these analyses, but not just as a control. Descriptions of these variables are in this form, although each was subjected to a z-score transformation for use in multivariate analyses. All measures used in the analyses in this study, the NELS variables from which they were constructed, and the psychometric properties of all composites are detailed in Appendix 2.

Analytic Approach

Descriptive analyses. The focus of this study is on school structure, defined by the three categories described earlier: unrestructured, traditionally reformed, and restructured schools. We present descriptive analyses as group mean differences for all model variables at two levels -- describing students and schools -- in the three categories. We tested observed mean differences between groups with one-way analysis of variance (ANOVA), in each case testing two contrasts: (1) between unrestructured and (2) restructured schools, each compared to traditionally reformed schools. We divided the descriptive analyses hierarchically: (a) for variables describing students (both independent and dependent variables) and (b) for variables describing schools.

Multivariate analyses. This type of study, which investigates how the organizational characteristics of schools influence the students attending

those schools, is often called "school-effects research." Beyond investigating adjusted school means for the outcomes considered here, we were also interested in considering how particular organizational features might affect the *distribution* of these outcomes across the social characteristics of students in each school. Increasingly common in "school effects" studies is the use of hierarchical linear modeling (HLM), a statistical approach we use here. Because the statistical theory and methodological approach of HLM are described in detail elsewhere (Bryk and Raudenbush, 1992; Lee and Bryk, 1989a), we only a very brief introduction to statistical models used here is provided in Appendix A. The technique's growing popularity for quantitative studies in education suggests that unfamilar readers who are interested in the theory underlying HLM could become acquainted with the technique, perhaps using the readings just cited. For those interested in more detail on the application of HLM to the analyses described here than we provide in Appendix A, we suggest contacting us directly. In this paper our discussions focus on the substantive rather than the statistical implications of the results.

Results

Descriptive Differences Among Students in Schools With Differing Structures

Table 2 displays descriptive differences between students attending unrestructured, traditionally reformed, and restructured schools. The number of sampled students attending unrestructured schools (n=1,280) is considerably below those in either traditionally reformed (n=5,353) or restructured schools (n=5,161).

Insert Table 2 about here

Dependent variables. For the outcome measures, there is a general pattern in the group mean differences between the three types of schools: the means of students attending unrestructured schools are significantly lower than those in traditional schools. Although mean achievement gains are somewhat larger in restructured than traditionally reformed schools, this difference is statistically significant only for science. In general, the magnitude of mean differences in achievement gains is quite small (the largest group differences are no more than .10 or .15 SD).

Independent variables. Students' social class (SES) and minority status distributions follow a pattern similar to that for the outcomes. Students in unrestructured schools are of lower SES and are more likely to be minority group members than students in traditionally reformed schools, whereas students in restructured schools are more advantaged on these measures than their peers in traditionally reformed schools. The same pattern applies to students' ability status and engagement at 8th grade -- students in unrestructured schools rank below, and those in restructured schools rank above, those attending schools defined as traditionally reformed. The proportions of female students is statistically equivalent for the three types of schools. In general, the magnitude of group differences in the control variables, while not large, is larger than the outcomes -- between .15 and .35 SD.

Descriptive Differences Among Schools With Differing Structures

Descriptive differences on several school characteristics between schools defined as unrestructured, traditionally reformed, or restructured are presented in Table 3. The number of unrestructured schools (n=97) is less than either traditionally reformed (n=346) or restructured schools (n=377).

Insert Table 3 about here

School demographics. Reflecting the characteristics of students in these three categories of schools shown in Table 2, in terms of social demographics (average SES, minority enrollment), unrestructured schools are significantly disadvantaged compared to traditional schools. Restructured schools are advantaged compared to the same group in terms of average SES (a .4 SD difference between restructured and unrestructured schools). The distribution of these schools by sector explains at least some of these social background differences. Although private schools represent 14 percent of the school sample, very low proportions of Catholic and elite private schools (i.e., members of the National Assocation of Independent Schools, or NAIS) schools are classified as unrestructured (2.3 and 1.5 percent, respectively). A relative high proportion of restructured schools are private (9.9 percent are Catholic, 9.7 percent are independent). In fact, 52 percent of Catholic schools, and 74 percent of independent schools

are classified as restructured by these criteria. Of course, the overwhelming majority of all groups (over 80 percent) are public schools.

Academic emphasis and size. Also of interest is the academic character of the schools, defined by students' coursework in mathematics and science. Unrestructured schools are typified by significantly less, and restructured schools by significantly more coursetaking. Coupled with differences in academic emphasis are differences in the homogeneity of coursetaking patterns. Unrestructured schools are characterized by more variability in coursetaking than traditionally reformed schools, whereas restructured and traditionally reformed schools have similar variability. The school size distribution does not follow the same pattern as other school characteristics. Traditionally reformed schools are smaller than either restructured or unrestructured schools, although the average size of schools in these groups are reasonably similar (between 1,000 and 1,500 students). We note that the distribution of school size is positively skewed, i.e., there are more small than large high schools (although, of course, most students attend large schools.

There are considerable differences between the students and schools who attend schools which we have characterized as unrestructured, traditionally reformed, and restructured -- demographic differences among students and schools, sector differences, differences in initial cognitive and non-cognitive status, and differences in the academic emphasis of the schools. In fact, differences across the groups between the control variables we chose to include in our multivariate analyses exceed differences in students' cognitive gains.

Multivariate Analyses

Unconditional HLM models. In any HLM analysis, the first step involves examining the variances of dependent measures in a model which includes no control variables for either schools or students. Here we partition the variance in each outcome into its within-school and between-school components, and estimate reliabilities. The effects of school restructuring may be evaluated only on the proportion of each outcome's variance which occurs between schools. Table 4 displays results for these unconditional HLMs.

Insert Table 4 about here

For each outcome, the proportion of variability between schools (i.e., the *intraclass correlation*) is modest (20 percent or less).[10] Although this proportion is adequate to proceed with multilevel analyses, it is clear that most of the variation in engagement and achievement gains is between students within schools, rather than between schools. The intraclass correlation for in engagement is lowest (.13), and highest for science gain (.20). As expected, HLM-estimated reliabilities for these gain scores (and also for engagement) are quite low (ranging from .27 for engagement to .50 for science gain).[11] It is clear that these outcome variables have less than optimal psychometric properties. Nonetheless, the fact that they measure change in each curriculum area for each student over the first two years of high school suggests that as a set they represent ideal outcomes on which to evaluate the effects of school restructuring on *student learning*. We acknowledge a tradeoff between less-than-perfect measures in a statistical sense for almost-perfect measures in a conceptual sense.

Within-school HLM models. Table 5 displays within-school HLM models for each of the five outcomes in this study -- school engagement and gains in mathematics, reading, history, and science.[12] Conceptually, within-school HLMs are somewhat analagous to many small within-school OLS regressions -- one in each of the 820 schools. The within-school model for each outcome contains almost identical control variables. Each model estimates *within each school* the effect of student demographics (social class, minority status, and gender) and of engagement at 8th grade. The ability control, although somewhat different for each outcome, is meant to control for ability at the point of high school entry. In all HLM analyses (Tables 5 and 6), the magnitudes of effects are displayed as effect sizes [ES] (i.e., SD units).[13] Because of small within-school sample sizes (averaging about 10 students/school), our HLM models estimate only one "slope as outcome," the SES effect on each outcome, as a random variable in between-school models (i.e., SES is "free"). Variability in the other control variables is constrained to vary only within each school (i.e., each of these variables is "fixed." Procedurally, this means that the SES variable is centered around the sample mean (i.e., it is a z-score variable), whereas the other controls are centered around their respective school means.

Insert Table 5 about here

Many of the control variables' effects are quite strong. SES is positively related to 10th-grade engagement (ES=.77) and to gains in mathematics and science (ES about .13), but only marginally related to gains in reading and history. Minority status is positively related to engagement (ES=.82), but negatively related to gains in reading (ES= -.23) and science (ES= -.68) and only marginally related to gains in history (again, negative). Gender effects are not consistent: while females show very positive engagement (ES=1.22), males are favored in science gains (ES=-.37). Other gender effects, while negative for females, are small and non-significant. Effects of 8th grade engagement are positive and statistically significant on all outcomes. The effect is, unsurprisingly, very large on 10th grade engagement (ES=1.16).

Effects of cognitive status at 8th grade on these outcomes are modest in size on achievement gains. Ability is most strongly and positively related to 10th-grade engagement (ES=.48), and moderately related to gains in science (ES=.21). On the other hand, ability is significantly, modestly, and negatively associated with gains in mathematics and history (ES less than -.10). Ability is unrelated to gains in reading. We may conclude that initially more able students become even more engaged with school life in high school, and learn somewhat more in science. However, initially less able students seem to learn a bit more in mathematics and history. Negative correlatons between prior ability and gain are well known. The small magnitudes suggest that IRT scaling has been successful.

