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

Valerie E. Lee  
University of Michigan  
Julia B. Smith  
University of Rochester

This study assessed the impact on 10th-grade students of attending high schools whose practices are consistent with the school-restructuring movement. Using data on a sample of 11,794 sophomores in 830 high schools from the first two waves of the National Educational Longitudinal Study of 1988, the authors evaluated restructuring effects on students' gains in engagement and achievement in four subjects and the social distribution of those gains. High schools with several practices consistent with restructuring and those with none of the 30 practices that were considered were contrasted with schools that engaged in only traditional reforms. School size was evaluated as an independent structural feature. The results revealed that students' gains in achievement and engagement were significantly higher in schools with restructuring practices and lower in schools without reforms. Higher and more socially equitable engagement and achievement were consistently associated with smaller high schools.

The movement to restructure American education continues to be popular in the lexicon of contemporary school reforms and reflects the nation's ongoing dissatisfaction with its schools. However, the reform movement embodied by the term restructuring rests on thin and inconsistent theory. Recommendations for reforms that have been grouped under this heading have been drawn from diffuse sources, including the narrow research base on effective teaching practices, production models formulated in industrial settings, and popular calls for empowering disenfranchised interest groups.

Rather than forge a coherent theoretical base from such disparate sources, we have attempted to locate this reform within a theoretical contrast that has been reinvigorated in the past few decades to describe differences in secondary schools: bureaucratic versus communal organizational forms. The Center on Organization and Restructuring of Schools (CORS) at the University of Wisconsin-Madison, sponsored by the U.S. Department of Education, has been the major impetus for developing the theory that undergirds restructuring the organization of schools (Newmann 1991). Its mission is also to study such reforms. Our efforts to provide an appropriate theoretical grounding are meant to be consistent with its work.

Two factors make it difficult to evaluate the effects of school restructuring on students. First, few high schools can be defined as either restructured or even actively restructuring. Recent efforts (see, for example, Berends and King 1994; Brown 1993; CORS 1992) to evaluate the degree of implementation of restructuring reforms have concluded that the fundamentally restructured school is rare. Second, for the more numerous schools that are in the process of implementing reforms in the spirit of restruc-
turing, it seems premature to assess the effects of these early organizational changes.

In the study presented here, we took an eclectic approach to this problem by studying the effects of organizational forms that are consistent with the intent of the restructuring movement, although they are not necessarily newly implemented (not always restructured from a different form). This approach represents an attempt to demonstrate that practices that are consistent with the restructuring movement have an impact on students' learning. Thus, although we evaluated how such practices affect changes in students, we did not study the effects of organizational change on students. Our study identified high schools by the organizational practices they follow. Practices that fit our definition of restructuring capture two related ideas: (1) they represent a movement from the bureaucratic form of American schooling toward the communal form and (2) they are consistent with a definition of restructuring laid out by CORS.

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) pointed out, the separation of secondary schooling from the traditional village school into a more formal and systematized, and ultimately bureaucratic, organization was itself a significant educational reform in the late 19th and early 20th centuries. In the 1980s, social critics (such as Boyer 1983; National Commission on Excellence in Education 1983; Sizer 1984) described the secondary school as so fundamentally linked to the success of this country's economic and political future that its failure jeopardizes the future of the nation itself. The multiple ailments of high schools have prompted constant and strident calls for reform, revision, and restructuring. Almost no one argues that U.S. high schools are "working."

Although calls for reforms to meet existing goals have been continuous, the goals of American education have themselves undergone substantial revision over the 120 years of secondary schooling. A major thrust of the Progressive Movement was the establishment of a national network of large high schools, designed to conform to such typically American ideals as efficiency, differentiation, specialization, depersonalization, and standardization; in effect, this network was a highly managed, smoothly functioning, and well-oiled machine whose goal was the production of human capital. Few educational reform efforts have "succeeded" as well as the comprehensive high school. Since the late 1960s, however, the basic foundations of the American high school have been questioned, and high schools have increasingly been described as offering inappropriate responses to the human concerns of organizational members (see, for example, Newmann and Oliver 1967; Oliver 1976; Tyack 1974).

Big schools may not be better, in terms of either economic efficiency (Fox 1981; Guthrie 1979) or academic development (Goddard and Wagner 1985; Haller, Monk, Spotted Bear, Griffith, and Moss 1990; Monk 1987). Moreover, a standard of equity in the distribution of school outcomes has been joined with the classical standard of excellence in assessments of the effectiveness of schools (Coleman et al. 1966; Edmonds 1984; Purkey and Smith 1983). As the nation rethinks which goals are appropriate for a democratic educational system, it must revise not only the goals, but the means for achieving them. High schools, as vital agents of education in a democracy, should come under special scrutiny.

BACKGROUND

Two Theories of School Organization

The structure of an organization refers to the relationships among members around its technical core of work (Perrrow 1967; Simon 1976). Since our study linked organizational form to students' achievement, we invoked a theory of how the "core technology" of schools is organized: as a bureaucracy or as a community. Although the grounding for this theoretical contrast was developed by classical sociological writing (such as
Effects of High School Restructuring

by Weber, Durkheim, Parsons, Sorokin, and Tonnies), Bryk and Driscoll (1988) articulated the theory in a form that applies to schools; this article draws heavily on their work. Since the two organizational forms incorporate different conceptions of the core technology, they are based on opposite assumptions about knowledge, learning, and teaching: the bureaucratic form, on a routine, clear, and stable technology, and the communal form, on a nonroutine core technology (Burns and Stalker 1961; Rowan 1990). How do these assumptions play out in the organizational structure of schools?

The bureaucratic form. Bureaucratic organizations characteristically include specialized and differentiated work roles, a top-down hierarchy of decision making, and a formalization of goals and expectations into affectively neutral rules and codes of behavior (Bryk and Driscoll 1988; Newmann and Oliver 1967; Rowan 1990). In schools with this form, the routine technical activity, instruction, breaks down knowledge into a curriculum composed of discrete and fixed subjects. The aim of teaching is to impart specialized knowledge, and instruction is organized into a standardized and sequenced pattern within subjects. Learning is assessed by measuring the mastery of subject matter, and students are sorted into specialized instructional treatments aimed at appropriately matching their ability and interests to the subject matter. The organization of instruction into departments and tracks is consistent with a specialization model. Such reforms as increasing academic standards, tightening graduation requirements, and teacher-proof curricular materials fit the bureaucratic model.

The communal form. In communal organizations, tasks are less certain and conditions are more changeable and unpredictable (Bryk and Driscoll 1988; Newmann and Oliver 1967; Rowan 1990). These organizations typically emphasize shared responsibility for work, shared commitment to a common set of goals, lateral communication and power in decision making, and expectations and behavior framed by greater personalization and individual discretion. In schools with this form, typical of small high schools in the early 20th century (especially in rural areas), knowledge is seen as multidimensional and interdisciplinary, and teaching is responsive to students' opinions, talents, and tastes. Learning is built more around concrete "problems" than abstract "subjects," and assessment is more flexible and less standardized. Organizational responses to the communal school's core technology may include independent study, interdisciplinary teaching, flexible scheduling, cooperative learning, and mixed-ability classes.

Moving away from the dominant form. The theories defining these alternative forms of teaching and learning are well established in American education and have undergirded historical and theoretical debates about the proper direction of school reform for at least a century (Cuban 1984, 1990). However, one form has been dominant in secondary schools since reforms of the Progressive Era moved schools in a bureaucratic direction toward a "perfect" product, the comprehensive high school. Some reforms have aimed to make the traditional system "work better," but the bureaucratic structure itself has been assumed to be solid until recently. Despite a history of discussion of both theories, the bureaucratic form still constitutes the "tradition" against which current structural reform efforts are targeted, including the call to "restructure" the American high school. We view the restructuring movement through the lens of these two theories of teaching and learning. Calls to restructure schools suggest a fundamental shift from the bureaucratic model toward the communal organizational model.

The difficulty of real reform. The modern comprehensive high school is a conservative organization, in which fundamental change is difficult. Reforms

---

1 The contrast laid out by Bryk and Driscoll was a major theoretical foundation in several of our recent writings about school organization and its effects on students and teachers: Bryk, Lee, and Holland (1993); Bryk, Lee, and Smith (1990); Lee, Bryk, and Smith (1993); and Lee and Smith (1993).
aimed at nudging the organization from a bureaucratic toward a communal form would require real effort to initiate, would not take hold easily, would be resisted by many, and would thus be firmly in place in few high schools. One definition of school restructuring—a substantial departure from conventional practice—captures this idea. Thus, even with secondary school reform on the front burner of educational policy, the probability that a particular reform practice would be adopted in any high school should be associated with its form. Although the reform “bandwagon” is moving fast, reforms with a communal form are probably less common than those that are not meant to disturb the heart of an essentially bureaucratic organization. This contention is supported by recent research that attempted to measure the extent of fundamental school restructuring. Although schoolwide reform is rare (Berends and King 1994; Brown 1993; CORS 1992), the implementation of practices reflecting the spirit of reform may be more common.

Student’s achievement under each form. We assume that the primary goal of all schools is to impart knowledge to students. Despite similar goals, however, schools’ organizational forms are likely to influence the means used to accomplish this goal. One way of capturing this difference is to examine how schools with the two forms might respond to diversity among students (Bryk and Driscoll 1988; Bryk, Lee, and Holland 1993). Under the bureaucratic form, the avowed purpose of matching instruction (and content) to students’ ability is to maximize the amount of demonstrable learning for each individual, in essence, to develop as much human capital as possible, based on the available “raw material.” Diversity would be defined mainly in cognitive terms, with instruction specialized to respond best to differences in students’ ability.

Although schools that are organized as communities also aim to maximize learning, they would more likely define and emphasize a common set of academic needs seen as appropriate for all students, regardless of cognitive capacity. The curriculum would not be dependent on students’ ability, preparation, or plans, although the means for arriving at the goal might be responsive to students’ diversity (the pace for moving through the same material could vary). Students would be expected to experience a set of common academic activities in communally organized schools. These different approaches to maximizing students’ learning might affect the average level of achievement differently and might have different influences on the distribution of learning among students, based on their social background characteristics.

