

# Chapter 13

## MULTILEVEL MODELS FOR SCHOOL EFFECTIVENESS RESEARCH

RUSSELL W. RUMBERGER

GREGORY J. PALARDY

One of the major topics for social science research is the study of school effectiveness. Beginning with the first large-scale study of school effectiveness in 1966, known as the “Coleman report” (Coleman et al., 1966), literally hundreds of empirical studies have been conducted that have addressed two fundamental questions:

1. Do schools have measurable impacts on student achievement?
2. If so, what are the sources of those impacts?

Studies designed to answer these questions have employed different sources of data, different variables, and different analytic techniques. Both the results of those studies and the methods used to conduct them have been subject to considerable academic debate.

In general, there has been widespread agreement on the first question. Most researchers have concluded that schools indeed influence student achievement. Murnane’s (1981) early review captured this consensus well:

There are significant differences in the amount of learning taking place in different schools and in different

classrooms within the same school, even among inner city schools, and even after taking into account the skills and backgrounds that children bring to school. (p. 20)

Another reviewer concluded more succinctly, “Teachers and schools differ dramatically in their effectiveness” (Hanushek, 1986, p. 1159). Despite this general level of agreement on the overall impact of schools, how much impact schools and teachers have is less clear, an issue we address later in this chapter.

It is the second question, however, that has generated the biggest debate. Coleman et al. began this debate with the publication of their report in 1966 by concluding that schools had relatively little impact on student achievement compared to the socioeconomic background of the students who attend them. Moreover, Coleman (1990) found that “the social composition of the student body is more highly related to achievement, independent of the student’s own social background, than is any school factor” (p. 119). The publication of the Coleman report also marked the beginning of the methodological debate on how to estimate school effectiveness, a debate that has continued to this day. The Coleman study was criticized on a number of methodological grounds, including the lack of

AUTHORS’ NOTE: We would like to acknowledge the helpful comments of David Kaplan and especially Michael Selzter.

controls for prior background and the regression techniques used to assess school effects (Mosteller & Moynihan, 1972).

Since the publication of the original Coleman report, there have been a number of other controversies on sources of school effectiveness and the methodological approaches to assess them. One debate has focused on whether school resources make a difference. In a major review of 187 studies that examined the effects of instructional expenditures on student achievement, Hanushek (1989) concludes, "There is no strong or systematic relationship between school expenditures and student performance" (p. 47). As noted earlier, Hanushek does acknowledge widespread differences in student achievement among schools but does not attribute these differences to the factors commonly associated with school expenditures—teacher experience, teacher education, and class size. A recent reanalysis of the same studies used by Hanushek, however, reaches a different conclusion: "Reanalysis with more powerful analytic methods suggests strong support for at least some positive effects of resource inputs and little support for the existence of negative effects" (Hedges, Laine, & Greenwald, 1994, p. 13).

Another debate has focused on the effectiveness of public versus private schools. Several empirical studies found that average achievement levels are higher in private schools, in general, and Catholic schools, in particular, than in public schools, even after accounting for differences in student characteristics and resources (Bryk, Lee, & Holland, 1993; Chubb & Moe, 1990; Coleman & Hoffer, 1987; Coleman, Hoffer, & Kilgore, 1982). Yet although some (Chubb & Moe, 1990) argue that all private schools are better than public and thus argue for private school choice as a means to improve education, other researchers have argued that Catholic schools, but not other private schools, are both more effective and more equitable than public schools (Bryk et al., 1993). Still other researchers find little or no Catholic school advantage (Alexander & Pallas, 1985; Gamoran, 1996; Willms, 1985). Moreover, it has been suggested that controlling for differences in demographic characteristics may still not adequately control for fundamental and important differences among students in the two sectors (Witte, 1992, p. 389).

Much of the debate about school effectiveness has centered on methodological issues. These issues concern such topics as data, variables, and statistical models used to estimate school effectiveness. Since the research and debate on school effectiveness began almost 50 years ago, new, more comprehensive

sources of data and new, more sophisticated statistical models have been developed that have improved school effectiveness studies. In particular, the development of multilevel models and the computer software to estimate them have given researchers more and better approaches for investigating school effectiveness. This chapter reviews some of the major methodological issues surrounding school effectiveness research, with a particular emphasis on how multilevel models can be used to investigate a number of substantive issues concerning school effectiveness.<sup>1</sup>

We will illustrate these issues by conducting analyses of a large-scale national longitudinal study that has been the source of a lot recent research on school effectiveness, the National Education Longitudinal Study of 1988 (NELS). NELS is a national longitudinal study of a representative sample of 25,000 eighth graders begun in 1988. Base year data were collected from questionnaires administered to students, their parents and teachers, and the principal of their school. Follow-up data were collected in 1990, 1992, 1994, and, most recently, in 2000 on a subset of the original sample (Carroll, 1996). Students were also given a series of achievement tests in English, math, science, and history/social studies in the spring of 1988, 1990, and 1992, when most respondents were enrolled in Grades 8, 10, and 12, respectively. In this chapter, we will use a subsample of the NELS data for 14,199 students with valid questionnaires from the 1988, 1990, and 1992 survey years who attended 912 high schools in 1990.<sup>2</sup> The appendix provides descriptive information on the variables in the data set that were used to test the models in this chapter.

We begin this chapter by presenting a conceptual model of schooling that can be used to frame studies of school effectiveness. Next we discuss several issues regarding the selection of data and variables used to test multilevel models. Then we review various types and uses of multilevel models for estimating school effectiveness. Finally, we review techniques for identifying effective schools. For each topic, we will explain some of the important decisions that researchers must make in undertaking school effectiveness studies and how those decisions can influence the outcomes and conclusions of the study.