Between-school HLM analyses. The final HLM models focus on ten school-level outcomes, each adjusted for the set of within-school controls shown in Table 5 (engagement, ability, SES, minority status, and gender). The outcomes are of two types: (1) a set of five adjusted school means for engagement and gains in each curriculum area; and (2) a set of within-school estimates of the relationship between SES and each outcome. The outcomes in set 1 -- school mean engagement and cognitive gain -- are what might be described as "effectiveness" parameters. That is, schools which are high on these parameters are more effective (students are more engaged, learn more). By our definition, school characteristics which are *positively* associated with these school means may be seen as typifying "good schools" (i.e., effective schools). We describe the outcomes in set 2, the relationship between social class and cognitive gains or engagement, as "equity"

parameters.[14] Thus, we expand our definition of "good schools" to those which simultaneously have a high effectiveness parameter and also a low equity parameter (schools with high learning rates that are equitably distributed among students from differing SES levels). Here were would see characteristics of schools as "good" if they were *negatively* associated with the SES slopes.

The full HLM models, which include estimates of the effects of the school characteristics described earlier on these 10 outcomes, are shown in Table 6. Although each analysis also includes the full set of within-school controls shown in Table 5, we omit them from the results displayed in Table 6. Effects are almost identical to the values shown in Table 5. In order to highlight the effects of primary focus in this study, we divided the school-level independent variables into two groups: those tapping school restructuring (below the line) and those representing demographic and academic controls (above the line). Each column in Table 6 represents an HLM analysis on one dependent variable. Results for school means on each outcome (the "effectiveness parameters") are found in the upper section of Table 6, and the social distribution of these outcomes (the "equity parameters") are shown in the bottom section. We note that these effects are estimated simultaneously in a fully multivariate HLM model. To discuss tests of the hypotheses posed earlier, we organize our discussion around effects of particular independent variables across all the outcomes (i.e., by row) rather than by column (i.e., for each outcome).

Insert Table 6 about here

Effects of School Restructuring

Testing Hypotheses 2 and 3: School restructuring contrasts. Recall that schools classified as "not restructured" and those classified as "restructured" are each compared to those classified as "traditionally reformed." The pattern of effects is consistent -- restructured schools have strong, positive, and significant effects on their students' cognitive gains in the first two years of high school. Effect sizes range from .35 on gains in history to .59 on gains in science. These findings confirm Hypothesis 2 -- that students attending restructured schools would learn more and be more engaged with school. The effects of unrestructured schools

(Hypothesis 3), while not as large, are consistently negative. Students attending unrestructured schools are less engaged, and gain less in cognitive areas, than those in traditionally reformed schools. Effect sizes, in the .1-.2 range, are generally smaller than those for restructured schools. Thus, Hypothesis 3 -- which suggested positive effects on students for traditionally compared to unrestructured schools -- is confirmed, although less dramatically than Hypothesis 2.

Noteworthy is the highly stratified distribution of these outcomes by SES. The large positive coefficients representing average SES slopes at the top of the lower panel of Table 6 suggest that high-SES students are significantly advantaged on all outcomes. Particularly strong are the relationships of SES with engagement (ES=1.59) and with science gain (ES=.61). Again, the pattern of school restructuring effects on social equity in these outcomes is consistently favorable. In general, restructuring is negatively related to these outcomes (i.e., restructured schools are more equitable), with effect sizes large (ES=-.54) on the social distribution of engagement and moderate (ESs of .3-.4) on the social distribution of achievement gains. Displaying a similarly consistent pattern, unrestructured schools are more stratifying in the distribution of achievement gains by SES (i.e., effects are positive and significant, ranging from .18 on the social distribution of mathematics gain to .54 on the distribution of reading gain). The findings on social equity also confirm our hypotheses.

The consistent pattern of effects on learning and engagement shown in Table 6 suggests that school restructuring as conceptualized in this study has important consequences for the students who attend institutions organized in these ways. High schools that engage in the practices we have called "restructured" fit the set of double qualifications for "good schools" -- schools like this are simultaneously more effective and more equitable in terms of student engagement and learning.

Testing Hypothesis 4; School size. The results from Table 6 demonstrate consistently and conclusively that students do better in smaller schools, findings which confirm Hypothesis 4. Across all "effectiveness" parameters -- engagement and cognitive gains -- the direct effects for larger schools are negative and significant. Effects on cognitive gains are particularly notable (range of ES from -.32 on reading gains to -.39 on mathematics gain). In addition to consistently higher levels of learning and engagement for students in small schools, smaller school size is also

related to more social equity in these gains within schools (the lower panel of Table 6). This is demonstrated by the pattern of positive and significant effects of school size on the relationship between SES and achievement gains (the "equity" parameters). The magnitude of significant effects ranges from small (ES=.03) on the slope of SES on mathematics gain to moderate (ES=.34) on the slope of SES on reading gain. To our knowledge, the effects of school size on the level and distribution of cognitive performance shown here are among the largest and most consistent of any published studies on the effect of school organization.[15]

Other noteworthy school effects. **Academic emphasis.** The primary purpose for including variables that measure the level and variability of academic emphasis of high schools was to control for coursetaking. Although raising levels of academic coursetaking is not part of the restructuring effort investigated in this study, there have been reform efforts directed to increasing enrollment in academic courses for high school students in recent years. Effects on the outcomes in this study are consistent -- academic emphasis is positively and significantly related to all outcomes (ESs range from small [.13] on history gain to large [.56] on science gain). In general, more variability in coursetaking within schools is negatively associated with these outcomes, especially engagement (ES=-.17). In one instance, however -- gains in science -- more variability in coursetaking has a *positive* effect (ES=.56). Not all effects on the equity parameters are statistically significant, but the pattern is generally similar to the restructuring effects. Schools with more academic emphasis are more equitable environments (significant effects range from moderate [-.27] on math gains to large [-.94] on engagement). Coursetaking variability effects, although small, are positive -- suggesting that more variability within schools is associated with social stratification in learning and engagement.

School sector. Results in Table 3 showed that both types of private schools (Catholic, NAIS) were somewhat more likely to be classified as restructured. Private secondary schools are also typically considerably smaller than public schools, academic emphasis is higher, and coursetaking variability lower (Coleman, Hoffer, and Kilgore, 1982). Including controls for school sector -- above and beyond the practices and structure typical of them -- was meant to address a possible alternative explanation for the findings described here. The residual effects of school sector in Table 6

are not consistent. For example, while NAIS schools have a large positive effect on their students' engagement, there is no residual Catholic school effect. Catholic schools show a significant positive effect on math gain but not on other cognitive gains. NAIS effects on cognitive gains are all positive, some large, but none significant. Similarly inconsistent are sector effects on equity. Although effects are often large and usually negative (suggesting more social equity in such schools), few are statistically significant. We conclude that including several important features of private school organization in our models, (e.g., smaller size, more academic emphasis) has largely explained away the residual sector effects on effectiveness and equity shown in other school effect studies. Thus, we draw no substantive conclusions from the sector effects shown in Table 6.

School social composition. In the gain-score analyses, school SES and minority concentration are generally unimportant. One exception is the moderate effect of school SES on science gain (ES=.27). Our interpretation rests on our view of average SES as a proxy measure of school resources. The finding suggests that students learn more science in schools with higher resources. Given the expense of laboratories and science equipment, this result makes sense. Schools enrolling more affluent students also appear to be much more stratifying in their distribution of engagement (ES=.63). Although other compositional effects on the equity parameters are sizeable in some cases, none reach statistical significance. We thus conclude that school social composition generally has little effect on student learning, above and beyond the considerable effects of individual students' social background on achievement gains (Table 5). One exception is science learning, where resource differences between schools are important.

Discussion

Summary of the Findings

Gains in achievement. Using the theoretical contrast of school organization between the organic and the bureaucratic, we conceptualized school restructuring along several dimensions in this study. One dimension focused on practices defining the organizational structure of high schools. We used the form of the practices we considered (organic or bureaucratic), as well as the idea of "conventional departure from conventional practice,"

to categorize schools as either restructured, traditionally reformed, or unrestructured. A second dimension focused on high school size. Here we considered smaller size a positive feature of school restructuring in an organic direction.

Considering school restructuring in terms of either school structural practices or school size, the pattern of effects on the cognitive learning of students in the early years of high school found in this study is clear, consistent, and strong. Students attending schools that are restructured learn more in mathematics, reading, history, and science achievement. Those who attend schools which are not restructured learn less. Equally important are the findings about school size. Students in smaller schools gain more in these important areas of the curriculum. These findings are *net* of differences in schools' academic and social character, as well as the academic and social characteristics of their students.

Social equity in student learning. Besides a focus on students' achievement gains in the early high school years, we also considered how different forms of school restructuring affect the distribution of these gains within each school among students from different social backgrounds. Effects here are also consistent. Achievement gains in the four curriculum areas we considered are more equitably distributed in smaller schools. Restructured schools are more equalizing environments in terms of the social distribution of cognitive gains, and unrestructured schools are more stratifying environments. The consistent pattern of findings allows us make quite unequivocal statements about the organizational structure of high schools: students learn more in schools which are restructured (by our definition) and in smaller high schools. Schools organized in this way are also more equitable environments in terms of the distribution of cognitive learning. Conversely, in schools that have not restructured, and in larger schools, students learn less and learning is more stratified. While general levels of learning are lower in those schools, socially disadvantaged students learn even less.

