USING MULTILEVEL METHODS TO INVESTIGATE RESEARCH QUESTIONS THAT INVOLVE NESTED DATA: EXAMPLES FROM EDUCATION*

Valerie E. Lee
Professor of Education at the University of Michigan, Ann Arbor, MI, USA

Resumo
A aprendizagem na escola envolve interações sociais em grupos – por exemplo, em salas de aula, em escolas. Este artigo apresenta uma discussão sobre como métodos multiníveis podem ser usados para investigar questões de pesquisa relacionadas com eficácia escolar e promoção de eqüidade pela escola. Apresenta ainda diversos exemplos envolvendo efeitos da escola sobre eficácia e sobre promoção de eqüidade. Os exemplos envolvem os temas do tipo de escola freqüentada, do tamanho das escolas e da colaboração entre professores.
Palavras-chave: Modelos multiníveis; métodos quantitativos; sociologia da educação.

Resumen
El aprendizaje en la escuela envuelve inter acciones sociales en grupos, por ejemplo, en clases y en escuelas. Este artículo presenta una discusión sobre como los métodos de múltiplos niveles pueden ser usados para investigar cuestiones de pesquisa relacionadas con la eficacia escolar y equidad educacional. Presenta también diversos ejemplos envolviendo los efectos de la escuela sobre la eficacia y promoción de la equidad. Los ejemplos envuelven temas relacionados con tipos de escuelas frecuentadas, el tamaño de las escuelas y la colaboración entre profesores.
Palabras-clave: Modelos de múltiplos niveles; métodos cuantitativos; sociología de la educación.

Abstract
Learning in schools involves social interaction in groups – for instance, in classrooms or in schools. This paper presents a discussion on how multilevel methods can be used for investigating research questions related to efficacy and equity in education. In addition, several examples involving school effects on efficacy and equity are provided. The examples include issues related to the type of school attended, school size and cooperation among teachers.
Keywords: Multilevel models; quantitative methods; sociology of education.

* Visit and seminar sponsored by Pontifical Catholic University of Rio de Janeiro and National Research Council of Brazil.

Part I
Multilevel Analysis in Education

Contexts for Learning

When children are very young, the family is the major context for learning. As they grow, children's educational activities occur with increasing frequency in more formal settings. Within the context of preschool and elementary school, young children experience much of their education in small groups, either grouped into classrooms or into smaller groups within classrooms (in the U.S., young children's classes are often differentiated by skill level in reading or mathematics). Each grouping – indeed, almost any setting where formal or informal learning experiences occur – represents an educational context. Self-selected peer groups may also serve as contexts in which both formal and informal learning occurs.

The nature of formal educational contexts changes as children mature. Younger children spend much of their school time in a single classroom, with a modest number of peers and a single teacher who instructs them in several subjects. As children mature, they still experience their education in classrooms. But for U.S. adolescents in secondary school this is seldom in the same classroom all day. In secondary schools, students are exposed to several teachers and a more numerous and diverse group of classmates compared to elementary school. The logic of specialization drives such changes in the organization of U.S. schooling. The knowledge to be imparted is more complex as students mature, so older children are seen as being best served by teachers with specialized expertise in particular subjects and skills. Most secondary school students visit several classrooms every day, each typically staffed by a different teacher.

Using this logic to study schooling and how it influences individual students' learning, it makes sense to focus on the school itself as an organization unit able to define the educational context for adolescents. Although schools may influence students in many ways, here (and in most of my research) I focus on academic outcomes – achievement or learning. That means that my purpose is to identify school-by-school differences in how much students learn. Although my focus here is on how schools influence learning, it is well known that learning is also influenced by students' own background characteristics (their gender, their race or ethnicity, their social class, or their ability, for example). Though how schools are organized may influence how much individuals in them learn, schools can also influence how that learning is distributed among the students based on their background characteristics – what my colleagues and I have come to call "the social distribution of achievement." Exploring learning and its social distribution as functions of school organizational and structural characteristics combines the common notions of **effectiveness** (high achievement or learning) and **equity** (learning that is relatively unrelated to students' social background).

Studying School Effects

The type of research I am discussing is often called "school effects research". Although this approach is becoming rather well known among sociologists of education, it is much less familiar to educational psychologists. The quantitative methods that are used more and more often to conduct school effects studies are multilevel, in particular Hierarchical Linear Models or HLM. Here I describe two published school effects studies that make use of the HLM methodology. Both of these studies focus on adolescents and high schools. One study focuses on a structural feature of high schools – *their size, or the number of students they enroll* – and the other focuses on a feature of secondary schools' social organization – what we call *"collective responsibility for learning."* Both of these studies were conducted using large and nationally representative samples of about 800 U.S. secondary schools and about 10,000 students who attend them. Both make use of longitudinal data collected by the U.S. Department of Education, "The National Educational Longitudinal Study of 1988 (NELS: 88)." Nationally representative samples of students, their parents, their teachers, and their schools were surveyed when the same students were in 8th grade, 10th grade, and 12th grades. These studies focus only on students who stayed in high school until graduation (in 1992), so that those who dropped out are not included. These data are available without charge from the National Center for Educational Statistics, in the U.S. Department of Education in Washington, D.C.

Both of these studies were co-authored by Julia Smith and myself. They were published in two different U.S. academic journals, as well as being summarized more recently in a new book published by Teachers College Press, entitled Restructuring High Schools for Equity and Excellence: What Works. I have two purposes for describing these studies. First, I want to introduce you to the use of multilevel methods in educational research. Second, I want to provide you with some concrete examples that bring together multilevel research questions posed in educational contexts.