Effects of Restructuring

We admit to only modest success in embedding our empirical research on school restructuring within a theoretical framework drawn from the sociological tradition of research and writings in which schools are seen as organizations. Recent works on the movement to direct reform to the structure or organization of schools (including the school-restructuring movement) have typically not drawn on the theoretical framework laid out here but have focused on practical and political issues. Little theory has been invoked in these writings, and what has been used has been drawn from disparate sources with inconsistent recommendations: research on effective teaching practices, models from industrial settings, and research and advocacy for the empowerment of disenfranchised groups. An exception is the five years of research sponsored by CORS. CORS-sponsored researchers (such as Fred Newmann and Anthony Bryk) have found the bureaucracy-community contrast generally useful, and several have been active in developing this theory.

Although recent writings about the school restructuring movement have been helpful in conceptualizing the issues (see, for example, Conley 1993; Elmore 1990; Murphy and Hallinger 1993), we located only one study (in one school district) that evaluated the effects of restructuring on learning (Kyle 1993). Schools in Jefferson County, Kentucky, have been engaged in a serious restructuring effort for over a decade. Kyle studied 42 of the 157 district schools,
categorizing them according to their commitment to reform. Restructured schools were matched to comparison schools by demographic composition, mobility rates, and level (elementary, middle, and secondary). Schools in Group I had a sustained commitment (three to five years) to restructuring. Group II schools (an exploratory stage) had only short-term experience with a wide range of reforms, and Group III schools were satisfied with current practices and not committed to reform.

The groups were compared on three outcomes: basic improvement (achievement on standardized tests), annual improvement rate (1988–91), and positive involvement (attendance, dropout and suspension rates, and parental involvement). The effects were consistent across outcomes and levels: Compared to the Group II or Group III schools, the Group I schools had higher rates of improvement and involvement. It is interesting to note that the Group II schools, in which reform was neither stable nor consistent but in which a multitude of reforms had been tried, had less favorable outcomes than did the Group III schools. Despite some design weaknesses (small samples, no statistical tests, and possible selection bias), the willingness of Jefferson County to evaluate its commitment to reform and to publicize the results is commendable. It is encouraging that restructuring has positive effects on students' learning, at least in a district with a long and strong commitment to reform.

School Size as a Reform Issue

Although school size is a potential organizational correlate of restructuring, this structural feature is seldom seen as a policy issue per se. Findings about the effects of school size have been inconsistent because of weaknesses in the research: inconsistent definitions, inappropriate methodology, and (primarily) an unclear focus about what may be affected by a change in school size and on the process through which those effects may work. We restricted our inquiry to the influence of school size on learning. Previous research on this topic (Lee, Bryk, and Smith 1993 summarized this work) 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.

Research contrasting bureaucratically and communally organized schools has found negative effects of school size (direct or indirect) on both mean values and the social distribution of educational outcomes (Bryk and Driscoll 1988; Bryk, Lee, and Holland 1993). Indirectly, size affects school organization by influencing role specialization, social relations, differentiation of the curriculum, or all three. Small size is seen as important in reforming urban schools (Fine 1994). Prominent on a list of suggested school reforms for “expanding opportunities for mobility and enrichment” was the Carnegie Foundation’s (1992:79) recommendation of “breaking up large schools into smaller units.”

Most research on the effect of school size on students’ development has supported a shift toward smaller high schools.

Research Hypotheses

Although reforms of high schools have been considered for many decades, evaluations of the effects of such reforms have been attempted only recently. One reason for this “evaluation delay” is the plethora of difficulties in doing the type of research that we characterize as “school-effects research.” Appendix A elaborates on some of the technical difficulties involved in the research we present in this article: selecting the right unit (or units) for analysis, appropriately timing the collection of data on students’ learning, making use of the best research designs, and using the most appropriate statistical methods. We tried to address each of these difficulties in our study. Our general purpose was twofold: (1) to investigate how reform practices that are consistent with the school-restructuring movement are taking hold in the nation’s high schools and (2) to evaluate how these organizational practices affect students’ learning. Within this evalua-
tion framework, we tested several hypotheses.

**Hypothesis 1.** Hypothesis 1 focuses on the prevalence of organizational practices that are aimed at changing the structure of high schools. Specifically, it suggests that the probability of an average high school adopting fundamental reforms aimed at moving schools toward the communal form is lower than the probability of it attempting reforms to effect change within the bureaucratic form.

**Hypothesis 2.** The second and third hypotheses focus on the nature of organizational practices. Hypothesis 2 proposes that students who attend schools with educational practices of the communal form will be affected more favorably than those who attend schools with reform practices that retain the more traditional, bureaucratic form. Furthermore, the effects will be positive in both the cognitive and noncognitive domains—on learning in several subject areas and on engagement with academic life—as well as on the social distribution of these outcomes in the direction of more social equity.

**Hypothesis 3.** Hypothesis 3 is a variation on the restructuring theme. It suggests that students in schools with structural practices that are consistent with the bureaucratic form are favored, compared to those in schools engaging in none of the practices identified as structural reforms. Effects are suggested on the same outcomes described in Hypothesis 2.

**Hypothesis 4.** Hypothesis 4 concerns the size of schools. It suggests that students who attend smaller high schools are favored by that experience, above and beyond practices considered under the "restructuring" or "traditional" definitions. It also suggests that positive effects of attending small schools occur on both cognitive and noncognitive outcomes and that attending smaller high schools engenders a more equitable distribution of learning and engagement.

**METHOD**

**Sample and Data**

**Sampling design.** We drew our sample from the first and second waves of the National Educational Longitudinal Study of 1988 (NELS), a general-purpose longitudinal study of the educational status and progress of U.S. students and schools sponsored by the National Center for Education Statistics (NCES). In 1988 about 25 eighth graders in each of about 1,000 middle-grade schools were sampled, 21,126 students in all (Ingels et al. 1989). Of that number, 17,424 students were traced to their 1,508 high schools in 1990, based on locator questionnaires they completed in 1988 (Ingels, Scott, Lindmark, Frankel, and Meyers 1992). Despite obvious difficulties, the response rates were reasonably high—81.2 percent.

**Data filters.** Our study included NELS 10th graders who fitted these data filters: (1) students had to have full cognitive test-score data from both the base-year and first follow-up; (2) there had to be data from their high schools and their teachers; (3) students had to be enrolled in public, Catholic, or elite private secondary schools; and (4) they must have been attending high schools with at least five NELS-sampled students in them. These data filters are similar to

---

2 Two types of data filters were applied to the 17,424 students and 1,508 schools in the first follow-up: We selected students with close-to-complete data of the types we needed and students in the types of schools we wanted. The filters, with resulting unweighted sample sizes for students and schools, were (1) students with base-year and first-follow-up test scores (leaving 16,334 students in 1,448 schools); (2) students in schools with school-level data (reduced to 15,550 students in 1,267 schools); (3) public, Catholic, and independent private schools only (down to 13,603 students in 1,120 schools); and (4) sampling schools with at least 5 students in them (11,794 students in 820 schools). The filter dropped mostly private schools.

The last filter was meant to assure adequate within-school sample sizes for the analyses described later. The comparison of the retained and nonretained samples showed that demographic differences were modest. Moreover, the retained sample is slightly less advantaged. The bias introduced by these filters actually favors less-select students. With only modest bias, particularly in this direction, we are confident in generalizing to
those in our earlier study of school restructuring using base-year NELS data (Lee and Smith 1993). They resulted in a sample of 11,794 sophomores in 820 high schools, averaging 14.4 students per school. Most schools (717) were public, with smaller numbers of Catholic (54) and independent (49) secondary schools.

Design weights. The base-year NELS sampling procedure oversampled certain types of students and schools (private schools and schools with high enrollments of Hispanic and Asian students). Of course, this oversampling was sustained in the follow-up. The sampling procedure necessitated the use of design weights in all analyses of NELS follow-up data. Although NCES data generally include design weights for students and schools, the 1990 NELS data provided design weights only for students. The high schools were not selected in the NELS sampling frame, but were selected by NELS students. Since our research questions focused on variation among schools, we needed school weights. The lack of school-level design weights presented us with a dilemma that we solved by constructing our own schools weights (for details, see Appendix A).

Measures

School restructuring measures. The logic underlying our measures of the restructuring construct rests on two criteria: (1) the definition of communal organization from the literature and (2) definitions of restructuring developed by CORS (Newmann 1991). Both criteria contributed to defining and operationalizing the construct. From the NELS first follow-up school survey questionnaire, completed by the school principals, we identified a set of 30 practices describing school structure. The list captures important elements of American secondary-school reform, and some items tap practices used in our earlier work on this topic (Lee and Smith 1993). Dichotomous responses were invited, indicating whether each practice was in place in the secondary school (1 = yes, 0 = no).

Classification of structural practices. We grouped the 30 practices according to their adherence to or departure from a bureaucratic structural form. Practices that reinforce a top-down power flow (from administrators to teachers and students) include strong departmentalization, emphasis on teachers’ expertise and specialization, emphasis on formal instructional requirements for students, and recognition programs for teachers that operate within the worker-reward paradigm. A natural extension of this structure to the external community would view parents as recipients of information provided by teachers and administrators.

We also identified three domains of practices that shift schools from the bureaucratic toward a communal form: (1) those aimed at reorganizing instruction (mixed-ability classes, a cooperative learning focus, independent study in different curricular areas, and flexible time for classes), (2) those meant to alter authority and expertise in the school (interdisciplinary teaching teams, students’ evaluations of teachers, and staff solving school problems), and (3) those aimed at personalizing relationships within the school (using parent volunteers; keeping students in the same homeroom for several years, common planning time for teachers, and schools-within-schools). These practices have a conceptual clarity when they are sorted into these groups. Together, they represent movement toward a communal school organization in the areas of instruction, authority, and human relations. They are also in line with CORS’s criteria.