Student engagement. Our findings regarding students' engagement with academic learning in the early high school years are consistent with the findings for cognitive development. Students are more engaged in smaller schools, and engagement is more equitably distributed in restructured schools. Students attending schools that are not restructured are less engaged. The findings on engagement take into account students' academic

status as they entered high school and their prior engagement with school.

Caveats. We remind readers of cautions we mentioned earlier about these outcome measures. There are two caveats about the modest statistical properties of the outcome variables examined here: (1) low proportions of variability between schools, and (2) low reliability of gain scores. Since only between-school variabililty in outcomes may be explained by school organizational differences, we admit that our analyses show generally strong effects on the 13 to 20 percent of variability in engagement and learning that is between schools. We also suggested the need to balance the measures' modest psychometric properties against the conceptual clarity of their meaning. The substantive interpretation of our caveat about modest reliabilities is somewhat different. The IRT-adjusted gain scores used in these analyses measure exactly what we were interested in investigating -- students' cognitive learning in their first two years of high school. Despite low reliabilities, the conceptual clarity of these outcomes make the study's findings unusually unequivocal. Thus, these effect probably represent lower bounds of the effects we might obtain if our outcomes were measured more reliably.

Another caveat concerns timing. Having longitudinal data available on students allows us to have some confidence that the outcome variables we employ capture the effects of students' experiences in high school -- a strong advantage of the NELS data. However, we have little information on the dates when the reforms considered in this study were implemented. This raises the question of whether restructuring preceeded or succeeded the high-learning environment of the restructured high schools. Of course, the structure of our analyses supposes that reform came first, and student learning was the result. However, we admit to an alternative causal order. Hopefully, with the next wave of NELS data can untangle the confusion about temporal sequencing.

Revisiting the Concept of Restructuring

What sort of conclusions should be drawn from these results about how American schools should be reformed? Earlier in the paper, as we described the construction of our measures of school restructuring, we argued that the reforms we grouped together under the heading of "restructuring" were conceptually related, in that each represented a move toward an organic organizational form and away from a bureaucratic form. While admittied

that the restructuring reforms were aimed at changing somewhat different school functions (instruction, automony, or community), we also argued for their conceptual commonality. That our grouping of structural practices proved to have strong and consistent effects on important student outcomes provides empirical support for some internal consistency among the reforms.

How many reforms? We are troubled that the results might lead practitioners or policymakers to a conclusion we don't feel is justifiable: "Pick any three reforms from this list, and your high school will be more effective and more equitable." Earlier in this paper we provided a rationale for our decision to classify schools as "restructured" if they had implemented at least three of these reforms. In an effort to explore the implications of this descision, we conducted a sensitivity analysis -- where we estimated the effects of schools implementing particular numbers of restructuring reforms. The results of the sensitivity analyses are displayed in Figures 1 and 2. The results on the "effectiveness" parameters, i.e., gains in achievement and engagement, are shown in Figure 1. Figure 2 presents graphical results for the "equity" parameters, i.e., the social distribution of achievement and engagement. All outcomes are included together in each graph.[16]

Insert Figures 1 and 2 about here

Sensitivity analysis results clearly demonstrate that the simultaneous implemention of many restructuring reforms (more than 4, on most outcomes) is not advantageous in terms of either effectiveness (Figure 1) or equity (Figure 2). These results also suggest that implementing reforms one or two at a time may not be advantageous. In terms of *numbers*, these results support our decision to use three reforms as a cut point. It seems clear that schools with a commitment to restructuring as we have defined it should decide on a modest number of reform strategies, should work hard to see that these reforms are engaged profoundly in the school, should continue their commitment to those particular reforms over a sustained period, and should not attempt too many reforms simultaneously.

These findings -- particularly those indicating that adopting large numbers of reforms is not advantageous -- are supported by two recent studies that examined school restructuring in local contexts. The study in Jefferson County, Kentucky (Kyle 1993) reported that schools in the early

stages of reform, particularly those who tried a wide range of reforms, had less advantageous outcomes for students than schools which engaged in no reforms, and considerly less positive outcomes than schools with sustained commitments to fewer reform strategies. The Kentucky findings extended to schools at the elementary, middle, and secondary levels. Anthony Bryk and his colleagues studied the emerging effects of the Chicago school reform in that city's elementary schools (which are K-8). They described some schools as "Christmas trees": "'showcase' schools with many new programs, multiple 'add-ons' with little coordination, and little attention to strengthening the organizational core" (Bryk et al. 1993:15). Such schools, which the study characterized by their "unfocused academic initiatives," compared unfavorably with other schools which practiced more systematic approaches to school restructuring. The latter group was described as having a "shared, unified, coherent school vision; changes in place that affect most classrooms; extensive staff development; high teacher commitment, and institutionalized environmental changes" (p.15). It seems clear that a strategy of attempting to embrace too many reforms -- perhaps to give the appearance of climbing on the "reform bandwagon" -- is counterproductive.

What do the restructuring results mean? We encourage readers to define the policy implications of this study broadly, rather than narrowly and prescriptively. Clearly, something important is going on inside the schools we have collectively labeled as "restructured," because students in those schools demonstrate more learning which is also more equitably distributed. In our opinion, the findings provide solid empirical support for the value of school reforms which move schools in the direction of a more organic organizational form, and move them away from the bureaucratic form that has characterized the comprehensive high school for a century. The results also suggest that schools should target their reform efforts around a modest number of practices of the organic type -- practices which should probably be adopted neither singly and serially, nor in large numbers to "showcase" a school's superficial commitment to reform. Our results also provide solid support for the movement toward smaller learning environments. We clarify our interpretation of the results for school size below.

We admit to some puzzlement about our lack of empirical success in grouping restructuring reform practices into those which are targeted on instruction, on autonomy, and on community. Thus, we caution readers against trying to extract specific recommendations from our results about

which of these reforms, considered individually, are better than others, or even about *how many* reforms are optimal. Our work on this topic is ongoing. We intend to pursue the investigation of the internal workings of those schools which we have classified as "restructured," and to use the next wave of data from NELS (when students are high-school seniors) to further investigate the effects of school restructuring on students in these high schools. We are nevertheless confident, given the evaluative framework in which this study was carried out, to conclude that something important is going on in restructured schools, and that students who attend such schools are advantaged in important ways -- at least in the first half of their high school experience.

Is reducing school size really the issue? Although the restructuring practices investigated here have some implications for school reform policies (at least in terms of the conceptual direction of reform), does it make sense to consider the findings about school size in a similar "policy light"? That is, would reducing the size of high schools really "cause" students to learn more? Although the structure of our analyses -- which estimate the *direct* effects of school size -- might suggest that this would be so, we would not draw that conclusion from our results. In fact, we don't know whether any schools had recently reduced their size as a reform strategy. Rather, we interpret the positive findings for small schools as indicating that enrollment size acts as a facilitating or debilitating factor for other desirable practices. For example, we know that colleagiality among teachers, personalized relationships, less differentiation of instruction by ability -- to name a few organization features of schools -- are more common (and probably easier to implement) in small schools. Reducing school size, while a potential structural reform in its own right, would not increase student learning per se.[17]

We suggest that school size can only have an *indirect* effect on student learning and engagement. Were we to introduce a set of school organization measures such as colleagiality, personalized relationships, and the like into our analyses, the magnitude of the direct school size effects seen here would surely decline. We content that such a finding, although hypothetical, would not weaken the substantive importance of our findings about school size; rather, it would help us understand the complex organizational mechanisms through which school size affects students. Given the current fiscal constraints surrounding American education, it is unlikely that new

and small high schools will be constructed, no matter how strong the empirical link between "smallness" and learning. Rather, we believe that the "school within a school" reform -- already embraced by 15 percent of American high schools (shown in Table 1) -- is a feasible and cost-effective way to accomplish this structural reform and facilitate other useful organizational changes.

Survey data and research on school structure. As mentioned in a study of restructuring in middle-grade schools (Lee and Smith, 1993), the use of survey data to investigate the effects of school restructuring is limited in several respects. Many practices in our "restructured" category (Table 1) are hardly "cutting edge." We were distressed to find that so few American high schools engage in practices such using parent volunteers in a school, employing a focus on cooperative learning, team teaching across disciplines, mixed-ability classes in mathematics and science, teachers' common planning time, or flexible time for classes. Despite growing research support for the effectiveness of these practices, few American high schools were doing them in 1990. Thus, one discouraging finding is how bureaucratic the structure of the American high schools is. In real terms, restructuring as we would like to define it -- instead of how we have actually defined it -- is quite rare in American secondary schools.

Although survey methods are well suited to investigate the *frequency* of organizational practices, they are not particularly well suited to study their *implementation*. Although NELS fortunately includes reports from principals about whether the practices occur in each high school, it includes no information about intensity, pervasiveness, or support. Thus, in a school reporting that it offers independent study in math or science or that it has inter-disciplinary teaching teams, for example, these practices might actually affect only a very small number of students and/or teachers. Most teachers could oppose these reforms, with only a small and self-selected group actually doing these things. Moreover, we have no idea whether the practices which in theory involve students (e.g., mixed-ability classes, independent study) were actually engaged in by the sample of students in the NELS study. With an average of less than 15 sampled sophomores in each school in our sample, unless the practice were almost universal (which is certainly possible) sampled students might not have actually experienced the programs reportedly offered by the school.