1. Many of the concepts and techniques we discuss can be used to study the effectiveness of other types of organizations, such as hospitals.

2. To generate accurate school-level composition measures, we restricted the sample to respondents who had a valid school ID in 1990, had valid test scores in 1988 and 1990, and attended a high school with at least five students.

### 13.1. A CONCEPTUAL MODEL OF SCHOOLING

To undertake quantitative research on school effectiveness, we should have a conceptual model of the schooling process. A conceptual model can be used to guide the initial design of the study, such as the selection of participants and the collection of data, as well as the selection of variables and the construction of statistical models. Although several different conceptual frameworks have been developed and used in school effectiveness research over the years (e.g., Rumberger & Thomas, 2000; Shavelson, McDonnell, Oakes, & Carey, 1987; Willms, 1992), all have portrayed schooling as a multilevel or nested phenomenon in which the activities at one level are influenced by those at a higher level (Barr & Dreeben, 1983; Willms, 1992). For example, student learning is influenced by experiences and activities of individual students, such as the amount and nature of the homework that they do. But student learning is also influenced by the amount and nature of the instruction that they receive within their teachers' classrooms, as well as by the qualities of the schools they attend, such as school climate and the nature of the courses that are provided. Ignoring or incorrectly specifying these multilevel influences can yield misleading conclusions about their effects on student learning (e.g., Summers & Wolfe, 1977).

In addition to its multilevel nature, the process of schooling can be divided into distinct components. One framework is based on the sociological view of schooling (Tagiuri, 1968; Willms, 1992), which identifies four major dimensions of schooling: ecology (physical and material resources), milieu (characteristics of students and staff), social system (patterns and rules of operating and interacting), and culture (norms, beliefs, values, and attitudes). Another framework is based on an economic model of schooling (e.g., Hanushek, 1986; Levin, 1994), which identifies three major components of schooling: the inputs of schooling—students, teachers, and other resources; the educational process itself, which describes how those inputs or resources are actually used in the educational process; and the outputs of schooling—student learning and achievement.<sup>3</sup>

An example of a conceptual framework based on the economic model is illustrated in Figure 13.1. The framework shows the educational process operating at the three levels of schooling—schools, classrooms,

and students. It also identifies two major types of factors that influence the outcomes of schooling: (a) inputs to schools, which consist of structure (size, location), student characteristics, and resources (teachers and physical resources), and (b) school and classroom processes and practices. School inputs are largely “given” to a school and therefore are not alterable by the school itself (Hanushek, 1989). The second set of factors refers to practices and policies that the school does have control over and thus are of particular interest to school practitioners and policy-makers in developing indicators of school effectiveness (Shavelson et al., 1987).

#### 13.1.1. Dependent Variables

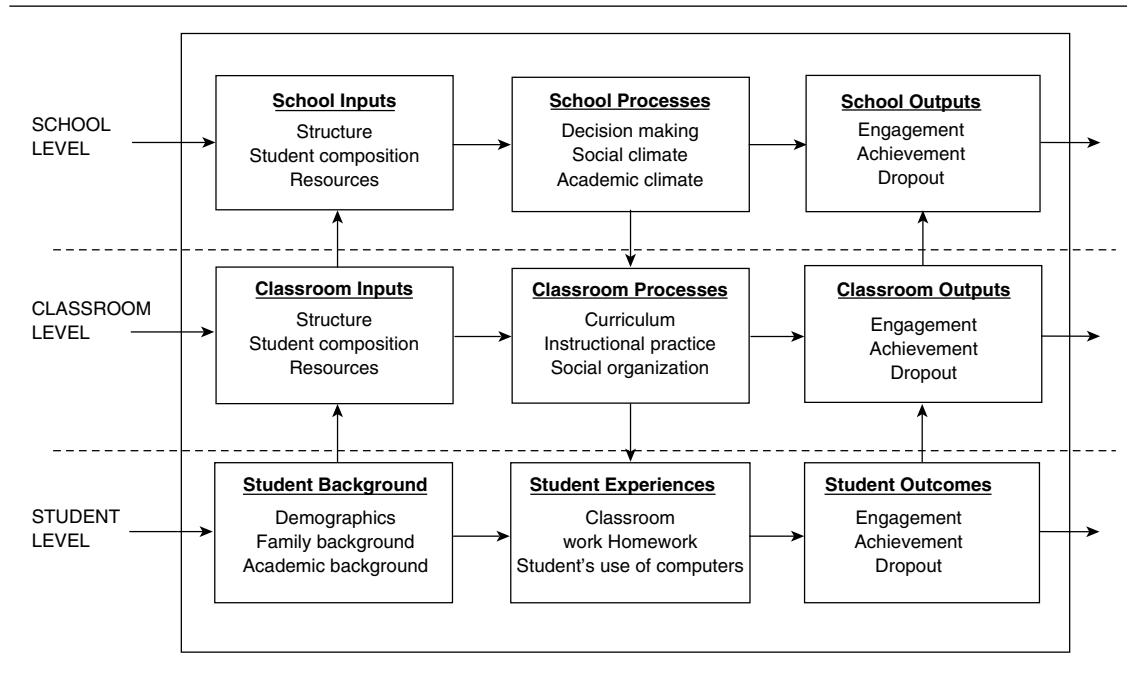
The framework suggests that school effectiveness research can focus on a number of different educational outcomes. The most common measure of school effectiveness is academic achievement, as reflected in student test scores, which is considered one of the most important outcomes of schooling. Although student academic achievement is affected by the background characteristics of students, research has clearly demonstrated that achievement outcomes are also affected by the characteristics of schools that students attend (Coleman et al., 1982; Gamoran, 1996; Lee & Bryk, 1989; Lee & Smith, 1993, 1995; Lee, Smith, & Croninger, 1997; Witte & Walsh, 1990).