3 Fred Newmann, the director of CORS, contributed to building the theory of schools as communities (see Newmann and Oliver 1967). He also directed the earlier U.S. Department of Education-sponsored National Center on Effective Secondary Schools, which sponsored Bryk and Driscoll’s (1988) study.

4 There certainly could be differences in the classification of these practices. For example, independent study could be seen as a way to differentiate instruction. However, we assumed that this practice showed a responsiveness to students’ interests and a desire to make instruction meaningful.
Testing Hypothesis 1

The dominance of the comprehensive high school model in American education suggests that communal reform practices would be less common than reforms within the traditional bureaucratic model. To test Hypothesis 1, we computed the probability of each practice in the 820 high schools in our sample (the proportion of average NELS high schools reporting the practice). Individual item probabilities ranged from .09 to .69. We found that the average school had 12 of the 30 reform practices in place. The individual structural practices, ranked by the probability with which they occur in U.S. high schools, are listed in Table 1, along with the names of variables in the NELS school data file that were used to identify them.\(^5\)

The results in Table 1 indicate that practices fitting our definition of restructuring are the least common reforms in U.S. secondary schools. The grouping of practices and the probabilities of their occurrence support Hypothesis 1: Communal reform efforts are indeed less common than bureaucratic reforms. We grouped together the structural practices that represent communal reforms, which are those representing a substantial departure from conventional practice (in terms of low frequencies), and labeled them "restructured practices." Practices that adhere to the bureaucratic 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 (a drop from probabilities of .56 to .46 in Table 1). However, neither reform category represents a shift from traditional forms of instruction, authority, or human relations. Nor are such practices unusual in high schools.

Classification of schools. Table 1 describes reform practices, but our study design required us to categorize schools in terms of current practices, particularly reforms that capture our definition of restructuring. Although our aim was to evaluate the effectiveness of the restructuring reform movement, it was difficult to do so because only a tiny fraction of American high schools could actually be called restructured. Therefore, we shifted our attention to evaluating the effects on students of attending schools that are implementing reform practices that are consistent with our definition of restructuring.

Exploratory analyses of schools focusing on the reform practices listed in Table 1 revealed that 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 that did so were also likely to adopt more than one restructuring practice and to adopt several traditional or moderate practices. On the other hand, we were surprised that more than a trivial number of schools reported having implemented none of the 30 practices listed in Table 1. On the basis of such exploratory analyses, we classified the 820 NELS schools into the following categories:

- **Unrestructured schools.** Twelve percent (97) of the 820 schools engaged in none of the 30 practices listed in Table 1.
- **Schools with traditional practices.** Forty-two percent (346) of the schools engaged in one or several moderate or traditional practices, but did not engage in a meaningful number of practices that we classified as consistent with restructuring.
- **Schools with restructuring practices.** Forty-six percent (377) of the schools reported having at least three of the restructuring practices in place, as well as several practices we listed as traditional or moderate.

\(^5\) The probabilities listed in Table 1 are not raw probabilities for the 820 schools in the sample but represent the probability that an average high school would engage in each practice. We computed these adjusted probabilities as follows: Since the average number of practices these schools engaged in was 12 (out of 30), we defined the average high school as one that had adopted 11, 12, or 13 of these practices. We then reestimated the frequency of each practice for these average high schools. These frequencies, represented as proportions, are the probabilities displayed in Table 1. Compared to raw probabilities, the order of the variables is identical.
Table 1. Frequency of Structural Practices in the Three Types of School \((N = 820 \text{ schools})\)

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

\(^a\) Each figure in this column represents the probability that an average high school (one that reported that it adopted 11-23 of the 30 reform practices listed here) engages in each practice.

\(^b\) School aggregate created from NELS variable F1T24. Schools coded 1 if at least 70 percent of the surveyed science and mathematics teachers responded that their classes were composed of students of "widely differing achievement levels"; coded 0 otherwise.

Any group that includes almost half the sample can hardly be called unusual. However, grouping schools in this way held together conceptually and statistically.\(^6\) Focusing on the practices listed in Table 1 as being consistent with restructuring, we investigated the proportion of schools that engaged in several of them simultaneously and found that as the number of practices increased, the proportions dropped steeply. This finding suggests that it is difficult to sustain many restructuring reforms. A fifth of the schools (20.2 percent) reported more than four practices on the list, only 6 percent reported more than six (or half the practices), and no school engaged in more than nine. This is not to say that adopting many reforms of this type, simultaneously, is necessarily good for students, only that it is rare.

Why did we choose three reforms? Our logic for deciding this cutoff point is explained by the procedure we followed (see Appendix A). We review the logic of this decision later, but here we preface that discussion by stating that a "choose any three from this list" strategy is not supported by our results.

We used this categorization of schools-by-practices to organize descriptive in-

\(^6\) We found that the proportions of schools and students in each of the three categories of schools was almost identical before and after the data filter dropping fewer than five NELS students was applied. This finding suggests that this restriction introduced no bias in terms of our major hypotheses.
formation on students and schools (see Tables 2 and 3), as well as to construct the major contrasts 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) restructuring (also compared to traditional reforms).

School size. A movement away from large comprehensive high schools is consistent with efforts to restructure schools along other dimensions. Thus, in addition to the restructuring contrasts, we also investigated the effects of the size of high schools. Because the distribution of the variable measuring total school enrollment was negatively skewed (there were a considerable number of small high schools), we used a logarithmic variable transformation for school size in multivariate analyses. We focused on school-enrollment effects because we consider small school size a facilitating factor for school restructuring. Simply, the communal organizational form is more common and probably easier to implement in smaller schools.

Outcomes. We considered five dependent variables: 8th–10th-grade gains in achievement in mathematics, reading, history, and science and engagement with school. The achievement measures, change scores drawn from standardized multiple-choice tests, are different from the authentic forms of student achievement that Newmann (1991) advocated as the ideal outcomes to evaluate restructuring effects. As cognitive outcomes, we used simple differences between 8th- and 10th-grade test scores from the NELS file. Our method of constructing these scores, discussion of the appropriateness of their use as outcomes, and some caveats about gains scores are detailed in Appendix A.

Our standardized factor measuring engagement includes eight components with several of the same qualities: They measure students’ behaviors and attitudes about their current high school classes, reflect the frequency of these attitudes and behaviors, and are associated with the same subject areas measured by the NELS tests. Two related constructs were included: working hard in school and feeling challenged in school. The composite had high reliability (alpha = .84). Because the NELS items measuring engagement in the 10th grade were different from the 8th-grade measures, our engagement analyses were structured as covariance models.

Control measures. All multivariate analyses included control variables on students and schools. Student-level controls were socioeconomic status (SES) (standardized at \(M = 0, SD = 1\)), minority status (Hispanic or Black = 1, nonminority = 0), and sex (female = 1, male = 0). We included controls for initial status (ability and engagement) in the eighth grade in analyses for all five outcomes. Analytic models also included a common set of controls for the demographic and structural characteristics of schools. School demographic controls were average school SES, school minority concentration, and sector (two dummy variables contrasting Catholic or independent 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 that characterize the academic nature of high schools. The first control measured academic emphasis as the average number of mathematics and science courses that students took in the 9th and 10th grades (summed and aggregated). The second, a measure of course-taking differentiation in schools, was the aggregated standard deviation of the academic emphasis variable. School size was included, which we view as a structural, rather than a control, variable. Details of all variables—their components, construction, and psychometric properties—are presented in Appendix B.

Analytic Approach

Descriptive analyses. The study’s focus was on school structure, defined by the reform practices that guided our creation of the three categories: schools with no reform practices, schools with traditional practices, and schools with restructuring practices. Descriptive anal-
yses are presented as means for variables used at two levels (students and schools) for the three groups of schools. We tested mean differences between groups with one-way analysis of variance (ANOVA) with two contrasts: (1) between schools without reform practices and (2) schools with restructuring practices, each compared to schools with traditional practices. Two sets of means are presented: for variables describing students (both independent and dependent variables) and for variables describing schools.

**Multivariate analyses.** In addition to investigating the influence of school structural features on adjusted school means of the outcomes considered here, we also considered how structural features of schools affect the social distribution of these outcomes in each school. Increasingly common in school-effects studies is the use of hierarchical linear modeling (HLM), the method we used. Because the statistical theory and methodological approach of HLM are described in detail elsewhere (Bryk and Raudenbush, 1992; Lee and Bryk, 1989a), only a brief summary of our HLM models is presented in Appendix C (details on HLM or the analyses are available from us on request). Therefore, the discussion focuses more on the substantive than on the statistical implications of the results.

### RESULTS

#### Students in Schools with Different Structures

Table 2 presents data on the students in the three types of schools. Considerably fewer students were in high schools without reform practices \((n = 1,280)\) than were in schools with either traditional \((n = 5,353)\) or restructuring practices \((n = 5,161)\).