Another disadvantage of using NELS data to study school restructuring

concerns the lack of information about *when* the practices were initiated. Thus, we have little idea of whether these practices represent recent innovations in any school, whether they were in place during the sampled students' two years in the high school, or whether they were adopted the day before the survey questionnaire was completed. Thus, we lack historical information about restructuring in these high schools -- actually, we don't know whether the practices represent *restructuring* at all.

The weaknesses of survey data are less applicable to our other featured structural variable -- school size. It is unlikely that school size changes much over two years (with the possible exception of the 15 percent of schools in Table 1 which describe themselves as "schools-within-schools"). School size is well measured in this survey.[18] Moreover, having large and nationally representative samples of students and schools represents an ideal venue for the investigation of size effects.

The very constraints of using survey methods to study the phenomenon of school size may underscore the importance of our findings. High schools that engage in as few as three of the practices that we have classified as "restructured" show quite powerful effects on their students' learning and engagement, and the rather small proportion of American high schools that engage in none of the practices -- even the rather ordinary ones -- have negative effects on their students. Thus, the school effects described in this study probably represent lower bounds for the actual effects of school restructuring on students' engagement and learning.

Breaking out of the bureaucratic mold. Why are these practices uncommon in American high schools? This question echoes a common theme -- why is profound change so difficult in our schools? As we suggested earlier, the answer lies in part in the entrenchement of rational-bureaucratic structural model in public secondary schooling in the United States. Bureaucratically organized schools are typified by a formal division of adult labor into specialized tasks, by rule-driven social interactions, by limited discretion for individuals, and by authority defined through roles and rules. By contrast, the organic or communal persective on schooling views these institutions as "small societies" where social interactions are informal, discretion is wide, and roles are defined by the people inhabiting them.

As we have stated, the restructuring practices described in this study, as well as small school size, are features of schools that are more typical

of the organic than the bureaucratic organizational model. The findings of this study sustain the empirical and theoretical contrast between bureaucratic and organic (or communal) organizations developed more fully elsewhere (e.g., Bryk, Lee, and Holland 1993; Lee, Bryk, and Smith 1993; Rowan 1990). We have argued here that this contrast is appropriate to impose on arguments about school reform. Shaking high schools out of the bureaucratic mold which has typified them since their beginnings around turn of the century has proven to be a difficult task. We hope that our results are seen as providing empirical support for a movement to restructure schools toward the "small society" model.

Technical Notes

1. This point was made by Lee and Bryk (1989b) about the selection of appropriate statistical controls in HS&B data. They stated that "there are certain categories of student outcomes...for which it is reasonable to hypothesize that school effects occur early and are in place by the end of students' sophomore year" (p.648).

2. Of the 21,126 NELS 8th graders described by NCES as eligible for being followed up, 17,424 (or 81.2 percent) were located in their 1,508 high schools and thus have student-level panel weights (i.e., this sample of students was selected at both waves and had complete data). Our sample of students has "survived" several subsequent data filters (see footnote 3).

3. Two types of data filters were applied to select our sample from the 17,424 students and 1,508 schools in the first follow-up. First, we selected students with close to complete data of the types we needed for our analyses. Second, we selected all retained students in the types of schools we wanted to analyze. The filters, with resulting unweighted samples sizes for students and schools, were as follows: (a) students with base-year and first-follow-up test scores (leaving 16,334 students in 1,448 schools); (b) students in high schools with school-level data (reduced to 15,550 students in 1,267 schools); (c) limiting our sample to public, Catholic, and independent (NAIS) schools (down to 13,603 students in 1,120 schools); and our final filter, (d) sampling only schools with at least 5 NELS-sampled students in them (11,794 students in 820 schools). The data filter dropped mostly private schools with one or two NELS students in them.

   The motivation for data filter (d) was to assure adequate within-school sample sizes for the hierarchal analyses described below. Examination and testing of differences between retained and non-retained samples of students showed that, except for school sector, demographic differences were modest. Moreover, the retained sample is slightly *less* advantaged that the non-retained group (the result of dropping many private school students). This suggests that the bias introduced by these data filters actually favors less select students. With only modest bias resulting from our data filters, particularly in this direction, we are reasonably confident in generalizing to the population of 1990 American high school sophomores and their high schools.

4. We developed the technique we used to construct these weights with the assistance of staff from the Sampling Division of the University of Michigan's Institute for Social Research. The construction method combined probabilities drawn from two major sources: (a) aggregations of the inverse of each student's follow-up weight (the NCES-computed weights for those students who also had base-year data); and (b) the probabilities of students in each school having experienced their 8th grade year in NELS public, Catholic, independent (NAIS), or other private schools, weighted by the total enrollment of each high school. Extreme values in our set of constructed school weights were trimmed, and the resulting weights were adjusted to a mean value of 1 for our sample of 1,199 schools.

We tested the validity of our technique by first constructing similar "pseudo-school weights" for the base year data. We then ran a typical hierarchical linear model from our base-year study (Lee and Smith, 1993) under 3 conditions: (a) unweighted, (b) using the NELS base-year school weights, and (c) using our constructed "pseudo-school weights". HLM parameter estimates using our constructed weights were quite similar to those obtained using the NELS school weights (and more divergent compared to the unweighted parameter estimates). These results supported our decision to proceed with constructed school weights for the NELS first follow-up. Details about construction of the weights and about our testing procedure are available from the authors.

5. The probabilities listed in Table 1 are not raw probabilities for the 820 schools in the sample, but rather represent the probability that "an average high school" would engage in each practice. The method to compute these adjusted probabilities is as follows. As the average number of practices engaged in by these schools was 12 (out of 30), we defined "the average high school" for this investigation as one which had adopted 11, 12, or 13 of these practices. We separated out those high schools, and re-estimated the frequency of each practice for these *average high schools*. It is these frequencies, represented as proportions, that are the probabilities displayed in Table 1. Compared to raw probabilities, the order of the variables is identical.

6. How do the proportions of students and schools in the three types of "restructuring" schools changed as a result of the data filter dropping schools with fewer than 5 NELS students? We found that the proportions of schools and students in each of the three categories was almost identical before and after this filter was applied, suggesting that this restriction introduced no bias in terms of our major hypotheses.

7. An IRT score is a non-linear transformation of the number of items correct, adjusted for three item parameters (differential difficulty, discriminating power, and the likelihood of correctly guessing in a multiple-choice format). IRT scaling has the effect of expanding the scale at its extremes. Two consequences of this rescaling are that the estimated gains between 8th and 10th graders will appear (a) larger for high-ability students in an IRT than a raw-score metric and (b) smaller for low-ability students. A major advantage of this scaling is that it eliminates a major stated disadvantage of gain-score analysis -- that those at the bottom will artifactually appear to gain the most (a variation on the "regression to the mean" theme). See Lord (1980) or Hambleton (1989) for more detail on item response theory.

8. The "tailoring" of the reading test (which had two forms) was meant to counteract the time burden of reading passages for slow readers in the short time frame provided for testing. In the case of mathematics, the purpose of "tailoring" (three different forms) was so that the tests would be responsive to the diversity of exposure to coursework which could be expected by 10th grade (especially algebra and geometry). Students were selected for forms of the tailored tests based on their 8th grade scores in the same curriculum area. In the case of the "tailored" tests, IRT scaling becomes especially important, since the relative difficulty of different test items is adjusted for with IRT.

A core of common "anchor" items among all forms of the test made IRT scaling possible (Ingels et al., 1992:Appendix I, p.17-18.

9. In general, the correlations of initial status with gains were very modest (under .2). Except for gains in science, correlations were negative.

10. The intraclass correlation is computed as the proportion of total variablity (tau + sigma-squared) represented by the between-school variability (tau). These figures have been adjusted for attenuation due to low reliability, in that sigma-squared (pooled within-school variance) is adjusted for the reliability of each outcome.

11. Reliabilities estimated with HLM are not the same as the classic Cronbach's alpha, although both aim to estimate the degree to which the observed score measures the "true score." Although Cronbach's alpha is an estimate of internal consistency for a composite measure, the HLM reliability estimate is a function of variability in means across schools and the within-group sample size (Bryk and Raudenbush, 1992).

12. As suggested in the footnotes in Tables 5 and 6, effects are presented in SD units, in order to allow comparison across outcomes with different metrics. Thus, group mean differences on these tables are not numerically comparable to those presented in Tables 2 and 3.

13. Eighth grade engagement, SES, and ability are normally distributed z-score variables; minority status and gender are dummy-coded variables. Thus, dividing by the HLM-estimated standard deviation in each outcome results in the standard effect-size metric. In common usage, effect sizes of .1-.2 SD are small effects, .2-.5 SD are medium, and over .5 SD are large (Rosenthal and Rosnow, 1984:360).