Other student outcomes have also been examined in studies of school effectiveness. One of these is school dropout, which studies have shown is also affected by the characteristics of schools that students attend (Bryk et al., 1993; Bryk & Thum, 1989; Coleman & Hoffer, 1987; McNeal, 1997; Rumberger, 1995; Rumberger & Thomas, 2000). Other studies have examined the impact of school characteristics on absenteeism (Bryk & Thum, 1989), engagement (Johnson, Crosnoe, & Elder, 2001), and social behavior (Lee & Smith, 1993). One reason for examining alternative student outcomes is that schools and school characteristics that are effective in improving student performance in one outcome may not be effective in improving student performance in another outcome (Rumberger & Palardy, 2003b).

#### 13.1.2. Independent Variables

The conceptual framework suggests that several types of variables are valuable in constructing statistical models of school effectiveness. We provide a very brief review of some of these variables.

3. In his landmark study of school effectiveness, sociologist James Coleman employed an input-output model of the schooling process (see Coleman, 1990).

**Figure 13.1** A Multilevel Conceptual Framework for Analyzing School Effectiveness

### 13.1.2.1. Student Characteristics

Research has demonstrated that a wide variety of individual student characteristics are related to student outcomes. These include demographic characteristics, such as ethnicity and gender; family characteristics, such as socioeconomic status and family structure; and academic background, such as prior achievement and retention. These characteristics have been shown to relate to such student outcomes as engagement, achievement (test scores), and dropout (Bryk & Thum, 1989; Chubb & Moe, 1990; Lee & Burkam, 2003; Lee & Smith, 1999; McNeal, 1997; Rumberger, 1995; Rumberger & Palardy, 2003b; Rumberger & Thomas, 2000).

Student characteristics influence student achievement not only at an individual level but also at an aggregate or social level. That is, the social composition of students in a school (sometimes referred to as *contextual effects*) can influence student achievement apart from the effects of student characteristics at an individual level (Coleman et al., 1966; Gamoran, 1992). Studies have found that the social composition of schools predicts school engagement, achievement, and dropout rates, even after controlling for the effects of individual background characteristics of students (Bryk & Thum, 1989; Chubb & Moe, 1990; Jencks & Mayer, 1990; Lee & Smith, 1999; McNeal, 1997; Rumberger, 1995; Rumberger & Thomas, 2000).

### 13.1.2.2. School Resources

School resources consist of both fiscal resources and the material resources that they can buy. As mentioned earlier, there is considerable debate in the research community about the extent to which school resources contribute to school effectiveness. But there is much less debate that material resources matter, particularly the number and quality of teachers. Yet the exact nature of teacher characteristics that contribute to school effectiveness, such as credentials and experience, is less clear (Goldhaber & Brewer, 1997). Beyond the quality of teachers, there is at least some evidence that the quantity of teachers—as measured by the pupil/teacher ratio—has a positive and significant effect on some student outcomes (McNeal, 1997; Rumberger & Palardy, 2003b; Rumberger & Thomas, 2000).

### 13.1.2.3. Structural Characteristics of Schools

Structural characteristics, such as school location (urban, suburban, rural), size, and type of control (public, private), also contribute to school performance. Although widespread achievement differences have been observed among schools based on structural characteristics, what remains unclear is whether structural characteristics themselves account for these differences or whether they are related to differences

in student characteristics and school resources often associated with the structural features of schools. As we pointed out earlier, this issue has been most widely debated with respect to one structural feature: the difference between public and private schools. More recently, there has been considerable interest in another structural feature of schools: school size (Lee & Smith, 1997).

#### 13.1.2.4. School Processes

Despite all the attention and controversy surrounding the previous factors associated with school effectiveness, it is the area of school processes that many people believe holds the most promise for understanding and improving school performance. Although most *individual* schools, or at least most public schools, have little control over student characteristics, resources, and their structural features, they can and do have a fair amount of control over how they are organized and managed, the teaching practices they use, and the climate they create for student learning—features referred to as *school processes*. Some researchers have also referred to them as “Type B effects” because, when statistical adjustments are made for the effects of other factors, they provide a better and more appropriate basis for comparing the performance of schools (Raudenbush & Willms, 1995; Willms, 1992; Willms & Raudenbush, 1989). A number of school processes have been shown to affect student achievement, such as school restructuring and various policies and practices that affect the social and academic climate of schools (Bryk & Thum, 1989; Croninger & Lee, 2001; Gamoran, 1996; Lee & Smith, 1993, 1999; Lee et al., 1997; Phillips, 1997; Rumberger, 1995).

## 13.2. DATA AND SAMPLE SELECTION

### 13.2.1. Data

Like all quantitative studies, school effectiveness research requires suitable data. The conceptual framework discussed earlier shows that student outcomes are influenced by a number of different factors operating at different levels within the educational system, including student factors, family factors, and school factors. Generally, insightful school effectiveness research requires data on all those factors. Moreover, as we discuss below, longitudinal models are useful for addressing certain research questions and required

repeated measurements of student outcomes over time. For these reasons, the data requirements of multilevel school effectiveness models can be extensive.