**Dependent variables.** For the outcomes, the means of students in schools without reform practices generally are significantly lower than those for students in traditional-practice schools. Although achievement gains are somewhat larger in restructuring than in traditional-practice schools, this difference is statistically sig-

<table>
<thead>
<tr>
<th>Variable</th>
<th>Schools without Reform Practices ((n = 1,280))</th>
<th>Schools with Traditional Practices ((n = 5,353))</th>
<th>Schools with Restructured Practices ((n = 5,161))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engagement (10th)</td>
<td>(.055^d)</td>
<td>(-.051)</td>
<td>(-.003)</td>
</tr>
<tr>
<td>Mathematics gain ((8 \rightarrow 10))</td>
<td>(4.74^d)</td>
<td>(5.28)</td>
<td>(5.49)</td>
</tr>
<tr>
<td>Reading gain ((8 \rightarrow 10))</td>
<td>(2.18^d)</td>
<td>(2.53)</td>
<td>(2.39)</td>
</tr>
<tr>
<td>History gain ((8 \rightarrow 10))</td>
<td>(1.92^d)</td>
<td>(2.21)</td>
<td>(2.33)</td>
</tr>
<tr>
<td>Science gain ((1 \rightarrow 10))</td>
<td>(2.11)</td>
<td>(2.26)</td>
<td>(2.57^e)</td>
</tr>
<tr>
<td><strong>Independent Variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engagement (8th)</td>
<td>(-.07^d)</td>
<td>(-.11)</td>
<td>(.24^e)</td>
</tr>
<tr>
<td>SES (^c)</td>
<td>(-.23^d)</td>
<td>(-.101)</td>
<td>(.048^e)</td>
</tr>
<tr>
<td>Minority status (%)</td>
<td>(34.9^d)</td>
<td>(15.9)</td>
<td>(20.8^e)</td>
</tr>
<tr>
<td>Female (%)</td>
<td>(52.5)</td>
<td>(50.6)</td>
<td>(50.2)</td>
</tr>
<tr>
<td><strong>Ability controls (8th grade)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test composite (^c)</td>
<td>(-.20^d)</td>
<td>(-.05)</td>
<td>(.14^e)</td>
</tr>
<tr>
<td>Reading control (^c)</td>
<td>(-.21^d)</td>
<td>(-.04)</td>
<td>(.14^e)</td>
</tr>
<tr>
<td>History control (^c)</td>
<td>(-.20^d)</td>
<td>(-.05)</td>
<td>(.13^e)</td>
</tr>
<tr>
<td>Mathematics control (^c)</td>
<td>(-.20^d)</td>
<td>(-.05)</td>
<td>(.13^e)</td>
</tr>
<tr>
<td>Science control (^c)</td>
<td>(-.21^d)</td>
<td>(-.05)</td>
<td>(.14^e)</td>
</tr>
</tbody>
</table>

\(^a\) School sample sizes were computed unweighted.
\(^b\) Variable means and contrasts were computed with the NELS student-level design weights.
\(^c\) Variables are standardzied, \(M = 0, SD = 1\). Other variables are presented in raw metrics.
\(^d\) Contrast of no-reform versus traditional-reform schools is statistically significant at \(p = .05\) or below.
\(^e\) Contrast of restructured versus traditional-reform schools is statistically significant at \(p = .05\) or below.
significant only for science. In general, the magnitude of differences in achievement gains is small (the largest group differences are no more than .10 or .15 SD).\textsuperscript{7}

**Independent variables.** Students’ SES and minority-status distributions follow a pattern similar to the outcomes. Students in schools without reform practices are of a lower SES and are more likely to be minority-group members than are students in traditional-practice schools, whereas students in restructuring-practice schools are more advantaged than are those in traditional-practice schools. The same pattern applies to students’ ability and engagement in the eighth grade: Students in schools with no reform practices rank below, and those in restructuring-practice schools rank above, those in schools with traditional reforms. The distribution of female students is comparable for the three types of schools. In general, the magnitude of group differences in the control variables, though not large, is greater than the outcomes—between .15 and .35 SD.

The observed means of several school characteristics for the three types of schools are presented in Table 3. There were fewer schools without reform practices ($n = 97$) than with either traditional ($n = 346$) or restructuring-reform practices ($n = 377$).

**School demographics.** Reflecting the students’ characteristics in Table 2, schools without reform practices are significantly disadvantaged compared to traditional-practice schools in social demographics (average SES and minority enrollment). Restructuring-practice schools are advantaged compared to the same group in average SES (a .7 SD difference between no-

\textsuperscript{7} The means in Table 2 are between students in the three types of schools. Thus mean group differences presented as effect sizes (ESs) are not comparable to those presented later, where ESs are between schools. To compute ESs in the two instances, we used different SD units in the denominators. In general, school-level SDs are smaller, so the effects are larger. Although both types are appropriate for the circumstances in which we used them, this distinction should be kept in mind.

reform and restructuring-practice schools), in minority enrollment (34 percent in no-reform schools and 16.4 percent in reform-practice schools are minority), and in average eighth-grade achievement (a .75 SD advantage for restructuring-reform-practice schools).

The distribution by sector explains some of the social-background differences. Although private schools represent 14 percent of the school sample, less than 1 percent of Catholic and elite private schools (members of the National Association of Independent Schools, NAIS) are classified as schools without reform practices. A relatively high proportion of restructuring-practice schools are private (10.4 percent Catholic and 7.0 percent independent). In fact, 52 percent of the Catholic schools and 74 percent of the independent schools are so classified. 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’ course work in mathematics and science. Schools without reform practices have significantly less course taking, and restructuring-practices schools have significantly more. Along with the variation in academic emphasis are differences in the homogeneity of course taking: Schools without reform practices are characterized by more variability in course taking than are traditional schools, whereas restructuring and traditional schools have similar variability.

School size does not follow the same pattern. Traditional-practice schools are smaller than either restructuring or no-reform schools, although the average size of schools is reasonably similar across groups (1,000–1,500 students). The distribution of school size is positively skewed; that is, there are more small than large high schools (of course, most students attend large schools).

Tables 2 and 3 demonstrate the considerable differences in the three types of schools and the students who attend them: demographic differences among students and schools, sector differences, differences in initial cognitive and non-cognitive status, and differences in the
Table 3. Means of Variables Describing the Characteristics of the Three Type of Schools (N = 11,820 students)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Schools without Reform Practices (n = 97)a,b</th>
<th>Schools with Traditional Practices (n = 346)a,b</th>
<th>Schools with Restructured Practices (n = 377)c,b</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average SES</strong></td>
<td>-.41d</td>
<td>-.18</td>
<td>.30e</td>
</tr>
<tr>
<td>% Minority enrollment</td>
<td>33.5d</td>
<td>12.9</td>
<td>16.4e</td>
</tr>
<tr>
<td><strong>School Sector</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Public</td>
<td>98.7d</td>
<td>95.5</td>
<td>82.6e</td>
</tr>
<tr>
<td>% Catholic</td>
<td>0.8d</td>
<td>4.1</td>
<td>10.4e</td>
</tr>
<tr>
<td>% NAIS</td>
<td>0.5</td>
<td>0.5</td>
<td>7.0e</td>
</tr>
<tr>
<td><strong>Average Achievement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8th-grade mathematics and reading</td>
<td>-.48d</td>
<td>-.15</td>
<td>.27e</td>
</tr>
<tr>
<td><strong>Average number of mathematics and science courses</strong></td>
<td>2.37</td>
<td>2.48</td>
<td>2.69e</td>
</tr>
<tr>
<td>Variability in mathematics and science course taking (SD)</td>
<td>1.46d</td>
<td>1.32</td>
<td>1.16e</td>
</tr>
<tr>
<td>School size</td>
<td>1.095d</td>
<td>633</td>
<td>764e</td>
</tr>
</tbody>
</table>

a School sample sizes were computed unweighted.
b Variable means and contrasts were computed with the NELS constructed school weight.
c Variables are standardized, M = 0, SD = 1. Other variables are presented in raw metrics.
d Contrast of no-reform versus traditional schools is statistically significant at p = .10 or below.
e Contrast of restructured versus traditional-reform schools is statistically significant at p = .10 or below.

academic emphasis of the schools. In fact, differences across the groups between the control variables we included in our multivariate analyses exceed differences in students' cognitive gains. Such variability among the characteristics of students and schools indicates the importance of multivariate analysis.

**Multivariate Analyses**

Unconditional HLM models. In any HLM analysis, the first step is to partition the variance of each dependent measure into within- and between-school components by estimating an HLM model that includes no control variables for either schools or students. The effects of school structure may be evaluated only on the portion of variance that is between schools. Table 4 presents the results for these unconditional HLM models.

For each outcome, the proportion of variance between schools (the intraclass correlation) is not large (20 percent or less). Although this proportion is adequate for multilevel analysis, clearly the majority of variance in achievement gains and engagement is within schools, rather than between schools. The intraclass correlation is the lowest for engagement (.13) and the highest for gains in science (.20). HLM-estimated reliabilities for the gain scores (and engagement) are low (ranging from .27 for engagement to .50 for gains in science). The outcomes obviously have less-than-optimal psychometric properties. However, they measure change in each subject for each student over the first two years of high school, which suggests that they represent ideal outcomes on which to evaluate the effects of school-restructuring.

---

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

9 Reliabilities estimated with HLM are not the same as those estimated with 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).
Table 4. HLM Estimates of Psychometric Properties of Dependent Variables (N = 11,794 students)\(^a\)

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>Academic Engagement</th>
<th>Gain in Mathematics</th>
<th>Gain in Reading</th>
<th>Gain in History</th>
<th>Gain in Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within-school variance</td>
<td>.95</td>
<td>40.81</td>
<td>25.55</td>
<td>11.75</td>
<td>13.62</td>
</tr>
<tr>
<td>(sigma-squared)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between-school variance</td>
<td>.04</td>
<td>2.89</td>
<td>1.22</td>
<td>0.76</td>
<td>1.69</td>
</tr>
<tr>
<td>(tau)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliability</td>
<td>.27</td>
<td>.36</td>
<td>.28</td>
<td>.34</td>
<td>.50</td>
</tr>
<tr>
<td>Intraclass correlation(^b)</td>
<td>.13</td>
<td>.16</td>
<td>.15</td>
<td>.16</td>
<td>.20</td>
</tr>
</tbody>
</table>

\(^a\) All parameters in this table are computed with a fully unconditional HLM model (a model that 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.

practices on student learning. We acknowledge a tradeoff between less-than-perfect measures in a statistical sense for close-to-perfect measures in a conceptual sense. However, the modest psychometric properties could constrain the ability to locate school effects.