14. We borrowed the notions of "effectiveness" and "equity" parameters in HLM from Chapter 1 of Bryk and Raudenbush (1992).

15. We investigated the possibility that these substantial effects of school size could be an artifact of the weighting scheme we devised, since weights were constructed proportional to reported school enrollments. All HLM analyses in this paper were, thus, run with and without weights. Consistently, the patterns of effects was sustained (in terms of probability levels in statistical tests), although the magnitudes of some coefficients changed. There was no consistent direction to the changes in coefficient magnitude.

16. The analyses whose results are displayed in Figures 1 and 2 were conducted with HLM methods on the set of outcomes, using identical within-school controls as displayed in Table 5 and used in Table 6. The school-level HLM model on each outcome contained a series of dummy variables classifying schools by the number of restructuring reform practices they reported -- from 1 to over 6. The comparison group for these dummy variables were the schools which did not engage in a single restructuring reform (but they may or may not have traditional reforms in place). Unweighted distributions of schools were: 0 reforms (the comparison group): 20.6 percent of schools; 1 reform: 17.0 percent; 2 reforms: 16.5 percent; 3 reforms: 13.7 percent; 4 reforms: 12.1

percent; 5 reforms: 9.4 percent; 6 reforms: 5.0 percent: over 6 reforms; 5.9 percent. This distribution of reforms argued against constructing a "number of reforms" continuous variable.

17. Although our findings provide strong empirical support for smaller schools, we did not investigate a "lower bound" below which high school size should not go. As the school size variable was transformed, the results here suggest a linearity in relationships which may not be entirely reflective of reality. It is also unclear whether the 15 percent of principals who reported "schools-within-schools" reported the "within-" or the overall school size. Thus, our results should be seen as providing general support for reducing the size of high schools, rather than suggesting specific figures about the optimal size of high schools.

18. At the time this study was conducted, use of the NELS first-follow-up school data file was restricted to researchers who were licenced by the National Center for Education Statistics for use of confidential data (the first author holds such a licence) . As the data become available as a public-use file, the continuous measure of school size will still be restricted to licencees only It is the continuous version of this variable that we employed here.

References

Alexander, K.L., Cook, M., and McDill, E.L. (1978). Curriculum tracking and educational stratification: Some further evidence. American Sociological Review, 43(1):47-66.

Anderson, C.S. (1982). The search for school climate: A review of the research. Review of Educational Research, 52(3):368-420.

Barker, R. and Gump, P (1964). Big School, Small School: High School Size and Student Behavior. Stanford, CA: Stanford Univ. Press

Berends, M. and King, M.B. (1994). A description of restructuring in nationally nominated schools: Legacy of the iron cage? Educational Policy, 8(1):28-50.

Bidwell, C. and Kasarda, J. (1975) School district organization and student achievement. American Sociological Review, 40(1):55-70.

Blau, P.M. and Scott, W.R. (1962) Formal Organizations: A Comparative Approach. New York: Chandler Publ.

Boyer, E.L. (1983) High School: A Report on Secondary Education in America. New York: Harper and Row.

Bridges, E.M. and Hallinan, M.T. (1978). Subunit size, work system interdependence, and employee absenteeism. Educational Administration Quarterly, 14:24-42.

Brown, R. (1993). Measuring the Progress of Systemic Education Reform. Denver, CO: Education Commission of the States.

Bryk, A.S., and Driscoll, M.E. (1988) The School as Community: Theoretical Foundations, Contextual Influences, and Consequences for Students and Teachers. Madison, WI: National Center on Effective Secondary Schools, Univ. of Wisconsin.

Bryk, A.S., Easton, J.Q., Kerbow, D., Rollow, S.G., and Sebring, P.A. (1993). A View from the Elementary Schools: Reform in Chicago. Chicago, IL: Consortium on Chicago School Research, University of Chicago (February 1993).

Bryk, A.S., Lee, V.E., and Holland, P.B. (1993). Catholic Schools and the Common Good. Cambridge, MA: Harvard University Press.

Bryk, A.S. and S.W Raudenbush (1992). Hierarchical Linear Models: Applications and Data Analysis Methods. Newbury Park, CA: Sage.

Burns, T. and Stalker, G.M. (1961). The Management of Innovation. London: Tavistock Publications.

Carnegie Foundation for the Advancement of Teaching (1992). School Choice: A Special Report. Princeton, NJ: Author.

Center on Organization and Restructuring of Schools [CORS] (1992 Fall).

Estimating the extent of school restructuring. Brief to Policymakers, Brief No. 4. Madison, WI: Author.

Chambers, J.G. (1981). An analysis of school size under a voucher system. Educational Evaluation and Policy Analysis, 3(2):29-40.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., and York, R. (1966). Equality of Educational Opportunity Report. Washington D.C.: United States Government Printing Office.

Coleman, J.S., T. Hoffer, and S.B. Kilgore (1982). High School Achievement: Public and Private Schools Compared. New York: Basic Books.

Conant, J.B. (1959). The American High School Today. New York: McGraw Hill.

Conley, D.T. (1993). Roadmap to Restructuring: Policies, Practices, and the Emerging Vision of Schooling. Eugene, OR: ERIC Clearinghouse on Educational Measurement, University of Oregon.

Cuban, L. (1984). How Teachers Taught: Constancy and Change in American Classrooms, 1890-1980. New York: Longman.

Cuban, L. (1990). Reforming again, again, and again. Educational Researcher, 19(1):3-13.

Edmonds, R. (1984). School effects and teacher effects. Social Policy, 15:37-9.

Elmore, R.F. (Ed.) (1990). Restructuring Schools: The Next Generation of Educational Reform. San Francisco, CA: Jossey-Bass Publishers.

Forsyth, P.B. and Hoy, W.K. (1978). Isolation and alienation in educational organizations. Educational Administration Quarterly, 14(1):80-96.

Fox, W.F. (1981). Reviewing economices of size in education. Journal of Education Finance, 6:273-296.

Friedkin, N.E. and Necochea, J. (1988). School size and performance: A contingency perspective. Educational Evaluation and Policy Analysis, 10(3):237-249.

Fuller, B., Wood, K., Rapoport, T., and Dornbusch, S.M. (1982). The organizational context of individual efficacy. Review of Educational Research, 52(1):7-30.

Garet, M.S. and DeLany, B. (1988). Students, courses, and startification. Sociology of Education, 61(2):61-77.

Godding, R.Z. and Wagner, J.A. III (1985). A meta-analytic review of the relationship between size and performance: The productivity and efficiency of organizations and their subunits. Administrative Science Quarterly, 30:462-481.

Gottfredson, G.D. and Daiger. D. (1979). Disruption in Six Hundred Schools (Rep. 289). Baltimore: John Hopkins University, Center for Social Organization of Schools.

Grant, G. (1988). The World We Created at Hamilton High. Cambridge, MA: Harvard Univ. Press.

Guthrie, J. (1979). Organizational scale and school success. Educational Evaluation and Policy Analysis, 1(1):17-27.

Haller, E.J., Monk, D.H., Spotted Bear, A., Griffith, J., and Moss, P. (1990). School size and program comprehensiveness: Evidence from High School and Beyond. Educational Evaluation and Policy Analysis, 12(2): 109-120.

Hambleton, R.K. (1989). Principles and selected applications of Item Response Theory. In R.L. Linn (Ed.) Educational Measurement, Third Edition: 147-200. New York: Macmillan.

Heyns, B.L. (1974). Social selection and stratification within schools. American Journal of Sociology, 79(6):1934-1951.

Ingels, S.J., S.Y. Abraham, S.Y., Rasinski, K., Karr, R., Spencer, B.D. and Frankel. M.R. (1989). National Education Longitudinal Study of 1988. Base Year: Student Component Data File User's Manual. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement. NCES 90-464.

Ingels, S.J., Scott, L.A., Lindmark, J.T., Frankel, M.R., and Meyers, S.L. (1992). National Education Longitudinal Study of 1988. First Follow-up: Student Component Data File User's Manual. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement. NCES 92-030.

Kyle, R.M.J. (1993). Transforming Our Schools: Lessons from the Jefferson Country Public Schools/Gheen Professional Development Academy 1983-1991. Louisville, KY: Jefferson Country Public Schools.

Lee, V.E. and Bryk, A.S. (1988). Curriculum tracking as mediating the social distribution of high school achievement. Sociology of Education, 61 (2):78-94.

_____ (1989a). A multilevel model of the social distribution of high school achievement. Sociology of Education, 62 (3):172-192.

_____ (1989b). Effects of single-sex schools: Response to Marsh. Journal of Educational Psychology, 81(4):647-650.

Lee, V.E., Bryk, A.S., and Smith, J.B. (1993). The Effects of High School Organization on Teachers and Students. In L. Darling-Hammond (Ed). Review of Research in Education, 19:171-268. Washington, D.C.: AERA Publ.

Lee, V.E. and Smith, J.B. (1993). Effects of school restructuring on the

achievement and engagement of middle-grade students. Sociology of Education, 66(3):164-187.

Lightfoot, S.L. (1984). The Good High School: Portraits of Character and Culture. New York: Basic Books.