Meeting these extensive data requirements necessitates considerable resources, which are not often available to small-scale studies. For this reason, the federal government has invested in the design and collection of several large-scale longitudinal studies that have been the basis for most school effectiveness studies conducted over the past 40 years or so. Early studies were based on national and some local (state) longitudinal surveys conducted on cohorts of high school students (e.g., see Alexander & Eckland, 1975; Hauser & Featherman, 1977; Jencks & Brown, 1975; Summers & Wolfe, 1977). The U.S. Department of Education conducted the 1972 National Longitudinal Study of the High School Class of 1972, the 1980 High School and Beyond study of 10th- and 12th-grade students, the 1988 National Education Longitudinal Study of 8th graders, and, most recently, the 1998 Early Childhood Longitudinal Study (ECLS) of the kindergarten class of 1998–1999 and the birth cohort of 2000, as well as the 2002 Educational Longitudinal Study of 10th graders.<sup>4</sup> All these survey programs involve large samples of students and schools along with student, parent, teacher, and school surveys as well as specially designed student assessments of academic achievement. One drawback of these studies is that they rarely have adequate classroom-level sample sizes, which makes investigations of teaching and classroom effects problematic. Until recently, all the federal education studies focused on middle and high school students, which has resulted in an inordinate proportion of the school effectiveness research in the past 20 years being directed at middle and high schools. With the availability of ECLS data, that focus seems to be shifting toward elementary schools.

### 13.2.2. Sample Selection

Once an appropriate set of data is selected, the next step in conducting a school effectiveness study is to select an appropriate sample. In addition to selecting a set of data and a sample based on the types of research questions that are to be addressed, two other issues are important to consider: missing data and sampling bias.

4. For further information, visit the National Center for Education Statistics Web site at <http://nces.ed.gov/surveys/>.

### 13.2.2.1. Missing Data

Missing data are a reality in social research and especially problematic in longitudinal analyses in which attrition tends to exacerbate the problem. In panel studies, attrition may occur when families move or students drop out between waves or students cannot be located for some other reason at the follow-up survey. Another situation is nonresponse on certain items. Deciding how to deal with missing values is a common dilemma. Perhaps the most widely used approach is to omit cases with missing data, although the general consensus is that deletion is only an appropriate course of action when data are missing completely at random (see Little & Rubin, 1987, for a detailed treatment of types of “missingness” and remedies). Deletion of cases in other situations can bias the sample and parameter estimates. For that reason, it is important to consider alternatives to deletion.

### 13.2.2.2. Sampling Bias

Sampling bias arises when some part of the target population is inadequately represented in the sample. This problem is often an outcome of deleting cases with missing data and, as mentioned above, can lead to distorted results.<sup>5</sup> Other times, researchers may choose to exclude some valid cases for one reason or another. For example, dropouts and mobile students may be excluded from a school effectiveness evaluation analysis because their achievement growth cannot be attributed to a single school. Whether cases have missing data or are being considered for removal for another reason, deletion is an option that should only be considered after establishing that those cases do not differ systematically from the rest. In general, the larger the percentage of cases being excluded, the greater the potential for selection bias. However, to be safe against sampling bias, cases with missing values should not be deleted but rather handled using an appropriate missing value routine.

As the title of this chapter suggests, school effectiveness research generally necessitates a multilevel model because students are nested in classrooms and schools. The previous discussion of selection bias focused on omission of student cases. Omissions at the student level can also bias the school-level sample. A simple example of this is the effect of deleting students with

5. The problem can also arise due to sampling techniques often used in collecting multilevel longitudinal studies, such as the large-scale federal studies mentioned earlier. Such studies typically provide sampling weights that researchers can use to produce accurate estimates of population parameters (e.g., see Carroll, 1996).

missing achievement data. If the omitted cases have lower achievement levels than the retained cases, mean achievement estimates at the school level will also be biased. Furthermore, omitting cases at the student level decreases the average number of students per school, which generally reduces the reliability of the fixed and random coefficients in the model.

## 13.3. USING MULTILEVEL MODELS TO ADDRESS RESEARCH QUESTIONS

A wide range of multilevel models can and have been used to conduct school effectiveness research. The choice of models depends both on the questions the investigator wishes to answer and on the data available to answer them. Two key aspects of the data are relevant in selecting models: whether the data represent measures at a single point in time (cross-sectional) or multiple points in time (longitudinal) and whether the outcome measures are continuously distributed (e.g., standard test scores) or categorical (e.g., dropout rates). In this section, we review a number of different models. We group the models by the types of dependent or outcome variables used in the models and whether the data are cross-sectional or longitudinal:

- achievement (cross-sectional) models with continuous outcomes,
- achievement growth (longitudinal) models with continuous outcomes,
- models with categorical outcomes.

For each group of models, we pose a series of research questions and the models most suited to address them. Then we illustrate the procedures for using them with the sample NELS data.

### 13.3.1. Achievement Models

The most commonly used type of multilevel model for school effectiveness is one in which the dependent variable is student achievement at a single point in time. One reason for the popularity of these models is that they only require one round of data collection, which is both easier and less expensive than multiple rounds of data collection found in longitudinal studies. Moreover, even though there are some inherent limitations in these models, as we discuss below, they can still be used to address a wide range of research questions.

Student achievement models typically specify two distinct components or submodels: (a) models for

student-level outcomes within schools, known as *within-school models*, and (b) models for school-level outcomes, known as *between-school models*, in which the parameters from the within-school model serve as dependent variables in the between-school model. Because the within-school model may contain a number of parameters, each parameter produces its own between-school equation. In most applications, a series of models are estimated that begin with relatively simple models and then add parameters to develop more complete models. Each model is useful for addressing particular types of research questions, so school effectiveness studies typically employ a number of distinct models.