Within-school HLM models. Table 5 presents within-school HLM models for the five outcomes: school engagement and gains in mathematics, reading, history, and science.\(^10\) Conceptually, within-school HLMs are somewhat analogous to many small ordinary least-squares (OLS) regressions, one in each of the 820 schools. Each model estimates the effect of students’ demographics (SES, minority status, and gender) and of engagement in eighth grade (in HLM, these are beta coefficients) within each school. We included a control for ability at the point of high school entry, although the control was slightly different for each outcome (see Appendix B for details). HLM effects in Tables 5 and 6 are presented as ESs (SD units).\(^11\) Because of modest samples within schools, we estimated only one “slope as outcome,” the relationship between SES and each outcome. The between-school variance in other control variables is set to zero. Procedurally, this means that we centered the SES variable on each school’s mean and other controls on the population means (a z-score variable).

Many effects of control variables are large. SES is positively related to 10th-grade engagement (ES = .77) and to gains in mathematics and science (ES = .14 and .12, respectively) but has only marginal effects on 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 (ES = -.12). Gender effects are inconsistent: Although females are very positively engaged (ES = 1.22), males are favored in science gains (ES = -.37). Other small and nonsignificant gender effects favor males. Engagement in 8th grade is positively related to all outcomes and is especially large for engagement in 10th grade (ES = 1.16).

\(^10\) Effects in Tables 5 and 6 are not directly comparable to those that could be computed as group differences from Tables 2 and 3 for two reasons. First, effects in Tables 5 and 6 are in SD units (to allow comparison across outcomes with different metrics), whereas gain scores in Tables 2 and 3 are in a raw IRT (Item Response Theory) metric. Second, and more important, the HLM analyses displayed in Tables 5 and 6 use only the between-school proportion of the gain scores as outcomes. The gain scores in Tables 2 and 3, on the other hand, contain the within-school, the between-school, and the error variance in these outcomes.

\(^11\) 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 ES metric. In common usage, ESs of .1—.2 SD are small effects, .2—.5 SD are medium, and over .5 SD are large (Rosenthal and Rosnow 1984: 360).
Table 5. HLM Within-School Model for Academic Engagement and Achievement Gains (N = 11,794 students)

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Academic Engagement</th>
<th>Gain in Mathematics</th>
<th>Gain in Reading</th>
<th>Gain in History</th>
<th>Gain in Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (school average)</td>
<td>.00*</td>
<td>3.22***</td>
<td>2.22***</td>
<td>1.76***</td>
<td>2.86***</td>
</tr>
<tr>
<td>8th-grade engagement</td>
<td>1.16***</td>
<td>.16***</td>
<td>.08*</td>
<td>.05*</td>
<td>.07*</td>
</tr>
<tr>
<td>8th-grade ability</td>
<td>.48***</td>
<td>-.04*</td>
<td>.07*</td>
<td>-.07**</td>
<td>.21***</td>
</tr>
<tr>
<td>SES&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.77***</td>
<td>.14**</td>
<td>.04*</td>
<td>.11*</td>
<td>.12**</td>
</tr>
<tr>
<td>Minority status</td>
<td>.82***</td>
<td>-.04</td>
<td>-.23*</td>
<td>-.12*</td>
<td>-.68***</td>
</tr>
<tr>
<td>Gender (female)</td>
<td>1.22***</td>
<td>-.10</td>
<td>-.04</td>
<td>-.07</td>
<td>-.37***</td>
</tr>
</tbody>
</table>

<sup>p < .01, * p < .05, ** p < .01, *** p < .001.</sup>

<sup>a</sup> All effects in this table are presented in a standardized metric, computed by dividing the HLM gamma coefficient for each outcome by the adjusted school-level SD of that outcome computed by HLM. These SDs are displayed at the bottom of Table 6.

<sup>b</sup> The 8th-grade ability control is different for each outcome. For engagement, it is a composite of reading and mathematics 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.

<sup>c</sup> In the HLM model, SES is allowed to vary randomly between schools, while the other controls are used as fixed parameters. SES is centered on the sample mean, while the other controls are centered on their respective school means.

The effects of cognitive status (ability) are modest for most achievement gains but strong for engagement (ES = .48) and moderate for gains in science (ES = .21). Ability is significantly and negatively associated with gains in mathematics and history (ES = -.04 and -.07, respectively) and unrelated to gains in reading. These findings indicate that initially, more-able students become even more engaged with school in high school and learn somewhat more in science (ES = .21); however, initially less-able students seem to learn a bit more in mathematics and history. Negative correlations between prior ability and gain were expected, but their small magnitude suggests that IRT scaling was successful.

**Between-school HLM analyses.** The full HLM models focus on 10 school-level outcomes, each adjusted for the within-school controls shown in Table 5. The outcomes are of two types: five adjusted school means for engagement and gains in each subject area and five within-school estimates of the relationship between SES and each outcome. Set 1 outcomes—school mean engagement and cognitive gain—may be described as parameters of "effectiveness." Schools that are high on these parameters are more effective (students are more engaged and learn more), and school characteristics that are positively associated with school means typify "good" or "effective" schools. The outcomes in Set 2, relationships between SES and cognitive gains or engagement, are parameters of "equity" (see Bryk and Raudenbush 1992). The definition of good or effective schools is expanded to include schools that simultaneously have a high effectiveness parameter and a low equity parameter (schools with high learning rates equitably distributed among students of different SES levels). Characteristics of schools would be good if they were negatively associated with the SES slopes.

Full HLM models estimate the effects of school characteristics (gamma coefficients) on these 10 outcomes (see Table 6). Although each HLM includes all the within-school controls shown in Table 5, the effects of those controls are not displayed here. To highlight the primary focus of the study, we divided the gamma coefficients into two groups: those tapping restructuring practices (below the line) and controls for demographic and academic school characteristics (above the line). Each column in Table 6 represents the HLM model for a single outcome. The results for school means on each outcome (effectiveness parameters) are in the top section of
Table 6. HLM Between-School Model for Academic Engagement and Achievement Gains (N = 820 students)

<table>
<thead>
<tr>
<th>Effects</th>
<th>Academic Engagement</th>
<th>Gain in Mathematics</th>
<th>Gain in Reading</th>
<th>Gain in History</th>
<th>Gain in Science</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>On Mean Between-School Outcome</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Average intercept | −.89***<sup>b</sup> | 2.92***<sup>b</sup> | 1.94*** | 1.71***<sup>b</sup> | 2.95***
| School SES | −.08 | −.02 | .08 | .10 | .27***
| Minority concentration | .03 | .00 | .10 | .03 | −.23***
| Catholic high school | −.07 | .49*<sup>b</sup> | .32* | −.01 | .33
| NAIS high school | 1.23*** | .18 | .44*<sup>b</sup> | .03 | .45***
| Academic emphasis | .26***<sup>b</sup> | .21**<sup>b</sup> | .18* | .13* | .56***
| Course-taking variability | −.17*<sup>b</sup> | −.07 | −.06 | −.09*<sup>b</sup> | .26

| **On SES-Differentiation**<sup>b</sup> |                      |                     |                 |                 |                 |
| Average SES slope | 1.59***<sup>b</sup> | .43***<sup>b</sup> | .51*<sup>b</sup> | .36*<sup>b</sup> | .61***
| School SES | .63**<sup>b</sup> | .39*<sup>b</sup> | .19 | .20 | .06
| Minority concentration | .25 | .26 | .44 | −.27 | −.05
| Catholic high school | −.82 | −.26 | −.68*<sup>b</sup> | −.70 | −.06
| NAIS high school | .03 | .75 | .25 | −.28 | −.70
| Academic emphasis | −.94**<sup>b</sup> | −.27*<sup>b</sup> | −.34*<sup>b</sup> | −.13 | −.46***
| Course-taking variability | .34**<sup>b</sup> | .02 | .09 | .20 | .12

| School size | .10 | .03*<sup>b</sup> | .34**<sup>b</sup> | .22*<sup>b</sup> | .25*
| Schools without reform practices | .16 | .18*<sup>b</sup> | .54*<sup>b</sup> | .33*<sup>b</sup> | .47**
| Restructuring-practice schools | −.54***<sup>b</sup> | −.33**<sup>b</sup> | −.38*<sup>b</sup> | −.32*<sup>b</sup> | −.30**

<sup>a</sup> All HLM analyses presented here also include adjustment for all within-school variables in Table 5: 8th-grade engagement, 8th-grade ability, SES, minority status, and gender.

<sup>b</sup> All effects shown in this table are presented in a standardized metric, computed by dividing the HLM gamma coefficient for each outcome by the adjusted school-level SD computed by HLM, which are as follows:

<table>
<thead>
<tr>
<th></th>
<th>Academic Engagement</th>
<th>Gain in Mathematics</th>
<th>Gain in Reading</th>
<th>Gain in History</th>
<th>Gain in Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD-intercept</td>
<td>.164</td>
<td>1.762</td>
<td>1.248</td>
<td>1.417</td>
<td>1.065</td>
</tr>
<tr>
<td>SD-SES slope</td>
<td>.083</td>
<td>.667</td>
<td>.456</td>
<td>.393</td>
<td>.351</td>
</tr>
</tbody>
</table>

Table 6, and the social distribution of these outcomes (equity parameters) are in the bottom section. These effects are estimated simultaneously in fully multivariate HLMs. To focus on the hypotheses posed earlier, our discussion is organized around the effects of particular independent variables across all the outcomes (by row), rather than for each outcome (by column).

**Effects of Restructuring Practices**

**Testing Hypotheses 2 and 3.** Schools with no reform practices and those with restructuring practices were compared separately to schools with traditional practices. The effects are consistent: Restructuring-practice schools have strong, positive, and significant effects on their students’ cognitive gains in the first two years of high school (ESs range from .35 in history to .59 in science). Our findings provide strong support for Hypothesis 2, that students attending schools with reforms consistent with the restructuring movement will learn more and be more engaged with school than will students in schools with more traditional reforms. The effects for schools without reforms (Hypothesis 3), though smaller in magnitude, are consistently negative: Students in high schools with no reform practices are less engaged and gain less in cognitive areas than do their counterparts in schools with traditional reform practices (ESs = .1–.2 SD, generally smaller than for restructuring-practice schools, .3–.6 SD). Hypothesis 3, which stated that the effects on students
in traditional-reform schools are positive compared to those on students in schools without reforms, is also supported by the findings in Table 6.