Lord, F.M. (1980). Applications of Item Response Theory to Practical Testing Problems. Hillsdale, NJ: Laurence Erlbaum Associates.

March, J.G. and Olsen, J.P. (1976). Ambiguity and Choice in Organizations. Bergen: Universitets Forlaget.

Monk, D.H. (1987). Secondary school size and curriculum comprehensiveness. Economics of Education, 6(2):137-150.

Murphy, J. and P. Hallinger (Eds.) (1993). Restructuring Schooling: Learning From Ongoing Efforts. Newbury Park, CA: Corwin Press.

National Commission on Excellence in Education (1983, April). A Nation at Risk: The Imperative for Educatinal Reform. Washington, DC: U.S
National Commission on Excellence in Education (1983, April). A Nation at Risk: The Imperative for Educatinal Reform. Washington, DC: U.S. Department of Education.

Newmann, F.M. (1981). Reducing student alienation in high schools: Implications of theory. Harvard Educational Review, 51(4):546-564.

_____ (1991). Beyond Common Sense in Educational Restructuring: The Issues of Content and Linkage. Madison, WI: Center on Organization of Restructuring of Schools, University of Wisconsin.

Newmann, F.M. and Oliver, D.W. (1967). Education and Community. Harvard Educational Review, 37:61-106.

Oakes, J. (1985). Keeping Track: How Schools Structure Inequality. New Haven: Yale Univ. Press.

_____ (1990). Opportunities, achievement, and choice: Women and minority students in science and mathematics. In C.B. Cazden (Ed.), Review of Research in Education, 16:153-222. Washington, D.C.: American Educational Research Association.

Powell, A.G., Farrer, E. and Cohen, D.K. (1985). The Shopping Mall High School: Winners and Losers in the Educational Marketplace. Boston: Houghton Mifflin.

Purkey, S.C. and Smith, M.S. (1983). Effective schools: A review. The Elementary School Journal, 83(4), 427-452.

Rosenbaum, J.E. (1976). Inequality: The Hidden Curriculum of High School Tracking. New York: John Wiley and Sons.

Rowan, B. (1990) Applying conceptions of teaching to organizational reform. In R.E. Elmore (Ed). Restructuring Schools: The Next Generation of Educational Reform (pp.31-58). San Francisco:

Jossey-Bass.

Rutter, M., Maughan, B., Mortimore, P., Ouston, J., and Smith, A. (1979). Fifteen Thousand Hours: Secondary Schools and their Effects on Children. Cambridge: Harvard Univ. Press.

Sarason, S.B. (1990) The Predictable Failure of School Reform. San Francisco: Jossey-Bass.

Seeman, M. (1975). Alienation studies. Annual Review of Sociology (Vol. 1). Palo Alto, CA: Annual Reviews.

Shedd, J.B. and Bacharach, S.B. (1991). Tangled Hierarchies: Teachers as Professionals and the Management of Schools. San Francisco: Jossey-Bass.

Sher, J.P. and Tompkins, R.B. (1977). Economy, efficiency, and equality: The myths of rural school and district consolidation. In J.P. Sher (Ed.) Education in Rural America. Boulder, CO: Westview Press.

Simon, H.A. (1976). Administrative Behavior: A Study of Decision-making Processes in Administrative Organizations (3rd Ed.). New York, NY: Free Press.

Sizer, T.R. (1984). Horace's Compromise: The Dilemma of the American High School. Boston: Houghton Mifflin.

Tyack, D. (1974). The One Best System. Cambridge, MA: Harvard Univ. Press.

Waller, W. (1932). The Sociology of Teaching New York: Russell & Russell.

Wehlage, G., Stone, C., and Kleibard, H.M. (1980). Dropouts and Schools: Case Studies of the Dilemmas Educators Face. Madison: Univ. of Wisconsin Press.

Weick, K.E. (1976). Educational organizations as loosely coupled systems. Administrative Science Quarterly, 21:1-19.

Willett, J.B. (1994). "Measuring Change More Effectively by Modeling Individual Growth Over Time." In T. Husen and T.N. Postlethwaite (Eds.) The International Encyclopedia of Education, Second Edition. Oxford, UK: Pergamon Press.

Table 1: Frequency of Structural Practices in Secondary Schools, Classified as Traditional, Moderate, and Restructured (n=820 schools)

Structural Practice	NELS Variable	Probability[a]
Traditional Practices		
Departmentalization with chairs	F1C73D3	.85
Common classes for same curricular track	F1C73A3	.76
Staff development focusing on adolescents	F1C73P3	.66
PTA or PTO	F1C73N3	.64
Parent-teacher conferences each semester	F1C73O3	.64
Focus on critical thinking in curriculum	F1C73R3	.64
Common classes for different curr. tracks	F1C73I3	.62
Increased graduation requirements	F1C73S3	.62
Recognition program for good teaching	F1C52	.56
Parents sent info. on how to help kids study	F1C73M3	.56
Moderate Practices		
Parent workshops on adolescent problems	F1C73L3	.46
Student satisfaction with courses important	F1C47C	.42
Strong emphasis on parental involvement	F1C91E	.38
Strong emphasis on increasing academic req.	F1C91C	.35
Stu. evaluation of course content important	F1C47B	.35
Outstanding teachers are recognized	F1C53B	.34
Emphasis on staff stability	F1C91B	.34
Emphasis on staff development activities	F1C91D	.32
Restructured Practices		
Students keep same homeroom throughout HS	F1C73G3	.30
Emphasis on staff solving school problems	F1C91A	.29
Parents volunteer in the school	F1C73K3	.28
Interdisciplinary teaching teams	F1C73E3	.24
Independent study, English/social studies	F1C73B3	.23
Mixed-ability classes in math/science	HTRGRP[b]	.21
Cooperative learning focus	F1C73H3	.21
Student evaluation of teachers important	F1C47A	.20
Independent study in math/science	F1C73C3	.18
School-within-a-school	F1C73Q3	.15
Teacher teams have common planning time	F1C73F3	.11
Flexible time for classes	F1C73J3	.09

a. Each figure in this column represents the probability that an average high school (one which reports that it has adopted 11-13 of the 30 reform practices listed here) engages in each practice.

b. School aggregate created from NELS variable F1T2_4. Schools coded "1" if at least 70% of the surveyed science and math teachers' responded that their classes were composed of students of "widely differing achievement levels;" coded "0" otherwise.

Table 2: **Early High School Restructuring Study: Means of Variables Describing High-School Sophomores Attending Schoools With No Reform Practices, With Traditional or Moderate Practices, and With Restructured Practices (N=11,794 Students)**

Variable	Schools Without Reform Practices	Schools With Traditional Practices	Schools With Restructured Practices
Sample Size (a)	n=1,280	n=5,353	n=5,161
Dependent Variables			
Engagement (10th)[c]	.055[d]	-.051	-.003
Math gain (8->10)	4.74[d]	5.28	5.49
Reading gain (8->10)	2.18[d]	2.53	2.39
History gain (8->10)	1.92[d]	2.21	2.33
Science gain (1->10)	2.11	2.26	2.57[e]
Independent Variables			
Engagement (8th)	-.07[d]	-.11	.24[e]
Social class[c]	-.232[d]	-.101	.049[e]
Minority status (%)	34.9[d]	15.9	20.8[e]
Female (%)	52.5	50.6	50.2
Ability Controls (8th Grade):			
Test composite[c]	-.20[d]	-.05	.14[e]
Reading control[c]	-.21[d]	-.04	.14[e]
Math control[c]	-.21[d]	-.04	.12[e]
History control[c]	-.20[d]	-.05	.13e
Science control[c]	-.21[d]	-.05	.14[e]

a. School sample sizes are computed unweighted.

b. Variable means and contrasts are computed with NELS student-level design weight.

c. Variables are standardized, M = 0, SD = 1.

d. Contrast of No Reform vs. Traditionally Reformed Schools statistically significant at probabibility .05 or below.

e. Contrast of Restructured vs. Traditionally Reformed Schools statistically significant at probabibility .05 or below.

Table 3: Early High School Restructuring Study: Means of Variables Describing Characteristics of Schools With No Reform Practices, With Traditional or Moderate Practices, and With Restructured Practices (N=820 Schools)

Variable	Schools Without Reform Practices	Schools With Traditional Practices	Schools With Restructured Practices
Sample Size (a)	n=131	n=465	n=524
Average SES (c)	-.22[d]	.01	.17[e]
% Minority Enrollment	39.3[d]	24.4	27.7
School Sector			
% Public	96.2[d]	86.9	80.3[e]
% Catholic	2.3[d]	9.7	9.9
% NAIS	1.5	3.4	9.7[e]
Average No. of Math & Science Courses	2.36[d]	2.59	2.79[e]
Variability in Math, Science Course Taking (SD)	1.47[d]	1.26	1.23
School Size	1463[d]	1125	1264[e]

a. School sample sizes are computed unweighted.

b. Variable means and contrasts are computed with NELS constructed school weight.

c. Variables are standardized, M = 0, SD = 1.

d. Contrast of No Reform vs. Traditionally Reformed Schools statistically significant at probabibility .05 or below.

e. Contrast of Restructured vs. Traditionally Reformed Schools statistically significant at probabibility .05 or below.