### 13.3.1.1. Do Schools Make a Difference?

This is the most fundamental research question in school effectiveness research that focuses on how much of the variation in student achievement can be attributed to the schools that students attend. Coleman was the first researcher to address this question, and he did it by partitioning the total variation in student achievement into two components: One component consisted of the variation in individual test scores around their respective school means, and the other component consisted of the variation in school means around the grand mean for the entire sample (Coleman, 1990, p. 76). Coleman found that schools only accounted for a small amount of the total variation in student test scores, ranging from 5% to 38% among different grade levels, ethnic groups, and regions of the country (Coleman, 1990, p. 77).

This research question can easily be addressed using a multilevel *unconditional* or *null model*. The first model has no predictor variables in either the within-school or between-school model and is known as a null or one-way ANOVA model:

$$\text{Level 1 model: } Y_{ij} = \beta_{0j} + r_{ij}, r_{ij} \sim N(0, \sigma^2).$$

$$\text{Level 2 model: } \beta_{0j} = \gamma_{00} + \mu_{0j}, \mu_{0j} \sim N(0, \tau_{00}).$$

$$\text{Combined model: } Y_{ij} = \gamma_{00} + \mu_{0j} + r_{ij}.$$

In this case, the Level 1 model represents the achievement of student  $i$  in school  $j$  as a function of the average achievement in school  $j$  ( $\beta_{0j}$ ) and a student-level error term ( $r_{ij}$ ), and the Level 2 model represents the average achievement in school  $j$  as a function of the grand mean of all the school means ( $\gamma_{00}$ ) and a school-level error term ( $\mu_{0j}$ ). In addition to providing an estimate of the one fixed effect, the grand mean for

achievement ( $\gamma_{00}$ ), the model also provides estimates for the student-level ( $\sigma^2$ ) and at the school-level ( $\tau_{00}$ ) variance components, which can be used to determine how much of the total variance is accounted for by students and schools.

We can illustrate the usefulness of the null model with the NELS data using 10th-grade math test scores as the dependent variable. The estimated parameters from this model are shown in Table 13.1 (column 1).<sup>6</sup> The estimate for the grand mean of the mean math achievement ( $\hat{\gamma}_{00}$ ) among the sample of 912 high schools is 50.85, which is very close to the actual mean for the students in the sample (see appendix). The estimated values for the two variance components can be used to partition the variance in student math scores between the student and school levels, as shown as follows:

$$\text{Student-level variance } (\hat{\sigma}^2) : 73.88$$

$$\text{School-level variance } (\hat{\tau}_{00}) : 24.12$$

$$\text{Total variance: } 98.00$$

$$\text{Proportion of variance at school level : } .25$$

The results show that 25% of the total variance is at the school level, which suggests that schools do indeed contribute to differences in student math scores. This result is within the range that Coleman et al. found in their 1966 study<sup>7</sup> and the range found in other recent studies of student achievement using similar models (e.g., Lee & Bryk, 1989; Rumberger & Willms, 1992). Once the total variance is decomposed into its student and school components, subsequent models can be constructed to explain each component, much the way single-level regression models are used to explain variance.

### 13.3.1.2. To What Degree Does Mean Achievement Vary Across Schools?

This is a related question that allows the researcher to determine the extent of the variation in average school achievement among schools. This question can also be addressed by using the parameter estimates from the unconditional model to calculate a 95% confidence interval, referred to as a range of plausible values, under the assumption that the school-level variance

6. Because of space considerations, we only provide estimates of fixed and random effects. Raudenbush and Bryk (2002) also suggest that researchers examine other statistics, including reliability.

7. Coleman (1990) provides a summary of the findings in Table 3.22.1 on page 77.

**Table 13.1** Parameter Estimates for Alternative Multilevel Math Achievement Models

	<i>Null Model</i> (1)	<i>Means-as- Outcomes Model 1</i> (2)	<i>Means-as- Outcomes Model 2</i> (3)	<i>One-Way ANCOVA Model</i> (4)	<i>Random- Coefficient Model</i> (5)	<i>Intercepts- and Slopes- as-Outcomes Model</i> (6)
<i>Fixed effects</i>						
Model for school mean achievement ( $\beta_0$ )						
INTERCEPT ( $\gamma_{00}$ )	50.85** (0.18)	49.93** (0.17)	50.85** (0.12)	50.96** (0.12)	50.84** (0.18)	50.84** (0.11)
MEANSES ( $\gamma_{01}$ )			8.11** (0.25)			8.11** (0.25)
CATHOLIC ( $\gamma_{02}$ )		3.22** (0.62)	-0.21 (0.43)			-0.23 (0.43)
PRIVATE ( $\gamma_{03}$ )		9.35** (0.64)	0.76 (0.53)			0.73 (0.53)
Model for SES achievement slope ( $\beta_1$ )						
INTERCEPT ( $\gamma_{10}$ )				4.95** (0.10)	4.22** (0.12)	4.51** (0.13)
MEANSES ( $\gamma_{11}$ )						1.09** (0.30)
CATHOLIC ( $\gamma_{12}$ )						-1.78** (0.55)
PRIVATE ( $\gamma_{13}$ )						-3.55** (0.55)
<i>Variance components</i>						
Within school (Level 1) ( $\sigma^2$ )	73.88	73.91	73.95	66.55	65.88	65.97
Between school (Level 2)						
School means ( $\tau_{00}$ )	24.12**	17.33**	5.35**	9.00**	24.75**	5.93**
SES achievement slopes ( $\tau_{11}$ )					1.34**	0.82*
Proportion explained						
School means		.28	.77	.63		.75
SES achievement slopes						.29

NOTE: SES = socioeconomic status; PRIVATE = private schools; CATHOLIC = Catholic schools; MEANSES = mean socioeconomic status.  
\* $p < .05$ ; \*\* $p < .01$ .

is normally distributed (Raudenbush & Bryk, 2002, p. 71):

$$\begin{aligned} \text{Range of plausible values} &= \hat{\gamma}_{00} \pm 1.96 (\hat{\tau}_{00})^{1/2} \\ &= 50.85 \pm 1.96 (24.12)^{1/2} \\ &= (41.23, 60.47). \end{aligned}$$

These results indicate a substantial range in average achievement among high schools, with average achievement 50% higher in the highest performing (97.5th percentile) compared to the lowest performing (2.5th percentile) high schools.