**Equity effects.** The distribution of learning is highly stratified by students' SES. The large positive coefficients for average SES-gain slopes at the top of the lower panel of Table 6 indicate that upper-class students are highly advantaged on all outcomes (the SES effects on engagement, ES = 1.59, and gains in science, ES = .61, are particularly strong). Again, the pattern of restructuring effects on social equity is consistently favorable.

The effect of schools with restructuring practices is strongly, negatively, and significantly related to all outcomes (in such schools, gains are distributed more equitably)—a large effect on the social distribution of engagement (ES = -.54) and a moderate effect on the social distribution of achievement gains (ESs = .3-.4). Displaying a similarly consistent pattern, schools without reforms are more stratifying in the distribution of achievement gains by SES (positive and significant effects range from .18 on the distribution of gains in mathematics to .54 on the distribution of gains in reading). The findings on social equity also support Hypotheses 2 and 3.

The consistent pattern of effects on learning and engagement shown in Table 6 suggests that school organization can have important consequences for students' learning. High schools that engage in the practices consistent with the restructuring-reform movement fit the set of double qualifications for good schools: They are simultaneously more effective and more equitable in terms of students’ engagement and learning.

**Testing Hypothesis 4.** The results shown in Table 6 also demonstrate conclusively that students who attend smaller schools are favored and hence support Hypothesis 4. For all parameters of effectiveness (engagement and cognitive gains), the effects of school size are negative and significant. The effects on cognitive gains are particularly notable (ESs = -.3 to -.4). Besides more learning and engagement for students in small schools, school size is also associated with social equity (see lower panel of Table 6). A pattern of positive and significant effects of school size on the relationship between SES and achievement gains (the parameters of equity) is evident. Magnitudes of significant effects ranges from small (ES = .03) on the SES/mathematics-gain slope to moderate (ES = .34) on the SES/reading-gain slope. To our knowledge, our results for school-size effects on the level and distribution of cognitive performance are among the largest and most consistent of any published studies on school effects.

**Other Noteworthy School Effects**

**Academic emphasis.** The primary purpose of including variables that measure the level and variability of the academic emphasis of high schools was as statistical controls, but the findings are meaningful. Although increasing academic course requirements is not within the purview of the restructuring movement, other recent reform efforts have been directed toward increasing high school students' enrollment in academic courses. The results here are also consistent: Academic emphasis is positively related to all outcomes (ESs range from .13 on gains in history to .56 on gains in science). More within-school variability in course taking is negatively related to these outcomes, especially to engagement (ES = -.17). In one instance,

---

12 We explored the possibility that school size had differential effects on the outcomes in different types of schools (a series of interaction terms of school-level variables with school size). We found no evidence of any significant school-size interactions, either singly or as a group.

13 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. Thus, all HLM analyses were run with and without weights. Consistently, the pattern 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 the magnitude of coefficients.
science, variability in course taking has a positive effect (ES = .56). Although not all effects on the equity parameters are significant, the pattern is generally similar to the effects for restructuring-practice schools: Schools with a consistently greater academic emphasis are more equitable environments. Although the effects are smaller, the findings suggest that inconsistent course-taking patterns in schools increase social stratification in learning and engagement. High schools with constrained curricula, in which all students take a similar set of academic offerings, appear to increase the learning of all their students.

School sector. The results in Table 3 indicate that restructuring practices are especially likely to be part of the organizational structure of private high schools (both Catholic and NAIS). Private schools are also typically smaller than public schools, academic emphasis is higher, and course taking is less variable (Coleman, Hoffer, and Kilgore 1982). The purpose of controlling for school sector, beyond the practices and structure typical of them, was to address a possible alternative explanation for the findings of this study. The residual effects of school sector in Table 6 are not consistent.14 We believe that the inclusion of several key features of private-school organization in our models (such as their small size and greater academic emphasis) largely explained away the residual sector effects on effectiveness and equity shown in some studies of school effects. Therefore, readers should not draw substantive conclusions from the sector effects in Table 6.

School social composition. School SES and minority concentration are generally unimportant in predicting gains (see Table 6), except for a moderate effect for school SES on gains in science (ES = .27). School SES is, in our view, a proxy measure of school resources; thus, the finding suggests that students learn more science in schools with more resources.15 Given the expense of laboratories and scientific equipment, this result makes sense. The social distribution of engagement appears much more stratified in schools with more affluent students (ES = .63). Although compositional effects on the equity parameters are sizable in some cases, few attain statistical significance. We conclude that school social composition has few effects on students’ learning, net of the considerable stratification by students’ social background on achievement gains within schools (Table 5). An exception is science learning, for which more school resources are important.

DISCUSSION

Summary of Findings

Gains in achievement. Under the theoretical contrast of bureaucratic and communal organization, we used two dimensions to evaluate high school restructuring. The first dimension focused on organizational practices, which we categorized by their form (communal or bureaucratic) and included the idea of a “substantial departure from conventional practice.” We grouped high schools, by the number and type of reform practices, into three groups: restructuring-practice schools, traditional-practice schools, and schools with no reform practices. The second dimension focused on the size of high schools; here we considered smaller size as a feature of school structure that moved schools toward a communal form.

When school effects are evaluated in terms of the types of practices that tap organizational structure and school size, the pattern of effects on cognitive learning in the first two years of high school is clear and consistent: Students who attend schools with several practices in

14 For example, although 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 mathematics gain but not on other cognitive gains. NAIS effects on cognitive gains are all positive, some large, but nonsignificant.

15 Unfortunately, the NELS data include no direct measures of school resources, such as per-pupil expenditures. Thus, the proxy measure of schools’ average SES is the only available measure of resources.
place that are consistent with the restructuring movement learn significantly more in reading, mathematics, history, and science, whereas those who attend schools without reform practices learn significantly less. Equally important is the finding that students in smaller schools learn more in these important areas of the curriculum. These results are net of differences in the schools' academic and social character, as well as the academic and social characteristics of their students.

**Social equity in students’ learning.** In addition to achievement gains in the early years of high school, we also considered how attending certain types of schools affects the distribution of these gains within each school among students from different social backgrounds. The effects in this area are also consistent: Achievement gains in the four subjects we considered are more equitably distributed in smaller schools. Schools with restructuring practices are more equalizing environments in terms of the social distribution of cognitive gains, and schools with no reforms are more stratifying environments.

The consistency of the findings allows us to make some unequivocal statements about the organizational structure of high schools: Students learn more in smaller schools and in schools with several practices that are consistent with restructuring. Schools that are organized in this way are also more equitable environments in terms of the internal distribution of cognitive learning. Conversely, in schools without such reforms and in larger schools, students learn less, and learning is more stratified. General levels of learning are lower in those schools, and disadvantaged students learn even less.

**Student engagement.** Our findings regarding students’ engagement with academic learning in the first two years of high school are consistent with those for cognitive development. Students in small schools are more engaged in their courses, and engagement is higher and more equitably distributed in schools in which restructuring practices are common. Our findings on engagement take into account students’ academic status and engagement on their entry into high school.

**Caveats.** We offer some substantive and methodological cautions about our findings. A substantive limitation is that although we considered 30 reform practices, it is clear that this list of potential reforms that may be captured under the “restructuring” rubric is far from complete. Moreover, we used the bureaucratic-communal perspective to classify these practices—one that not everyone would agree with. Our list was, of course, confined to data available from the NELS survey of high school administrators. These limitations restrict the generalizability of our results to the restructuring practices we used and the classification of schools we followed.

Other caveats concern the modest statistical properties of the outcome variables examined here: (1) low variability among schools and (2) the modest reliability of the gain scores. Since school organizational properties may explain only between-school variability in outcomes, our results show strong effects on the moderate proportion of variability in engagement and learning that is between schools (13–20 percent). We also suggest the need to balance the measures’ modest psychometric properties against the conceptual clarity of their meaning—change over time in achievement is an essential element of learning. The substantive interpretation of our caveat about modest reliabilities suggests that these effects probably represent lower bounds to true effects that might be obtained if gain scores were measured with perfect reliability.

Another set of caveats concerns timing. Multiple measures of the same construct (achievement) available on the same students over time increases our confidence in the study’s results—a major advantage of the NELS design. Countering the study’s strength derived from longitudinal data is the lack of information about the timing of the implementation of the reform practices. The data structure suggests that we were studying the effects not of restructuring but of solidly established school organizations. The structure of our analyses supposes that reform came first and that
students’ learning was the result. We admit, however, to a possible alternative causal order. We hope that the data from the second follow-up of NELS can help untangle any confusion about temporal sequencing.

Revisiting the Concept of Restructuring

What conclusions can be drawn from these results about the reform of the organization of U.S. high schools? At the outset, we argued that the reform practices that we grouped together as being consistent with the restructuring movement represented a move toward a communal organizational form and away from a bureaucratic form. Although we admitted that these restructuring practices aim to change different school functions (instruction, autonomy, or community), we argued for their conceptual commonality when they are seen as efforts to reform schools in a particular direction. That grouping structural practices in this way had strong and consistent effects on important student outcomes provides empirical support for some internal consistency among the reforms.

How many reforms? We are concerned that the results may lead practitioners to a conclusion that we do not think is justifiable: “Pick any three reforms from this list, and a high school will be more effective and more equitable.” We provided a rationale for classifying schools in the restructuring-practices group if at least three reforms of this type were in place. To explore the implications of this decision, we conducted a sensitivity analysis, in which we estimated the effects on students of schools implementing particular numbers of these reforms. The results of the sensitivity analyses for the effectiveness parameters (gains in achievement and engagement) and the equity parameters (the social distribution of achievement and engagement) are presented in Figures 1 and 2, respectively; details of the procedures are in Appendix A.