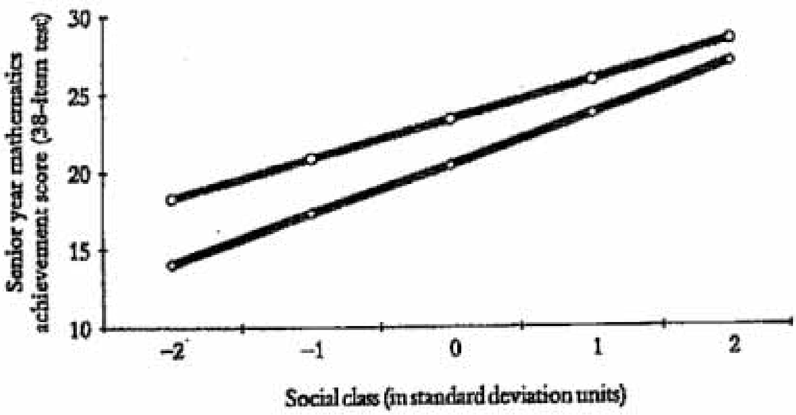
Picturing Effectiveness and Equity

I hope, but I don't assume, that readers share my interest in education. I talk about education because that is what I know best. HLM is very useful, in many cases absolutely essential, in educational research precisely because education occurs in groups – in schools, in classrooms, or even in ability groups within classrooms. The fact is that as a social intervention, education is almost always administered in groups. However, many other social interventions also occur in groups. Before I describe the studies I mentioned above, let me introduce some important ideas by using an example from the 1993 book, Catholic Schools and the Common Good co-authored by Anthony Bryk, Peter Holland, and myself. I think the major theme of the book can be captured in a single graph, which displays some simple numerical information about U.S. Catholic and public high schools presented at the bottom of the graph. The data come from a random sample of U.S. Catholic and public high schools in the early 1980s drawn from a nationally representative and longitudinal dataset called, "High School and Beyond", also collected by the U.S. Department of Education.

This graph is perhaps the most straightforward way to represent a phenomenon I mentioned before: the social distribution of achievement. In this case, the outcome is 12th graders' scores on a standardized test of mathematics achievement (displayed on the vertical [Y] axis). Socioeconomic status, or SES, is shown on the horizontal [X] axis. SES is scaled such that 0 (the mean) represents a U.S. student from a middle-class family. The two lines here represent "regression slopes." As you all probably remember from studying algebra, lines have two descriptive characteristics: a slope and an intercept. The slope, as you might remember, is represented by how many units Y changes for a one-unit change in X. Maybe you also remember the definition of the intercept: the value of Y (achievement) when X (SES) is zero. The top line represents these relationships for students in Catholic schools; the bottom line for those in public schools.

Figure 1

Catholic and public schools: slope and intercept differences for the relationship between social class and mathematics achievement



School Sector	*Correlation SES/Achievement*	*Slope*	*Intercept*	*Sample Size*
Catholic	.24	2.50	23.23	2035
Public	.34	3.23	20.40	1614

These two concepts – intercept and slope – are fundamental ideas in multilevel analysis. What do these two lines show us? This simple display represents the major theme of our research comparing U.S. Catholic and public high schools, in which we concluded that compared to public schools, Catholic schools are both more **effective** (that is, the intercept, or achievement is higher in Catholic schools), and they are also more **equitable** (that is, the slope of the line is flatter in Catholic than public schools. Not only do middle-class students in U.S. Catholic secondary schools learn more (at least in terms of their performance in this standardized test of mathematics achievement), all students' achievement scores are less strongly associated with their social background (in this case, SES). I have no idea whether the same phenomena characterize Catholic and public schools in Brazil, but these phenomena are now well accepted in the U.S. Our major aim in the book was to search for explanations for **why** this is so. We looked for explanations for this combination of phenomena – higher intercept, lower slope – in Catholic compared to public schools. In fact, this is how we defined "good" schools. In terms of HLM, we would try to identify, in the same analysis, the characteristics of schools that were simultaneously related to both the intercept and the slope. We would use both of these as outcomes in a school-effects analysis, using mathematics achievement as the outcome.

A Quick Trip Through HLM

Three steps. For readers who are unfamiliar with multilevel research methods, I provide a very brief introduction. In a typical HLM model used in a school-effects study, there are three steps. In *Step 1*, the researcher partitions the variance in the dependent variable into two parts – the proportion of variance that is among students in the same school (pooled across schools) and the proportion of variance in the dependent variable that lies systematically between schools. It is only between-school variance that may be analyzed as a function of school characteristics.

Determining the between-school proportion of variance, what we call the *intra-class correlation* (or ICC), is the first step in any HLM. In the case of the example shown in the graph, where we used a measure of students' achievement in mathematics as they completed high school as the outcome, that proportion was about 25 percent. Those results are meaningful in themselves (in that they suggest there is considerable stratification in U.S. high schools), and they also provide evidence that an HLM analysis might be fruitful.