### 13.3.1.3. What School Inputs

#### *Account for Differences in School Outputs?*

Another fundamental research question on school effectiveness concerns the relationship between school inputs and school outputs. Again, this is one of the

main questions that Coleman et al. (1966) addressed in their landmark study (summarized in Coleman, 1990, p. 2), and it continues to have importance for policy initiatives designed to address disparities in school inputs.

This research question can be addressed using a second type of multilevel model, known as a *means-as-outcomes model*. This model attempts to explain school-level variance, but not student-level variance, by adding school-level predictors to the model, as shown in the following example in which we add two indicator or dummy variables for school sector:

$$\text{Level 1 model: } Y_{ij} = \beta_{0j} + r_{ij}.$$

$$\begin{aligned} \text{Level 2 model: } \beta_{0j} &= \gamma_{00} + \gamma_{01}\text{CATHOLIC}_j \\ &+ \gamma_{02}\text{PRIVATE}_j + u_{0j}. \end{aligned}$$

In this example, there are three fixed effects: one for the mean math achievement in public high schools ( $\gamma_{00}$ ), one for the mean achievement difference

between public and Catholic schools ( $\gamma_{01}$ ), and one for the mean achievement difference between public and private, non-Catholic schools ( $\gamma_{02}$ ). The results of this model (see Table 13.1, column 2) show that mean student math achievement is 49.93 in public schools and averages more than 3 points higher in Catholic schools and more than 9 points higher in private schools. Both predictor variables are statistically significant.<sup>8</sup>

With these two predictors in the model, the school-level variance ( $\tau_{00}$ ) is now a conditional variance or the variance that remains after controlling for the effects of school sector (CATHOLIC, PRIVATE). Consequently, it is generally smaller than the variance in the unconditional model. The difference in the two variance estimates can be used to determine how much of the unconditional variance is explained by the model containing these two predictors:

Proportion of variance explained

$$\begin{aligned} &= [\hat{\tau}_{00}(\text{Model 1}) - \hat{\tau}_{00}(\text{Model 2})] / \hat{\tau}_{00}(\text{Model 1}) \\ &= [24.12 - 17.33] / 24.12 \\ &= .28. \end{aligned}$$

The results indicate that 28% of the total variance between schools in mean math achievement is accounted for by the two school sector variables.

Next we added a third predictor to the school-level model, mean socioeconomic status of students in each school (MEANSES<sub>*j*</sub>):

$$\begin{aligned} \text{Level 2 model: } \beta_{0j} &= \gamma_{00} + \gamma_{01}\text{MEANSES}_j \\ &+ \gamma_{02}\text{CATHOLIC}_j + \gamma_{03}\text{PRIVATE}_j + u_{0j}. \end{aligned}$$

In this example, there are four fixed effects: the mean math achievement in public high schools, where MEANSES is zero ( $\gamma_{00}$ );<sup>9</sup> the effect of school mean socioeconomic status (SES) on mean math achievement ( $\gamma_{01}$ ); the mean achievement difference between public and Catholic schools, holding constant school mean SES ( $\gamma_{02}$ ); and the mean achievement difference between public and private, non-Catholic schools, holding constant school mean SES ( $\gamma_{03}$ ). The results of this model (see Table 13.1, column 3) show that MEANSES has a large and statistically significant effect on mean math achievement ( $\hat{\gamma}_{01} = 8.11$ ,  $p < .01$ )—a one standard deviation increase in

MEANSES increases mean test scores by 4.22 ( $8.11 \times .52$ ) points. After controlling for school mean SES, the coefficients for Catholic and private schools are no longer statistically significant. This example illustrates the importance of correctly specifying a model to yield valid and unbiased results. Although this issue applies to all statistical models, it is particularly important in multilevel models because the researcher must draw on a broader array of research literature pertaining to both individual and school determinants of student achievement to correctly specify models at each level of analysis.

This model explains 77% of the school-level variance. In other words, only three predictors explain the majority of the variability in average achievement among schools.<sup>10</sup>

#### 13.3.1.4. What Difference Does the School a Child Goes to Make in the Child's Achievement?

This is another fundamental question that Coleman (1990, p. 2) addressed in his landmark study and one particularly important to parents. Parents are often interested in selecting a school that will improve their child's academic achievement. They are also aware that the average achievement varies widely among schools, in part because schools, state education agency Web sites, and newspapers often report such information. Yet, all the variance in student achievement at the school level cannot be attributed to the effects of schools. Some of that variance is due to the individual background characteristics of the students, which affect student outcomes no matter where they attend school.

This research question can be addressed using another type of multilevel model, known as a *one-way ANCOVA model*. One helpful technique to control for the effects of student background characteristics in this model is through “centering” student-level predictors around their grand or sample mean.

A simple illustration of this model is shown in the following model, in which a single student-level predictor, SES, is introduced and centered on the grand mean:

$$\text{Level 1 model: } Y_{ij} = \beta_{0j} + \beta_{1j}(\text{SES}_{ij} - \overline{\text{SES}}_{..}) + r_{ij}.$$

$$\text{Level 2 model: } \beta_{0j} = \gamma_{00} + u_{0j}.$$

$$\beta_{1j} = \gamma_{10}.$$

8. Hypothesis testing for both fixed and random effects is explained in detail in Raudenbush and Bryk (2002, pp. 56–65). The  $p$ -values shown in Tables 13.1 and 13.2 are from single-parameter tests, which are based on  $t$ -tests for fixed effects and chi-square tests for the variance components.