The results of the sensitivity analyses clearly demonstrate that the simultaneous implementation of many restructuring reforms (more than three on most outcomes) does not increase either effectiveness (Figure 1) or equity (Figure 2) and that implementing one or two reforms at a time may not be advantageous. In terms of numbers, these results support our decision to use three reforms as a cutoff point. It seems clear that schools with a commitment to the

Figure 1. Sensitivity Analysis on Effectiveness: Effects of the Number of Restructuring Reforms on Gains in Achievement and Engagement

Effect Size (SD)

<table>
<thead>
<tr>
<th></th>
<th>Engagement</th>
<th>Gain in Math</th>
<th>Gain in Reading</th>
<th>Gain in History</th>
<th>Gain in Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Reforms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Restructuring practices we defined should decide on a modest number of reform strategies and should not attempt too many reforms simultaneously.

These findings are supported by two recent studies of school restructuring in local contexts. The Kentucky study (Kyle 1993) reported that schools in the early stages of reform, particularly those that dabbled in a wide range of reforms, had less advantageous outcomes than schools that engaged in no reforms and considerably less positive outcomes than schools with sustained commitments to fewer reform strategies. These findings extended to the elementary, middle, and secondary levels.

Bryk, Easton, Kerbow, Rollow, and Sebring (1993) studied the emerging effects of school reform in Chicago’s elementary schools (K–8). Some schools were labeled “Christmas trees . . . ‘showcase’ schools with many new programs, multiple ‘add-ons’ with little coordination, and little attention to strengthening the organizational core” (p. 15). Such schools, characterized by their “unfocused academic initiatives,” compared unfavorably with other schools with a more systematic approach to reform. The latter group had 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 that attempts to embrace too many reforms—perhaps to give the appearance of climbing on the reform bandwagon—are counterproductive.

Is reducing school size really the issue? Although the restructuring practices investigated here have some implications for school reform (or the conceptual direction of reform), does it make sense to consider the findings about school size in a similar light? That is, would reducing the size of high schools really “cause” students to learn more? Although the structure of our analyses, in which we estimated the direct effects of school size, suggested that it would, we would not draw that conclusion from our results. In fact, we suspect that none of these schools had recently reduced their size as a reform strategy. Rather, the findings indicate that the size of enrollments acts as a facilitating or debilitating factor for other desirable practices. For example, collegiality among teachers, personalized relationships, and less differentiation of instruction by ability (to name a few organizational features of schools) are more common.

Figure 2. Sensitivity Analysis on Equity: Effects of the Number of Restructuring Reforms on SES Slopes on Gains in Achievement and Engagement

Effect Size (SD)

<table>
<thead>
<tr>
<th>Number of Restructuring Reforms</th>
<th>Engagement</th>
<th>Gain in Math</th>
<th>Gain in Reading</th>
<th>Gain in History</th>
<th>Gain in Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
and easier to implement in small schools. However, reducing the size of schools, although a potential structural reform in its own right, would not increase students' learning per se.\textsuperscript{16}

We suggest that school size can have only an \textit{indirect} effect on students' learning and engagement. If we had introduced a set of school organization measures, such as collegiality and personalized relationships, into our analyses, the magnitude of the direct effects of school size would surely have declined. 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 on 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 U.S. high schools (see Table 1)—is a feasible and cost-effective way to accomplish this structural reform and to facilitate other useful changes.

There is some danger that the school-within-a-school reform could be motivated by attempts to introduce more specialization into high school programs, rather than to create more personalized environments. In fact, some reform groups, such as the Coalition of Essential Schools (Sizer 1984), have argued against this reform for this reason. Thus, the purpose that drives high schools to embrace this reform should be examined.

\textbf{Survey data and research on school structure.} As was mentioned in our study of restructuring in middle 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 restructuring category (Table 1) are hardly cutting edge. It is distressing that so few high schools use parent volunteers, focus on cooperative learning, team their teachers across disciplines, offer mixed-ability classes in mathematics and science, provide teachers with common planning time, offer flexible time for classes, or keep students in the same homeroom over several years. Despite growing research support for the effectiveness of these practices, which as a group seem to transform high schools into more humane places in which students' experiences with adults and with one another are more personal, few American high schools were implementing them in 1990. A disturbing finding is how bureaucratic the structure of the American high school is. In real terms, restructuring (as we would like to define it and not how we have tried to tap it) is rare.

Although survey methods are well suited to investigate the frequency of organizational practices, they are not particularly good for studying the \textit{implementation} of such practices. Although NELS includes reports from principals about whether the practices occur in each high school, it does not include information on the intensity and pervasiveness of these practices or on the support for them. Thus, in a school that reported that it offers independent study in mathematics or science or has interdisciplinary teaching teams, for example, these practices may actually affect only a small number of students and teachers. Most teachers could oppose these reforms, with only a small and self-selected group actually attempting them. Moreover, we have no idea whether the students in the NELS study were actually engaged in practices that, in theory, involve students (such as mixed-ability classes and independent study). With an average of fewer than 15 sam-

\textsuperscript{16} 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. Since the school-size variable was transformed, the results suggest a linearity in relationships that may not entirely reflect reality. It is also unclear whether the 15 percent of principals who reported schools-within-schools reported the size of these "schools" or of the overall schools. Thus, our results should be seen as providing general support for reducing the size of high schools, rather than suggesting specific figures for the optimal size.
pled sophomores in each school in our sample, unless the practice was almost universal (which is certainly possible), the sampled students might not have actually experienced the programs that the schools reported they offered.

The very constraints of using survey methods to study the phenomenon of school reform may underscore the importance of our findings. High schools that engaged in as few as three restructuring practices had powerful effects on their students' learning and engagement, and the small proportion of American high schools that engaged in none of the practices had negative effects on their students. Thus, the school effects we described here probably represent lower bounds for the actual effects of school restructuring on students' engagement and learning.

CONCLUSION

The policy implications we draw from this study are broad, rather than narrow and prescriptive. Clearly, something important is going on inside the schools with several restructuring practices because their students demonstrate more learning that is also more equitably distributed. Thus, our results provide empirical support for a broad policy of school reform that would move schools toward a communal organizational form and away from the bureaucratic form that has characterized the comprehensive high school for over a century. They also suggest that schools should target their reform efforts to a modest number of communal practices—practices that probably should be adopted neither singly nor in large numbers. Smaller schools with more personal environments and greater commonality of students' academic and social experiences help to engage students in learning and keep them in school. This is powerful evidence arguing for a fundamental restructuring of school (Bryk 1994:7)

Another commentator focused on human relationships:

Structures such as those collected here under the rubric “restructuring practices” make a difference in student achievement and engagement when they support personal and sustained connections between students and adults in the school setting, and when they facilitate the sharing of knowledge about students as individuals and learners. (McLaughlin 1994:9)

APPENDIX A

TECHNICAL ISSUES

Unit of analysis. Until the 1980s, research on school effects misconceived the ques-
tions and methods for investigating them. Because such questions are by nature multi-level, efforts to quantify how schools affect students require methods that capitalize on this nesting. Until appropriate statistical methods were available, data were analyzed either at the level at which the major proportion of variation in outcomes occurred (between students), thereby assuming that students in the same school are independent of each other in how school factors influence them, or at the level at which the “treatment” was administered (between schools), assuming no variation among students in school effects. Statistical and substantive problems plague both approaches.

Timing of data collection. Studies that have used longitudinal data from High School and Beyond (HS&B), even those using appropriate methods, were restricted to estimating school effects during the last two years of high school. Thus, they ignored potentially powerful school effects at the beginning of high school. It seems possible that organizational effects may not be constant over the high school years, but they could accrue early and level off thereafter. This point was made by Lee and Bryk (1989b) about the selection of appropriate statistical controls in HS&B data. We hope that this study overcame these difficulties, for the reasons spelled out in the text and in this appendix.

Data and statistical needs. Quantitative research evaluating school effects on students has stringent data requirements: large samples at two nested levels (students and schools). To capture how schools affect students’ learning, we also needed measures of academic status at two or more time points, preferably (1) on entry into the school and (2) after a lengthy exposure to the organizations whose effects were being evaluated. In addition, school effects on learning may vary by subject area, so we needed measures of achievement across the curriculum. Since school structure may also influence noncognitive outcomes, we considered dependent measures other than achievement. Even with an ideal data structure—large samples of students and schools, achievement measured across the curriculum at more than a single point in time and noncognitive outcomes measured longitudinally—researchers doing school effects studies need statistical software that is able to analyze nested data. Fortunately, several multilevel software programs are now available. Their growing use in teaching and research and the burgeoning body of published studies that have used these methods allow researchers to investigate organizational questions like the ones posed here. We used one of these programs, HLM, in this study.

Technical Details of the Analyses

Constructing school weights. Our research questions suggested the need to use multi-level methods. Because the questions and method both focused on variation among schools, we needed school weights. Thus, we constructed a set of “pseudo-design weights” for the high schools attended by NELS sophomores. 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: aggregations of the inverse of each student’s follow-up weight (the NCES-computed weights for students who also had base-year data) and the probabilities that students in each school had spent their eighth-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 1.0 for our sample of 820 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 HLM from our base-year study (Lee and Smith 1993) under three conditions: unweighted, using the NELS base-year school weights, and using our constructed pseudo-school weights. HLM parameter estimates using our constructed weights were similar to those obtained using the NELS school weights (and more divergent from the unweighted parameter estimates). The results supported our decision to proceed with constructed school weights for the NELS first follow-up.

We used these school weights in all HLM analyses, adjusting them to a mean of 1 for appropriate significance testing in multivariate analyses. Only the descriptive statistics in Table 2 used the NELS student design weights supplied by NCES, also adjusted for appropriate statistical testing. Details about construction of the weights and about our testing procedure are available on request.

Constructing and using gain-score outcomes. Test scores in each subject at both the base-year and first-follow-up were scaled with IRT methods, to adjust for the relative
difficulty of items and other psychometric properties of multiple-choice tests (Ingels et al. 1992: 125). Simple differences between each student's IRT-adjusted test scores at the 10th and 8th grades represent growth measures that form an interval scale for all levels of student ability.