In *Step 2*, we investigate how the social and academic characteristics of students are associated with the outcome – what we call the *within-group* or in this case the within-school model. In the example shown in the graph, the only student characteristics we considered was SES. But in a full study we would probably also include statistical controls for students' race/ethnicity, gender, and ability. At Step 2, we can consider whether we should treat the relationship between SES and achievement as a random effect (because it is this relationship we want to explore between schools) and the relationships with the outcome of other characteristics of individual students (race, gender, ability) as fixed effects – as variables we want to include only for the purpose of statistical control. An important advantage of HLM is that we can directly explore slopes-as-outcomes, which allows us to examine determinants of social equity. In HLM, we differentiate between random and fixed effects in terms of our within-group independent variables – and the slopes that we wish to explore as outcomes we designate as random. This raises an issue called "centering". I'll not say anything more about centering here. However, this is something that is important consider in conducting an HLM analysis.

Only at *Step 3* (the final step) may we pose "school effects" questions. The example shown in the graph includes only one school-level variable: a dichotomous measure of whether the school is Catholic or public. In the rest of our analyses (discussed at length in the book and also in a 1989 article by Lee and Bryk published in the journal *Sociology of Education*), we searched for other characteristics of schools that would "explain away" the Catholic/public differences in both the slope and intercept. Our search for such school characteristics focused on the schools' structures, their academic organizations, and their social organizations.

Let me highlight a few issues which readers might want to keep in mind at this point. First, you should recognize the dual characteristics of these lines: the differences between Catholic and public schools in terms of both the intercept (average achievement) and the slope (the relationship between SES and achievement). These two phenomena have characterized much of my research in education. Second, it is useful to remember the three steps in a school-effects HLM: (1) partitioning the variance into within-and between-group parts; (2) estimating a within-group model (with random and fixed effects); and (3) the between-group model, where the researcher searches for characteristics of groups that are related to both the intercept and the slope.

It is possible to use HLM for more complex research questions. For example, we could explore a three-level HLM educational analysis, with students nested in classrooms and classrooms nested in schools. We might also explore another type of three-level HLM, where Level 1 would be a repeated measures change analysis (e.g., multiple test scores nested in students), Level 2 would explore how individual characteristics of students (e.g., race, class, gender, ability) would be related to achievement growth, and Level 3 might explore school effects. We might also use HLM for analyzing categorical or dichotomous outcomes – whether students drop out of high school, whether they go to university after graduating, or whether they follow a college-preparatory, a general, or a vocational curriculum in high school. I won't provide any details on these other applications of HLM here, but readers should be aware that HLM is quite flexible in the type of multilevel questions that may be explored. Of course, there are many applications for this methodology in fields other than education.

Part II
Examples of School Effects Studies

Study I: School Size and Learning

The research questions. A defining characteristic of any school is the number of students which it is responsible for educating. Especially in secondary schools, enrollment size has important implications for how the school is organized, the curriculum that is offered, and how school members interact. In the U.S., high-school size is currently a very "hot" policy issue. In this study, my colleague Julia Smith and I examined how this important school structural feature influences learning.

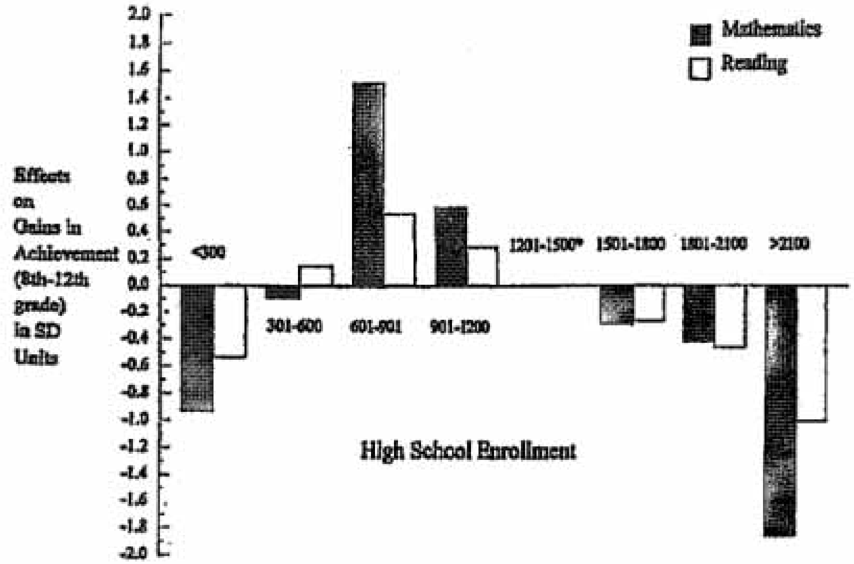
There are two strands of research on this topic. One focuses on the economic efficiency in bringing larger numbers of students into a single school. The second strand focuses on the sociological dimensions of schooling. The first research strand favors larger schools as a factor in economic efficiency. However, contemporary policy-oriented research has generally concluded that U.S. high schools (particularly public schools) are generally too large. However, there is also a concern that schools could become too small to serve all their students well, particularly in terms of offering a reasonable curriculum.

The aim of our study was to identify the optimal size for a high school. As we defined it, **optimal** was indicated by how much students learn over their four years in high school in mathematics and reading comprehension in English. We were also interested in identifying whether the same optimal size was applicable for (a) students of different social class backgrounds, and (b) schools of different social compositions (defined in economic and racial terms).

Data. As mentioned, the data were drawn from NELS: 88. All students were in 8th grade in 1988 and graduated from high school in 1992. They all remained in the same high schools between 10th and 12th grade. We estimated school size effects on learning in mathematics and reading, measured as the change in achievement in these two subjects between when students entered high school (the end of 8th grade) and when they graduated (the end of 12th grade). Our major school-level independent variable was school size, which we divided into eight categories (under 300, 300-600, 600-900, 900-1200, 1200-1500, 1500-1800, 1800-2100, and over 2100). We used categories because we hypothesized that the effects of school size on learning might be non-linear.