Table 4: **Early High School Restructuring Study: HLM Estimates of Psychometric Properties of Dependent Variables: Academic Engagement and Achievement Gains (N=11,794 Students)**

	Dependent Variables				
	Academic Engagement	Gain in Mathematics	Gain in Reading	Gain in History	Gain in Science
Within-school variance (sigma-squared)	.95	40.81	25.55	11.75	13.62
Between-school variance (tau)	.04	2.89	1.22	0.76	1.69
Reliability	.27	.36	.28	.34	.50
Intra-class correlation(b)	.13	.16	.15	.16	.20

a. All parameters in this table are computed with a fully unconditional HLM model (i.e., a model which includes neither within-school nor between-school variables).

b. These figures are computed with the within-school variance (sigma-squared) adjusted for the HLM estimate of its reliability.

Table 5: Early High School Restructuring Study: HLM Within-School Model for Academic Engagement and Achievement Gains (N=11,794 Students)

	Dependent Variables				
	Academic Engagement	Gain in Mathematics	Gain in Reading	Gain in History	Gain in Science
Independent Variables					
Intercept (School Average)	.00(a)	3.22***	2.22***	1.76***	2.86***
8th Grade Engagement	1.16***	.16***	.08*	.05*	.07*
8th Grade Ability (b)	.48***	-.04*	.07~	-.07**	.21***
Social Class(c)	.77***	.14**	.04~	.11~	.12**
Minority Status	.82***	-.04	-.23*	-.12~	-.68***
Gender(Female)	1.22***	-.10	-.04	-.07	-.37***

~ p < .10; * p < .05; ** p < .01; *** p < .001

a. All effects in this table are presented are in a standardized metric. These are computed by dividing the HLM gamma coefficient for each outcome by the adjusted school-level standard deviation (SD) of that outcome computed from by HLM. These SDs are displayed at the bottom of Table 5.

b. The 8th-grade ability control is different for each outcome. For engagement, it is a composite of reading and math achievement. For the achievement gains, the ability control is constructed as a composite of the 8th grade tests in the three curricular areas *not* measured by the gain score.

c. In the HLM model, SES is allowed to vary randomly between schools, while the other controls are employed as fixed parameters. SES is centered around the sample mean, while the other controls are centered around their respective school means.

Table 6: **Early High School Restructuring Study: HLM Between-School Model for Academic Engagement and Achievement Gains (N=820 Schools)**

	Dependent Variables				
	Academic Engagement	Gain in Mathematics	Gain in Reading	Gain in History	Gain in Science
Effects on Mean Between-School Outcome (a)					
Av. Intercept	-.89***(b)	2.92***	1.94***	1.71***	2.95***
School SES	-.08	-.02	.08	.10	.27***
Minority Conc'tn	.03~	.00	.10	.03	-.23~
Catholic HS	-.07	.49*	.32~	-.01	.33
NAIS HS	1.23***	.18	.44~	.03	.45~
Academic Emphasis	.26***	.21**	.18*	.13*	.56***
Course-taking Variability	-.17*	-.07	-.06	-.09~	.26**
School Size	-.19*	-.39***	-.32***	-.36***	-.37***
No Reform Practices	-.14*	-.21~	-.20~	-.15*	-.10
Restructured Reforms	.37**	.49***	.37***	.35***	.59***

Effects on SES-Differentiation (a)

Av.SES Slope	1.59***	.43***	.51*	.36*	.61***
School SES	.63**	.39~	.19	.20	.06
Minority Conc'tn	.25	.26	.44	-.27	-.05
Catholic HS	-.82	-.26	-.68*	-.70	-.06
NAIS HS	.03	.75	.25	-.28	-.70
Academic Emphasis	-.94**	-.27*	-.34*	-.13	-.46*
Course-taking Variability	.34~	.02	.09	.20	.12
School Size	.10	.03*	.34**	.22*	.25*
No Reform Practices	.16	.18*	.54*	.33*	.47~
Restructured Reforms	-.54***	-.33**	-.38*	-.32*	-.30**

~ p < .10; * p < .05; ** p < .01; *** p < .001

a. All HLM analyses shown here also include adjustment for all within-school variables shown on Table 5: 8th grade engagement, 8th grade ability, SES, minority status, and gender.

a. All effects shown on this table are presented in a standardized metric. These are computed by dividing the HLM gamma coefficient for each outcome it by the adjusted school-level standard deviation computed from by HLM, which are as follows:.

	Academic Engagement	Gain in Mathematics	Gain in Reading	Gain in History	Gain in Science
SD-Intercept	.164	1.762	1.248	1.417	1.005
SD-SES Slope	.083	.667	.456	.393	.351

Appendix 1: Brief Description of the Hierarchical Linear Models Used in This Study

Within-School Models

A simple form of Hierarchical Linear Models (HLM) used here consists of two equations, a within- and a between-school model. Some of the parameters estimated in the within-school model become outcomes to be explained in between-school equations. One within-school model investigates the gain in mathematics achievement of student i in school j, Y_{ij}, as a function of student background characteristics, X_{ij}'s (the X-variables considered here are ability, engagement, SES, minority status, and gender), and and random error, R_{ij}:

$$R_{ij} = ß_{j0} + ß_{j1}X_{ij1} + ß_{j2}X_{ij2} + \ldots + ß_{jk}X_{ijk} + R_{ij}$$

The $ß_{jk}$ regression coefficients are structural relations occurring within school j that indicate how achievement in each school is distributed across the measured student characteristics. In the HLM models investigated here, we are particularly interested in two ß parameters:

$ß_{0j}$ - the average gain in mathematics for students in school j; and
$ß_{1j}$ - the relationship between SES and math gain in school j. We refer to this as the SES-math learning slope.

While the other ß parameters (i.e., distributional effects) were also estimated in our HLM analyses, we were not interested in modeling these parameters as functions of structural parameters. As such, the other within-school controls (ability, engagement, minority status, and gender) are *fixed* in our HLM models. This means that we have "fixed" the between-school variability in these other ß parameters to 0 (i.e., they do not vary randomly between schools).

Between-School Models

In the second set of equations, we model the random-effect ß parameters, adjusted for student characteristics, as functions of school-level characteristics (W-variables). We estimate a single between-school model for for each outcome, estimating the effects of the three restructuring components on the outcomes ($ß_0$ and $ß_1$ for each outcome). For each model, we also adjust for the potentially confounding effects of school structure and demographics. A typical between-school model is as follows:

$$ß_{jk} = \gamma_{0k} + \gamma_{1k}W_{1j} + \gamma_{2k}W_{2j} + \ldots + \gamma_{pk}W_{pj} + U_{jk}.$$

The parameters of interest here are the effects associated with the school restructuring variables, W_{1j} -- the γ_{pk} coefficients. Since the error terms in this equation are complex, conventional linear model techniques may not be used. However, recent developments in statistical theory and computation, available through the HLM software, make this estimation possible. Briefly, the total variance in each outcome is partitioned into two components: parameter and error variance. It is only effects on the *parameter variance* which are estimated in HLM. This is an important development, since it is only variability in the structural parameters, $Var(ß_{jk})$, which can be explained by school factors. In general, previous efforts to estimate school effects with ordinary least squares regression have systematically underestimated school effects for this reason.

Appendix 2: Description of Variable Construction for all Measures Used in This Study

Dependent Measures

o *Achievement Gains*
+ F1TXMG -- Mathematics IRT-estimated gain between 8th and 10th grade.
+ F1TXRG -- Reading IRT-estimated gain between 8th and 10th grade.
+ F1TXHG -- History IRT-estimated gain between 8th and 10th grade.
+ F1TXSG -- Science IRT-estimated gain between 8th and 10th grade.

o *10th-Grade Academic Engagement*
Standardized factor-weighted composite (M = 0, SD = 1) of 8 items measuring student behaviors (related to their current courses). Composite created using principle components factor analysis. Item coding reflects students' assessment of the frequency with which they engage in each behavior, coded from 1 (never) to 5 (every day). Internal consistency (Cronbach's alpha) = .84. Factor eigenvalue = 3.76, percent of total variance in all items explained by the factor = 47.1. NELS student item components are:

+ F1S27A OFTEN WORK HARD IN MATH CLASS
+ F1S27B OFTEN WORK HARD IN ENGLISH CLASS
+ F1S27C OFTEN WORK HARD IN HISTORY CLASS
+ F1S27D OFTEN WORK HARD IN SCIENCE CLASS
+ F1S28A OFTEN FEEL CHALLENGED IN MATH CLASS
+ F1S28B OFTEN FEEL CHALLENGED IN ENGLISH CLASS
+ F1S28C OFTEN FEEL CHALLENGED IN HISTORY CLASS
+ F1S28D OFTEN FEEL CHALLENGED IN SCIENCE CLASS

Measures of School Restructuring

o *Restructuring Measures*
Two dummy-coded items were created, using the variables in Table 1 and the technique described in the paper. Measures constructed from items from NELS first followup school file, where principals reported whether or not schools engaged in a set of 30 school practices. Using those reports, two measures were:

+ NO REFORM PRACTICES: Schools which engaged in no reform practices coded 1, schools classified as those with traditional reforms coded 0.