An IRT score is a nonlinear transformation of the number of correct items, adjusted for three item parameters (differential difficulty, discriminating power, and the likelihood of correctly guessing in a multiple-choice format). IRT scaling expands the scale at its extremes. Two consequences of this rescaling are that the estimated gains between 8th- and 10th graders will appear larger for high-ability students in the IRT than a raw-score metric and 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 artifically appear to gain the most (a variation on the "regression to the mean" theme). See Lord (1980) or Hambleton (1989) for further details on IRT.

In science and social studies, the test items at the base-year and first follow-up were identical. In mathematics and reading, follow-up tests were tailored to students' ability. The tailoring of the reading test (which had two forms) was meant to counteract the burden of reading passages for slow readers in the short time frame provided for testing. The mathematics test was tailored (three different forms) so it would be responsive to the diversity of exposure to course work that 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 curricular area. IRT scaling is especially important for the tailored tests because 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).

There is some controversy about using gain (or difference) scores to measure change for the same persons between two points in time. Despite an intuitive appeal, 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. Their reliabilities are often modest primarily because they represent the difference between two scores that are less than perfectly reliable. When analyzing data on two points in time that measure changes in individuals, the researcher typically pursues one of two alternative strategies: (1) using the difference between posttest and pretest scores as an outcome (the gain score) and then perhaps including a measure of initial status as a covariate or (2) using the posttest as an outcome and controlling for the pretest and other potentially confounding factors (the covariance model).

We decided to use gain scores as outcomes, following the advice of Willett (1994). The fact that both pretests (8th-grade scores) and posttests (10th-grade-scores) have been subjected to IRT scaling convinced us that one common difficulty—correlation between gains and initial status—should be minimized. In general, the correlations of initial status with gains were modest (under .2). Except for gains in science, correlations were negative.

From reform-practices to restructuring-practice schools. After we divided the practices in Table 1 into the three groups conceptually, we considered whether to rank the restructuring practices in a "better-worse" order. Since we had neither theory nor evidence to support a ranking, we abandoned that strategy. We then wondered whether certain reforms (the instructional reforms, authority reforms, or human relations reforms) might be adopted by schools in groups based on that logic. We found that they were not. We then investigated the number of reforms that schools might engage in simultaneously. It was clear that schools with restructuring practices 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 cutoff point because it was at this number that the probability of adopting any single restructuring practice alone was exceeded by adopting it in concert with others.

Sensitivity analyses. The results of analyses displayed in Figures 1 and 2 were conducted using the same HLM methods, controls, and outcomes as those displayed in Tables 5 and 6. They represent a series of separate HLMs where the school-level model on each outcome contained a dummy variable classifying schools by the number of restructuring practices they reported (from 1 to over 6). The comparison group in each analysis was composed of the schools that did not engage in a single restructuring reform (but may or may not have had traditional reforms in place). Unweighted distributions of schools were 0 reforms
(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; and over 6 reforms, 5.9 percent. This decidedly nonnormal distribution of reforms argued against constructing a "number of reforms" continuous variable.

The findings for equity (Figure 2) suggest that when multiple reforms (five or more) are simultaneously adopted, equity effects diminish. There is also a slight trend for the effectiveness parameters to diminish with many reforms (Figure 1), especially for engagement and gains in history.

APPENDIX B
DESCRIPTION OF THE CONSTRUCTION OF VARIABLES FOR ALL MEASURES

Dependent Measures

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.

10th-Grade Academic Engagement

Standardized factor-weighted composite ($M = 0$, $SD = 1$) of eight items measuring students’ 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, percentage of total variance in all items explained by the factor = 47.1. NELS student item components are as follows:

- 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 felt challenged in science class

MEASURES OF SCHOOL RESTRUCTURING

Restructuring Measures

Two dummy-coded items were created, using the variables in Table 1 and the technique described in the article. Measures were constructed from items from the NELS first follow-up school file, in which principals reported whether or not schools engaged in a set of 30 school practices. These two measures were:

- No reform practices: Schools that engaged in no reform practices were coded 1, schools classified as those with traditional reforms were coded 0.
- Restructuring practices: Schools that engaged in at least three practices listed as restructured in Table 1 were coded 1, and schools classified as those with traditional reforms were coded 0.

School Size

- F1C2: total enrollment as of October 1989. The 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’s Background

- F1SES: SES composite.
- F1RACE: student’s race (recoded to 0 = White or Asian and 1 = Black, Hispanic, or Native American).
- F1SEX: student’s gender (recoded to 0 = male and 1 = female).

Academic Controls

Analyses included different controls for each curricular area. Controls were constructed as follows:

- For mathematics gain: Z-score of sum of BYTXRIRS, BYTXHIRS, and BYTXSIRS.
- For reading gain: Z-score of sum of BYTXMIRS, BYTXHIRS, and BYTXSIRS.
- For history gain: Z-score of sum of BYTXRIRS, BYTXMIRS, and BYTXSIRS.
- For science gain: Z-score of sum of BYTXRIRS, BYTXHIRS, and BYTXMIRS.

Engagement Control

A factor-weighted standardized ($M = 0$, $SD = 1$) composite of the following base-year student measures of students’ engagement. The composite’s internal consistency (Cronbach’s alpha) = .74.
Effects of High School Restructuring

- 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).
- BYS56E: Students in class see me as a troublemaker (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
- AVES: SES composite, aggregated to the school level.
- F1RACE: student’s race (recoded to 0 = White or Asian and 1 = Black, Hispanic, or Native American), aggregated to the school level and recoded to 1 = 40 percent or more and 0 = less than 40 percent minority.

Academic Emphasis

Summed 10th graders’ reports of course taking in academic courses in mathematics and science:
- F1S22C: How much course work in Algebra I.
- F1S22D: How much course work in geometry.
- F1S22E: How much course work in Algebra II.
- F1S22F: How much course work in trigonometry.
- F1S22G: How much course work in pre-calculus.
- F1S22H: How much course work in calculus.
- F1S23C: How much course work in biology.
- F1S23E: How much course work in chemistry.
- F1S22F: How much course work in physics.

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

Variability in Course Taking

The sum of students’ course taking (academic emphasis) was aggregated to the school level, using the SD operator in SPSSX. The variable was standardized \( (M = 0, SD = 1) \).

Sector

Created from G10CTRL2, the school control measure on the NELS first follow-up restricted school file. Public, Catholic, and NAIS schools were retained; other private schools were dropped. Created two dummy-coded variables:
- Catholic: coded 1 for Catholic and 0 for public and NAIS schools.
- NAIS: coded 1 for NAIS and 0 for public and Catholic schools.

APPENDIX C

HLM MODELS

Within-School Models

A simple form of HLM used in the study 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 investigated the gain in mathematics achievement of student \( i \) in school \( j \), \( y_{ij} \), as a function of students’ background characteristics, \( X_{ij} \)'s (the \( X \)-variables considered are ability, engagement, SES, minority status, and gender), and random error, \( R_{ij} \):

\[
y_{ij} = \beta_{0j} + \beta_{1j}X_{ij1} + \beta_{2j}X_{ij2} + \ldots + \beta_{nj}X_{ijn} + R_{ij}
\]

The \( \beta \) 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 were particularly interested in two \( \beta \) parameters, referred to as the SES-math learning slope:

\[
\beta_{0j} = \text{the average gain in mathematics for students in school } j \text{ and}
\]

\[
\beta_{1j} = \text{the relationship between SES and math gain in school } j.
\]

Although the other \( \beta \) parameters (distributional effects) were also estimated in our HLM analyses, we were not interested in modeling them as functions of structural parameters. As such, the other within-school controls (ability, engagement, minority status, and gender) were fixed in our HLM.
models. That is, we fixed the between-school variability in these other $\beta$ parameters to 0 (they do not vary randomly between schools).

**Between-School Models**

In the second set of equations, we modeled the random-effect $\beta$ parameters, adjusted for students’ characteristics, as functions of school-level characteristics ($W$-variables). We estimated a single between-school model for each outcome, estimating the effects of the three restructuring components on the outcomes ($\beta_i$ and $\beta_j$ for each outcome). For each model, we also adjusted for the potentially confounding effects of school structure and demographics. A typical between-school model is as follows:

$$
\beta_{jk} = \gamma_k + \gamma_{1k}W_{i1} + \gamma_{2k}W_{i2} + \ldots + \gamma_{pk}W_{pi} + U_{jk}
$$

The parameters of interest here are the effects associated with the school-restructuring variables, $W_{ij}$—the $\gamma_{pk}$ coefficients. Since the error terms in this equation are complex, conventional linear model techniques cannot 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. Only effects on the parameter variance are estimated in HLM. This is an important development, since only variability in the structural parameters $\text{Var}(\beta_{jk})$, can be explained by school factors. In general, previous efforts to estimate school effects with OLS regression systematically underestimated school effects for this reason.

**REFERENCES**


Valerie E. Lee, Ed.D., is Associate Professor, School of Education, University of Michigan, Ann Arbor. Her main fields of interest are quantitative research methodology and sociology of education. Her current research is focused on evaluating the effects of school restructuring and reform on students. She is also doing work that identifies organizational features of schools that reduce gender differences in educational outcomes.

Julia B. Smith, Ed.D., is Assistant Professor, Department of Administration and School Organization, Warner School, University of Rochester, Rochester, New York. Her main fields of interest are sociology of education and school organization. She is now studying size and restructuring and teacher professionalism and community.

This research was conducted under the auspices of the Center on Organization and Restructuring of Schools, University of Wisconsin-Madison, with funding from the Office of Educational Research and Improvement, U.S. Department of Education (Grant No. R117Q000005). The authors are grateful to 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; and David Cohen, Fred Newmann, Andrew Porter, and Brian Rowan, for useful suggestions for revising an earlier version. They accept full responsibility for any errors in analysis or interpretation that remain. Address all correspondence to Dr. Valerie E. Lee, School of Education, University of Michigan, 610 East University, Ann Arbor, MI 48109, or by e-mail at Velee@umich.edu.