Figure 2

Effects of high school size on achievement gains in mathematics and reading. Note: 1,201 to 1,500 students were used as the comparison group. By definition, effect size were zero in that category



Our within-school (Level 1) model took into account students' gender, SES, minority status (African-American and Hispanic students were considered minorities), and their scores on tests in both subjects in those subjects as they entered high school. We designated the SES slope as a random variable, so that the relationship between SES and gains in achievement could be examined as an outcome (the slope), as well as average achievement gain (the intercept). Our Level 2 model includes, besides the set of dummy-coded variables for school size, other controls for the average SES of the school, whether or not the school enrolled over 40 percent minority students, and whether the school was a public, Catholic, or elite private school. These within-school and between-school variables were meant to control for other features of students and schools that might serve as possible alternative explanations for the effects of school size.

Descriptively, test-score gains, as well as other characteristics of students, were related to school size. In general, schools that were in the mid-range for size (enrolling over 600 but less than 1,200 students) enrolled more able students, fewer minorities, and higher-SES students. This was reflected in the school demographic variables. Particularly striking was that in general, private schools were smaller than public schools. In the U.S., more schools were in the size category of 1,200-1,500 students, so we used this group as the uncoded comparison.

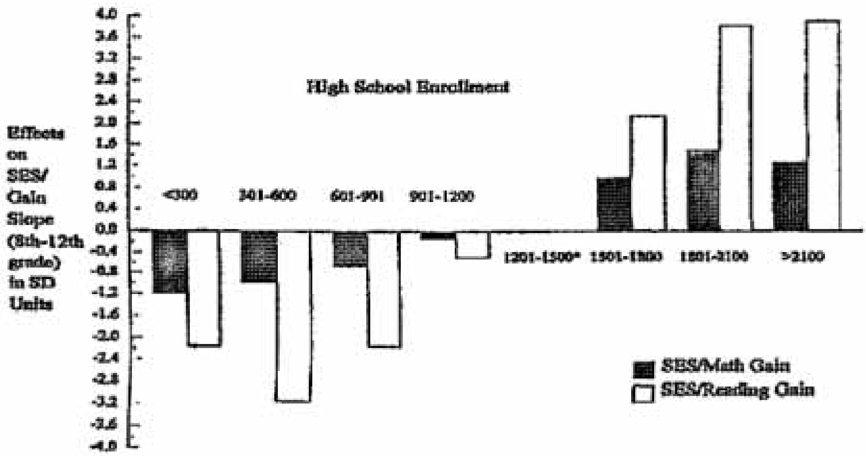
Size effects on learning in secondary schools. We chose to present these results in graphs, rather than statistical results in tables, because we were anxious that the analyses and findings be understood and meaningful to people who work in schools (practitioners) and to those who control schools (policymakers). Thus turned out to be a very good idea. We have sent out hundreds and hundreds of copies of this paper, and I have received e-mails and telephone calls from many, many policy makers. Thus, I want to highlight another message here, beyond the findings themselves and a substantive application of HLM. That is, I recommend that researchers try to present complex results in as non-technical a fashion as possible, so that a wide audience can make sense of them. In the journal article describing this work, we included the tables with the complete numerical results in an Appendix. The results in these graphs are presented in effect-size units. These are simply standard deviations, on achievement gains in the two subjects over the four years of high school. This metric is increasingly common in social science.

Findings about size and learning. Our Level-2 analyses estimated an HLM model that contained a series gamma coefficients for each of the school size dummy variables. All of these effects have been adjusted for the demographic and academic backgrounds of students, as well as the compositions and sectors of the high schools. The findings in Figure 3 focus on the intercepts, where size effects on mathematics learning are in dark bars, for reading learning in light bars. Remember that all comparisons on this graph are with schools in the 1,200-1,500 category (whose effects are, by definition, zero). As the heights of the bars represent learning over the four years of high school, higher bars represent a more effective school size. Three findings from Figure 3 are noteworthy:

- The optimal size category, for both mathematics and reading, are schools that enroll 600-900 students. That is, students learn more in both subjects in schools of that size;
- Size effects average learning more strongly in mathematics than reading; and
- Learning is particularly low in very large and very small schools. That is, school size does not have a linear effect on learning – students learn least in schools that are very large or very small.

Figure 3

Effects of high school size on the relation between SES and achievement gains in mathematics and reading. Note: 1,201 to 1,500 students were used as the comparison group.
By definition, effects sizes were zero in that category