+ RESTRUCTURING PRACTICES: Schools which engaged in at least three practices listed at restructured in Table 1 were coded 1, schools classified as those with traditional reforms coded 0.

o *School Size*
+ F1C2 TOTAL ENROLLMENT AS OF OCTOBER 1989
Principal's report of high school size (on NELS restricted school file) was transformed to its natural logarithm and standardized (M = 0, SD = 1).

Control Variables

Student Background

o *Socioeconomic Status*
  + F1SES -- socio-economic status composite

o *Minority Status*
  + F1RACE -- student race (recoded to: 0=white or Asian; 1=black, Hispanic, or Native American)

o *Gender*
  + F1SEX Student gender (recoded to: 0=male; 1=female)

o *Academic Controls*
  Analyses included different controls for each curriculum area. Controls were constructed as follows:
  + For math gain: Z-score of sum of BYTXRIRS, BYTXHIRS, BYTXSIRS.
  + For reading gain: Z-score of sum of BYTXMIRS, BYTXHIRS, BYTXSIRS.
  + For history gain: Z-score of sum of BYTXRIRS, BYTXMIRS, BYTXSIRS.
  + For science gain: Z-score of sum of BYTXRIRS, BYTXHIRS, BYTXMIRS.

o *Engagement Control*
  A factor-weighted standardized (M = 0, SD = 1) composite of the following base-year student measures of student engagement. Composite's internal consistency (Cronbach's alpha) =.74.

  + BYS69A LOOK FORWARD TO MATH
  + BYS70A LOOK FORWARD TO ENGLISH
  + BYS71A LOOK FORWARD TO SOCIAL STUDIES
  + BYS72A LOOK FORWARD TO SCIENCE
  + BYS69C MATH USEFUL IN MY FUTURE
  + BYS70C ENGLISH USEFUL IN MY FUTURE
  + BYS71C SOCIAL STUDIES USEFUL IN MY FUTURE
  + BYS72C SCIENCE USEFUL IN MY FUTURE
  + BYS55A SENT TO OFFICE FOR MISBEHAVING (REVERSED)
  + BYS55E PARENTS RECEIVED WARNING ABOUT MY BEHAVIOR (REVERSED)
  + BYS55F GOT INTO A FIGHT WITH ANOTHER STUDENT (REVERSED)
  + BYSF56E STUDENTS IN CLASS SEE ME AS A TROUBLE-MAKER (REVERSED)
  + BYS78A HOW OFTEN COME TO CLASS WITHOUT PENCIL OR PAPER (REVERSED)
  + BYS78B HOW OFTEN COME TO CLASS WITHOUT BOOKS (REVERSED)
  + BYS78C HOW OFTEN COME TO CLASS WITHOUT HOMEWORK (REVERSED)
  + BYS75 HOW OFTEN MISS SCHOOL (REVERSED)
  + BYS76 HOW OFTEN CUT OR SKIP CLASS (REVERSED)
  + BYS77 HOW OFTEN COME TO CLASS LATE (REVERSED)

School Demographics and Structure

o *Average Socioeconomic Status*
  + AVSES -- SES composite, aggregated to the school level.

o *Minority Concentration*
  + F1RACE -- student race (recoded to: 0=white or Asian; 1=Black, Hispanic, or Native American), aggregated to the school level, and recoded to a dichotomous variable (recoded to: 1=40% or more, 0=less

than 40% minority).

o *Academic Emphasis*

Summed 10th graders' reports of coursetaking in academic courses in mathematics and science:

+ F1S22C -- HOW MUCH COURSEWORK IN ALGEBRA I
+ F1S22D -- HOW MUCH COURSEWORK IN GEOMETRY
+ F1S22E -- HOW MUCH COURSEWORK IN ALGEBRA II
+ F1S22F -- HOW MUCH COURSEWORK IN TRIGONOMETRY
+ F1S22G -- HOW MUCH COURSEWORK IN PRE-CALCULUS
+ F1S22H -- HOW MUCH COURSEWORK IN CALCULUS
+ F1S23C -- HOW MUCH COURSEWORK IN BIOLOGY
+ F1S23E -- HOW MUCH COURSEWORK IN CHEMISTRY
+ F1S22F -- HOW MUCH COURSEWORK IN PHYSICS

Variable was then aggregated to the school level as a school mean, and standardized (M = 0 , SD = 1).

o *Variability in Coursetaking*

The sum of students' courstaking (academic emphasis, described above) was aggregated to the school level, using the standard deviation operator in SPSSX. Variable was standardized (M = 0, SD = 1).

o *Sector*

Created from G10CTRL2, the school control measure on the NELS first followup restricted school file. Public, Catholic, and NAIS schools were retained, other private schools were dropped. Created 2 dummy-coded variables:

+ CATHOLIC -- coded 1 for Catholic, 0 for public, NAIS schools.
+ NAIS -- coded 1 for NAIS, 0 for public, Catholic schools.

Figure 1. Sensitivity Analysis on Effectiveness:

Effects of Number of Restructuring Reforms On Gains in Achievement and Engaement

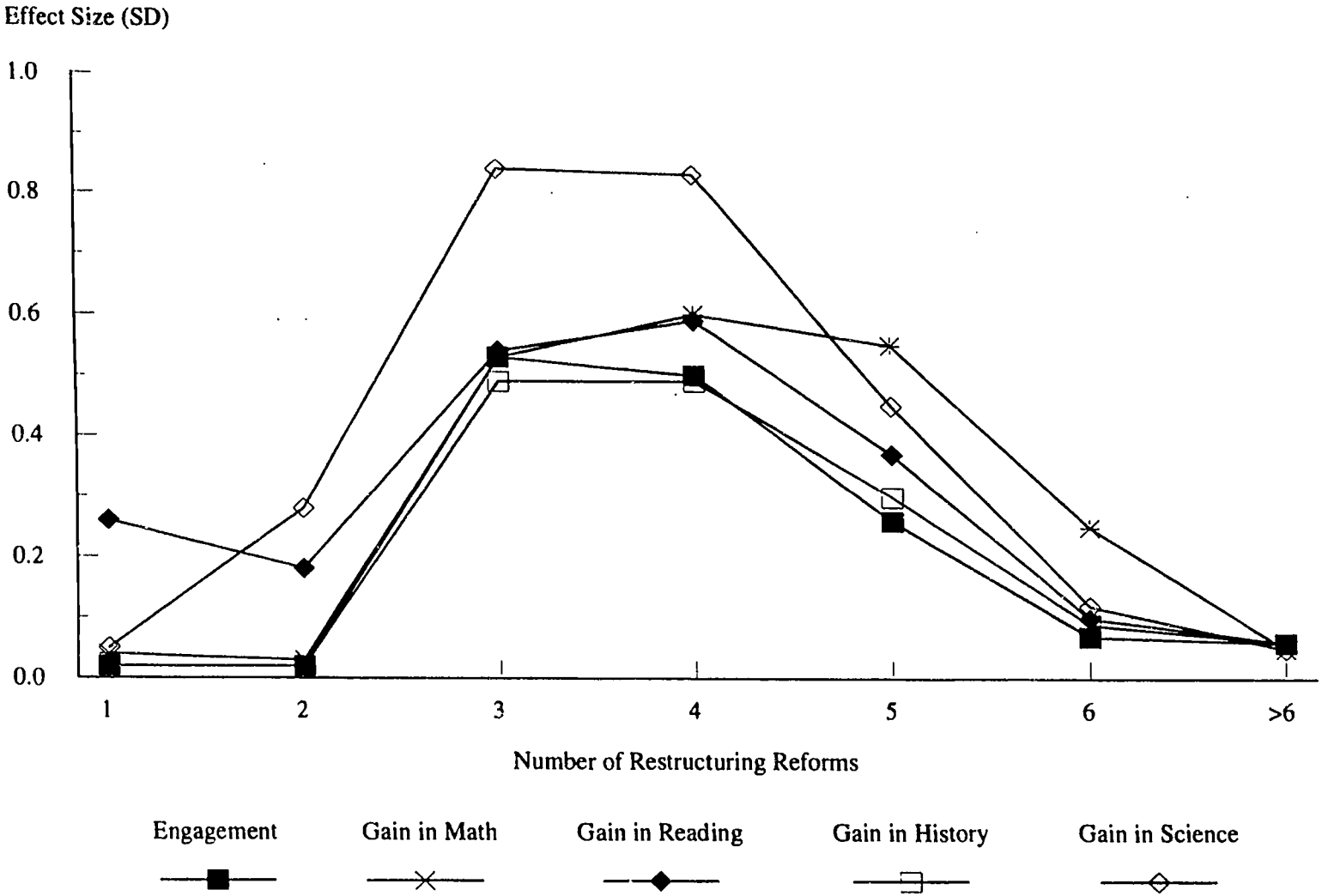

Figure 2. Sensitivity Analysis on Equity:

Effects of Number of Restructuring Reforms on SES Slopes on Gains in Achievement and Engaement