9. This is extremely close to the sample mean of .01.

10. In fact, mean SES alone explains 77% of the variance, which is why Coleman concluded that the social composition of the school is the most important school input.

Grand-mean centering alters the meaning of the intercept term ( $\beta_{0j}$ ). Instead of representing the actual mean achievement of students in each school, it now represents the expected achievement of a student whose background characteristics are equal to the grand mean of all students in the larger sample of students (Raudenbush & Bryk, 2002, p. 33). In other words, the school means are adjusted for differences in the background characteristics of the students attending them and now represent the expected achievement of an “average” student. In this example, there are two fixed effects: one for the school mean of the expected math achievement for students with mean SES ( $\gamma_{00}$ ) and one for the predicted effect of student SES on math achievement ( $\gamma_{10}$ ).<sup>11</sup> In addition, the equation for the student-level predictor is “fixed” at Level 2 in this model because no random school effect is specified, which assumes that the effect of student SES does not vary among schools (like a classical ANCOVA model)—an assumption that we test below. In this case, the student-level variance ( $\sigma^2$ ) represents the residual variance of student achievement after controlling for student SES, and the school-level variance ( $\tau_{00}$ ) represents the variance among schools in adjusted school means.

The estimated parameters of this model (see Table 13.1, column 4) show that student SES is a powerful predictor of academic achievement ( $\hat{\gamma}_{10} = 4.95$ ,  $p < .01$ ). A one standard deviation increase in student SES implies a 4-point ( $4.95 \times .81$ ) increase in student achievement. This single predictor, grand-mean centered, explains 63% of the school-level variance. In other words, almost two thirds of the observed variance in mean math achievement among schools can be explained by differences in the SES background of the students who attend them. The magnitude of this impact can also be illustrated by calculating the adjusted range of plausible values:

$$\begin{aligned} \text{Range of plausible values} &= \hat{\gamma}_{00} \pm 1.96(\hat{\tau}_{00})^{1/2} \\ &= 50.85 \pm 1.96(9.00)^{1/2} \\ &= (45.08, 56.84). \end{aligned}$$

These results indicate that for a student from an average SES background, his or her expected achievement would be about 26% higher in the highest performing compared to the worst-performing high school. Although such a difference is only about half of the

11. In cases in which student characteristics affect educational outcomes at both the individual and school levels, as we discuss below, then the student-level predictors in this model produce biased estimators of the within-school effects of those characteristics (see Raudenbush & Bryk, 2002, pp. 135–139).

range in the overall means shown earlier, it may still be considered meaningful.

### 13.3.1.5. Do the Effects of Student Background Characteristics Vary Among Schools?

In the preceding example, we assumed that the effects of the student-level predictors were the same across schools. In most cases, the investigator should test this assumption by first specifying them as random at the school level. If the variance of the random effect is not significantly different from zero, the researcher can “fix” the predictor by removing the random effect. If the variance is significantly different from zero, the researcher can then try to explain the variance by adding school-level predictors much the same way that school-level predictors are added to the intercept term.

This type of multilevel model is known as a *random-coefficient model*. To derive accurate estimates of all the variance parameters in this type of model, we must use a different form of centering known as *group-mean centering* (see Raudenbush & Bryk, 2002, pp. 143–149). In this case, the student-level predictors are centered at the mean for the students in their respective schools, and, by doing so, the intercept term ( $\beta_{0j}$ ) represents the unadjusted mean achievement for the school (Raudenbush & Bryk, 2002, p. 33).<sup>12</sup>

To illustrate this model, we estimated a model similar to the one above, but SES was group-mean centered, and a random term was added to its Level 2 equation:

$$\begin{aligned} \text{Level 1 model: } Y_{ij} &= \beta_{0j} + \beta_{1j}(\text{SES}_{ij} - \overline{\text{SES}}_{\cdot j}) + r_{ij}. \\ \text{Level 2 model: } \beta_{0j} &= \gamma_{00} + u_{0j}. \\ \beta_{1j} &= \gamma_{10} + u_{1j}. \end{aligned}$$

In this example, there are two fixed effects—the grand mean of the mean math achievement among schools ( $\gamma_{00}$ ) and the mean of the SES achievement slope among schools ( $\gamma_{10}$ )—and three random effects: the residual variance of student achievement after controlling for student SES ( $\sigma^2$ ), the variance in the average math achievement among schools ( $\tau_{00}$ ), and the variance in the SES achievement slopes among schools ( $\tau_{11}$ ). The results from this model (see Table 13.1, column 5) show similar parameter estimates for mean achievement and student SES compared to the previous ANCOVA model (column 4), but now the variance parameter for the intercept term is similar to that of the unconditional model (column 1), and there is a variance estimate for the SES equation,

12. In addition, group-mean centering provides an unbiased estimator of the student-level effects (see Raudenbush & Bryk, 2002, pp. 135–139).

which in this case is statistically significant.<sup>13</sup> This suggests that the effects of SES on achievement, sometimes referred to as the SES achievement slope, vary among schools. The magnitude of this variation can be illustrated by calculating a range of plausible values:

$$\begin{aligned} \text{Range of plausible values} &= \hat{\gamma}_{10} \pm 1.96 (\hat{\tau}_{11})^{1/2} \\ &= 4.22 \pm 1.96 (1.34)^{1/2} \\ &= (1.95, 6.49). \end{aligned}$$

The results suggest that the effects of student SES on achievement are more than three times as great in some high schools as in other high schools, which suggests that some schools are more equitable in that they attenuate the effects of student background characteristics on achievement.