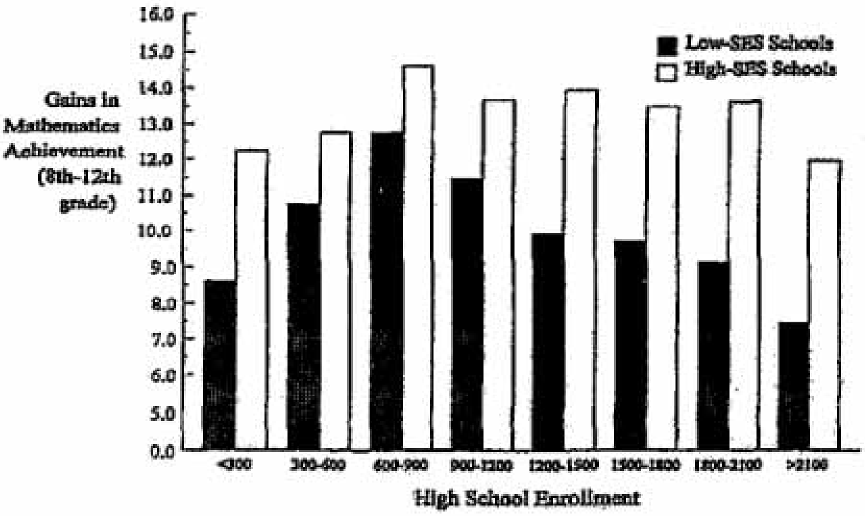


Findings about size and the equitable distribution of learning. Figure 4 displays the results for the analyses in which the SES/achievement gain slopes were outcomes. As before, results are in effect-size units, again compared to the mid-sized schools. The black bars again represent the relationship between SES and mathematics learning, the white bars the relationship between SES and reading learning. The "ideal" here is exactly opposite from the "effectiveness" analyses shown in Figure 3. Here, we would like the relationship between SES and learning to be small. As before, size effects are compared to the 1,200-1,500 category. Again, I highlight three findings:

- Size effects on equity are larger in reading learning than mathematics;
- In this case, the most equitable schools are smaller – the most equitable category is the 300-600 range, where compared to average size schools, the relationship between SES and learning is much smaller; and
- The most inequitable schools, in terms of the relationship between SES and learning, are the largest schools.

Figure 4

Average gains in mathematics achievement by high school size in low-SES and high SES high schools



Findings about school size and school social composition. These analyses are similar to the HLMs in which we estimated the results shown in Figures 1 and 2, except that here we added to our analytic models a set of interaction terms – between the school size categories and school average SES. These analyses show all the school size categories. The dependent variable was, again, gains in mathematics achievement. Low-SES schools were those that were at least one SD below the mean, and high-SES schools were those that were one or more SDs above the mean for average SES. We also conducted the same analyses for reading achievement gains, and also compared our results for schools with high and low minority enrollments. Because the results were quite similar, I show only the findings for math learning. Again, we three conclusions are drawn from these results:

- An important but perhaps not surprising finding is that learning is greater in high-SES schools (white bars) than low-SES schools;
- More surprising is that the optimal school size for both high- and low-SES schools is identical: 600-900 students; and
- School size appears to make more difference in low-SES than high-SES schools (i.e., the change in effect sizes due to school size is more striking). This is perhaps the most noteworthy finding here.

Implications of these analyses. Although my purpose in showing you results from this study is to demonstrate a use of HLM. I want to emphasize how important HLM was in conducting this kind of analysis – in fact, that is my major point. However, there are also some substantive conclusions we have drawn from these analyses that are hard to ignore. First, U.S. high schools should be smaller than they usually are, in that students learn more in smaller high schools. However, there seems to be a balance point – neither too large nor too small. We believe high schools need to be large enough to offer a credible curriculum but small enough so that school members can know one another well – so that schools can function as communities. Although the dual aims of effectiveness and equity do not converge on exactly the same school size, we conclude that and ideal size for a U.S. high school is 600-900 students. We were struck by the consistency for results for different subjects and in schools that enrolled different social compositions. Although school size seems important for learning for all students, the special importance of school size for disadvantaged students is noteworthy. The schools such students typically attend are typically the largest. A similar study in Brazilian high schools would be very interesting.

Why would school size influence learning? Is this really a direct effect, as we estimated it in this study? I suggest that school size is actually a proxy measure for several different organizational characteristics of high schools, both academic and social. I believe that if we actually were to include good measures of the many organizational features of high schools that are influenced by how many students attend, it is likely that the school size effects would be "explained away." Nevertheless, from these analyses it would be difficult to conclude that school size didn't matter much.

Study 2: School Social Organization and Student Learning

The research questions. Let us shift gears a bit. The main work of schools, teaching and learning, revolves around teachers. Many current reforms in the U.S., reforms that are often grouped under the title of "school restructuring," focus on changing the procedures, principles, rules, roles, and relationships within schools. Suggestions of the last two types of reform (roles and relationships) target teachers. This second study focuses on teachers. Whether the locale is inside or outside the classroom, whether the focus is on the academic or social dimensions of schooling, the work done by teachers is pivotal in the education of students. Moreover, the work teachers do (instruction) is influenced by the conditions in which they perform their tasks.

A common assumption of reforms targeting teachers' work is that, either implicitly or explicitly, such reforms will improve student learning. Study 2 tests this assumption by exploring how features of high school teachers' work lives influence how much their students actually learn. My co-author (again it is Julia Smith) and I captured the notion of teachers' professional lives in three ways:

(1) How teachers define their work and their students (collective responsibility);
(2) How they interact with their students, colleagues, and superiors (collaboration); and
(3) The degree to which teachers feel control over their work (control).

We used school averages of these three notions as measures of school social organization. That is, we aggregated composite variables measuring these ideas, created from teachers' responses to several survey items, to the school level. My main focus here is on the first issue, which captures school norms reflecting teachers' attitudes about their students. We called this **collective responsibility for learning**. We hypothesized that the attitudes that teachers have about their students may influence the students in two ways. Most obviously, when the teachers in the school believe that their own efforts are very important in the learning process, student learning might increase. Moreover, when responsibility is assumed for **all** students' learning, regardless of the students' academic qualifications or social backgrounds, learning may be more equitably distributed among students. This study explores both of these outcomes.

Data. The data used for this study are almost exactly the same as for the study of school size, again drawn from NELS: 88. However, in this study the data were drawn from the first two waves – students' achievement gains in the first two years of high school. Thus, the sample here was a bit larger than for the size study, in that some of the students in this study would have dropped out of high school by 12th grade. Here, our sample was 11,692 students in 820 schools. Most schools were public (650), 68 were Catholic, and 47 were elite private schools. Data on teachers' work lives were drawn from surveys of students' teachers at the 10th grade. Two of each student's teachers were surveyed, in four different subjects (9,904 10th grade teachers of English, math, social studies, and science).

Measures. The measures of teachers' work lives are factor-score composites constructed from individual teacher reports and then aggregated to the school level. The measure of collective responsibility for learning includes four related ideas: (1) teachers' internalizing responsibility for the learning of their students rather than attributing learning difficulties to low ability or deficient home conditions; (2) a belief that teachers can be successful with all students; (c) willingness to alter teaching methods in response to students' difficulties and successes; and (d) self-efficacy in teaching. Two other work life measures included in the analysis capture the degree of (a) cooperation among teachers and (b) control teachers feel they have over teaching and school policy. In comparison to Study 1, we investigated these effects on learning in four subjects (English, mathematics, social studies, and science). Again we use achievement gains in these subjects as outcomes, similarly to Study 1.

Analysis strategy and presentation. This multilevel research question also requires multilevel analysis methods. We used HLM to investigate not only whether collective responsibility is associated with average student learning in the four subjects but also whether this feature of the social organization of schools is associated with a socially equitable distribution of learning. As before, I present these results in graphic form. The effects shown in these graphs are drawn from numerical results from our four HLM analyses (one for learning in each subject). The results are presented as point-score estimates of two-year gains on standardized tests in the four subjects. In computing the intraclass correlation of our outcomes, we found that the proportion of the total variance in these gain-score measures that is between schools ranged from 15 to 20 percent, lower for reading and higher for science. Of course, our search for effects on student learning of collective responsibility involves only the proportion of the total variance that is between schools.

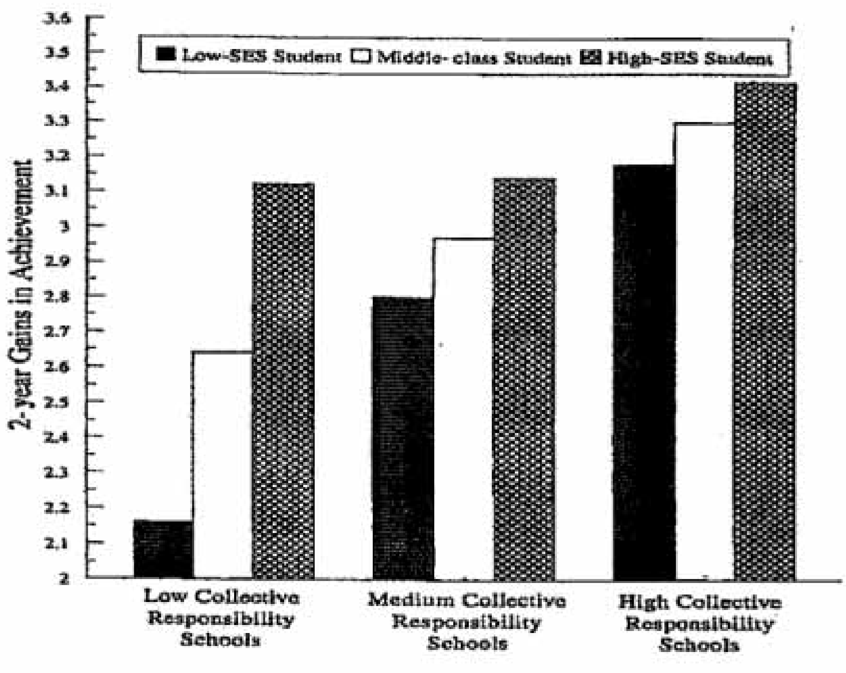
We included in our Level-1 (within-school) model statistical controls for students' social and academic background: students' engagement with school, their gender, minority status, and family SES. We also controlled for students' ability as they entered high school and for the curriculum track (academic, general, vocational) they followed in their high-school coursework. Similar to school size study, we considered student SES as a random variable, so that both the intercepts (average learning in the four subjects) and the social distribution parameters (the SES/learning slope in the four subjects) became outcomes in our between-school Level-2 analyses.

The major focus of our Level-2 models was on the effects of collective responsibility. Besides the average of this measure, we also included a measure of how these attitudes varied among a school's faculty (capturing the degree of agreement or disagreement among teachers). Measures of teacher cooperation and control were in the models, as well. We also took into account school composition, both demographic (average SES, percent of minority students) and academic (average achievement as students entered high school and the proportion of students in the academic curriculum track). As before, we also controlled for school sector (i.e., whether the school governance is public, Catholic, or elite private).

Effects of collective responsibility on learning and its equitable distribution. In our full 2-level HLM models, we found that of the several measures of teachers' work lives that we considered, only one – collective responsibility for learning – was consistently, positively, and significantly related to achievement gains. Thus, I concentrate on this measure here. As with Study 1, the results are again in graphical form. However, the graphs presenting these results are different from those in Study 1, as they combine our findings about how collective responsibility for learning influences both average learning (the intercept) and the relationship between SES and learning (the slope).

Figure 5.1

*Achievement gains in **mathematics** for students o varying levels os SES in schools with different levels of collective responsibility*



The heights of the bars in Figure 5.1 were all calculated from a full two-level HLM, including the many controls about students and schools. The outcome (on the Y axis) is measured in point-score gains over two years. Figure 5.1 describes findings for gains in mathematics achievement. Schools are grouped by whether than exhibit low, medium, or high collective responsibility. These are simply point estimates for schools at one SD below the mean on this variable (low), schools at the mean (medium), and those that are one SD above the mean on collective responsibility (high). We have also divided the students in the entire sample into three groups – low-SES students are those whose family SES is one SD below the mean; middle-class students are those whose SES is 0, and high-SES students are one SD above the mean. Figure 5.1 tells two major stories.

Story 1 focuses on **effectiveness**. Quite simply, learning in mathematics over the first two years of high school is related to the level of collective responsibility among the schools' teachers, with students learning more in schools characterized by more collective responsibility. The heights of the white bars, which display the results for middle-class students' learning, represent collective responsibility effects on average learning (i.e., the intercept). The heights of the three white bars demonstrate that in schools with higher levels of collective responsibility mathematics learning is greater.

Story 2 is drawn from the results shown in Figure 5.1 focuses on **equity**. Here we would compare the heights of the bars, within each level of collective responsibility, for students from different SES backgrounds. Notice that for high-SES students (the checked bars), a school's level of collective responsibility isn't too important (that is, the three checked bars are almost the same height). However, for low-SES students (the black bars), the relationship between collective responsibility and learning in mathematics is very strong. Low-SES students in schools with high levels of collective responsibility learn quite a lot, whereas low-SES students in schools with low levels of collective responsibility learn very little in their first two years of high school. That is, collective responsibility equalizes the social distribution of mathematics learning.

Figure 5.2

*Achievement gains in **reading** for students o varying levels os SES in schools with different levels of collective responsibility*

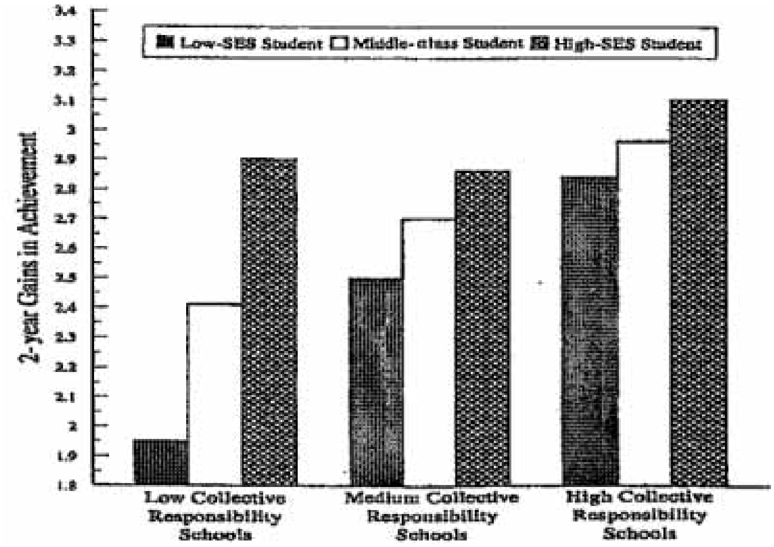


Figure 5.3

*Achievement gains in **history** for students o varying levels os SES in schools with different levels of collective responsibility*

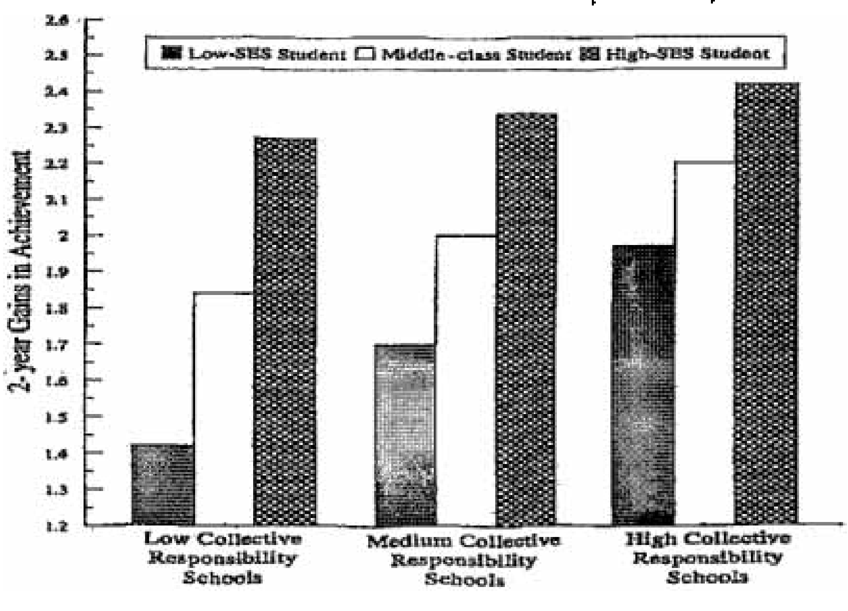
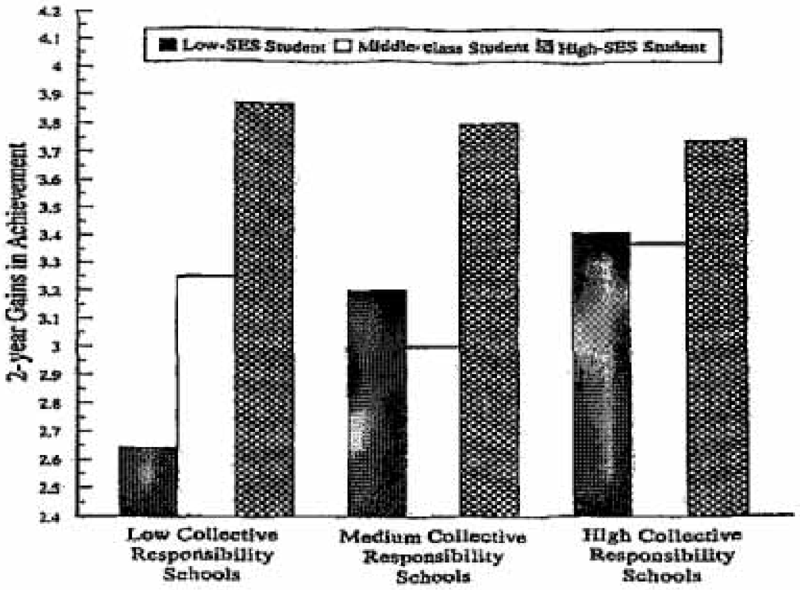


Figure 5.4

Achievement gains in ***science*** *for students o varying levels os SES in schools with different levels of collective responsibility*



Does this same pattern hold for learning in other subjects? The answer is that it does. Figures 5.2 to 5.4 represent the same analyses for achievement gains in reading (5.2), history (5.3), and science (5.4). As you can see, the patterns here are quite similar to those for achievement gains in mathematics. That is, in general, average achievement gains are higher in schools with higher levels of collective responsibility for learning (compare the white bars for students of average SES). However, school collective responsibility is again especially important for lower- SES students (the black bars). If you were draw a line between the bars in each SES group, the slopes would be the steepest for low-SES students. For higher-SES students, the collective responsibility level of the school is again not very important. As in mathematics, if we focus on SES differences in achievement gains for schools with high collective responsibility, we see that low-SES students do particularly well in such schools. However, in schools with low levels of collective responsibility, low-SES students learn very little.

Policy implications for Study 2. Technically, the results from this kind of analysis are quite satisfying. HLM worked well here, and we were able to demonstrate effects for both effectiveness and equity. In another sense, however, the strength and consistency of the findings of Study 2 are quite troubling, in that learning is quite strongly related to the attitudes held by a school's teachers about the students they teach. In schools that enroll many students who are socially disadvantaged, teachers beliefs about the limitations of their students' ability to learn and of teachers' ability to teach them effectively are quite prevalent (i.e., collective responsibility is quite low). If you believe, as I do, that secondary school teachers **should** assume responsibility for their students' academic progress – all their students – then the results demonstrate a phenomenon in schools that we would rather not find. Our results do demonstrate how important teachers' attitudes are to students' learning; however, how to change these attitudes is less obvious. More optimistically, the results of Study 2 offer support for the importance of teachers' beliefs for their students' learning. On a broader level, we may conclude that schools in which teachers interact as a professional community have benefits for their students, in terms of both learning and social equity.

Part III
Conclusions

Studying School Effects with Multilevel Methods

My aim in describing these two studies has been to provide examples of research that investigates how specific features of the school context can influence how much students learn. The particular characteristics of secondary schools explored in these studies included one feature of school structure – the number of students enrolled – and one feature of school social organization – the attitudes of teachers about the students they teach. Both studies were conducted using a large, relatively recent, longitudinal, and nationally representative sample of U.S. high schools and random samples of the students who attend them. Although these studies do not employ an experimental design (that is, students were not randomly assigned to schools of different sizes or different levels of collective responsibility), it is hard to imagine how any educational studies could actually use a randomized design – random assignment of students to schools would be socially quite unacceptable in the U.S.; we have a strong commitment to students attending local schools close to where they live.

A major strength of these two studies is their longitudinal designs, which allow us to investigate **learning** (which measures change over time in academic status) rather than **achievement** (which is a status measure). The ability to demonstrate how schools influence the students who attend them is strengthened by being able to take into account the status of these students at the point where they enter the schools (especially their academic status). Both studies focus on learning outcomes, measured in terms of both school averages and the social distribution of learning. I would argue that it is very appropriate to investigate how characteristics of schools influence both effectiveness and equity.

Even though the quality and structure of the data to study school effects are very important, the major message I would like impart is how crucial it is to use appropriate and multilevel methods to study multilevel questions. Such questions are quite common in education, because the educational experience for both students and teachers occurs in groups – classrooms and schools (of course, there are higher levels of educational groupings – school districts, states, and even countries).

Recent writings about educational evaluations have frequently argued that many educational studies arrive at incorrect conclusions because the wrong methods are used. This is particularly serious when social policies are made as a result of the studies. In general, contextual effects are systematically underestimated in single-level analyses. However, many educational policies are actually implemented at the group level – often at the school level. If we want to know whether particular educational policies and implementations attain the goals set for them, it is crucial to analyze data in the right way. Of course, we need to have good data. Equally important, we need more and more researchers to become familiar with, and skilled in using, multilevel research methods. If you draw no other message from this seminar, I hope you recognize this.