### 13.3.1.6. How Effective Are Different Kinds of Schools?

One of the most important policy questions concerns measuring school effectiveness. Policymakers are interested in identifying effective and ineffective schools to recognize the effective schools and intervene in the ineffective schools. But this is easier said than done. Schools should only be accountable for the factors that they have control over. In most cases, at least in the public sector, schools do not have control over the types of students who are enrolled in them (as well as other types of school inputs). As we demonstrated earlier, the background characteristics of students explain much of the variation in mean achievement among schools. In addition, student background characteristics can affect student outcomes at the school level, which are known as compositional or contextual effects (Gamoran, 1992). For example, the average SES of a school may have an effect on student achievement above and beyond the individual SES levels of students in that school. In other words, a student attending a school where the average SES of the student body is low may have lower achievement outcomes than a student from a similar background attending a school where the average SES of the student body is high. Data from the 2000 National Assessment of Educational Progress confirm this: Low-income students attending schools with less than 50% low-income students had higher scores in the fourth-grade math exam than middle-income students attending schools with more than 75% low-income students (U.S. Department of Education, 2003, p. 58).

School effectiveness may be judged not simply by determining which schools have higher average achievement, after controlling for certain inputs, but also by how successful they are in attenuating the relationship between student background characteristics and achievement, as we suggested earlier. Coleman (1990, p. 2) argued that there is another important question about school effectiveness: *How much do schools overcome the inequalities with which children come to school?* For example, some earlier studies found that not only did Catholic schools have higher achievement than public schools, even after controlling for differences in the average SES of students, but the relationship between student SES and achievement was lower, meaning that disparities between high and low SES students was lower (Byrk et al., 1993; Lee & Bryk, 1989). In other words, Catholic schools were found to be more equitable.

A type of multilevel model that can be used to assess both questions on school effectiveness is referred to as a *means- and slopes-as-outcomes model*. This model incorporates school-level predictors in both the intercept and random slopes equations. To generate accurate parameter estimates in these types of models, one must introduce a common set of school-level predictors in all the Level 2 equations (see Raudenbush & Bryk, 2002, p. 151). In addition, to disentangle the individual and compositional effects of student-level predictors, one should include school-level means of all the student-level predictors in the model (see Raudenbush & Bryk, 2002, p. 152).

An example of this model is the following:

$$\text{Level 1 model: } Y_{ij} = \beta_{0j} + \beta_{1j}(\text{SES}_{ij} - \overline{\text{SES}}_j) + r_{ij}.$$

$$\begin{aligned} \text{Level 2 model: } \beta_{0j} &= \gamma_{00} + \gamma_{01}\text{MEANSES}_j \\ &+ \gamma_{02}\text{CATHOLIC}_j + \gamma_{03}\text{PRIVATE}_j + u_{0j}. \end{aligned}$$

$$\begin{aligned} \beta_{1j} &= \gamma_{10} + \gamma_{11}\text{MEANSES}_j \\ &+ \gamma_{12}\text{CATHOLIC}_j + \gamma_{13}\text{PRIVATE}_j + u_{1j}. \end{aligned}$$

In this example, there are eight fixed effects and three random effects. The meaning of the student-level random effect and the effects for the model for school means ( $\beta_{0j}$ ) are similar to those described earlier. In the model for the SES achievement slope ( $\beta_{1j}$ ), there are now four fixed effects: the SES achievement slope in public high schools, where the school mean SES is zero ( $\gamma_{10}$ ); the effect of school mean SES on the SES achievement slope ( $\gamma_{11}$ ); the difference between public and Catholic schools in the SES achievement slope, holding constant school mean SES ( $\gamma_{12}$ ); and the difference between public and private, non-Catholic schools on the SES achievement slope,

13. The SES achievement slope in this model is lower than in the ANCOVA model (4.22 vs. 4.95), which suggests that there are both student-level and school-level effects of SES, something we confirm in the next model.

holding constant school mean SES ( $\gamma_{13}$ ). In this model, the variance ( $\tau_{11}$ ) now represents the residual variance in the SES achievement slopes after controlling for school sector and school SES.

The estimated parameters from this model (see Table 13.1, column 6) yield several important conclusions about differences in school effectiveness among public, private, and Catholic schools. First, unlike the earlier reported studies, the average achievement at private and Catholic schools is not significantly higher than the average achievement at public schools after controlling for the effects of school mean SES. Second, consistent with earlier studies, the effects of student SES on achievement are higher in high-SES schools than lower SES schools and lower in Catholic and private schools than in public schools. For example, the effect of student SES is 4.51 at public schools, with a school mean SES equal to zero; at a Catholic school, it is 2.73 ( $= 4.51 - 1.78$ ), and at a private school, it is 0.96 ( $= 4.51 - 3.55$ ). Third, the SES of students affects school achievement at both the individual and schools levels—that is, student SES has both individual and compositional or contextual effects on student achievement.<sup>14</sup>

### 13.3.2. Achievement Growth Models

Achievement models only examine the relationship between student outcomes and predictor variables at discrete points in time. A drawback of this approach is that it fails to account for the fact that an unknown proportion of the achievement that students demonstrate at a particular point in a school is due to learning that took place prior to their arrival at that school. Although this problem can be partially corrected by including measures of prior achievement in the model, using an outcome measure that isolates the student learning that occurred while students were actually attending that school is a far better choice.

Growth models are a special class of multilevel model in which repeated measurements are collected for each individual in the sample (Singer & Willett, 2003). Growth models are useful for understanding mean patterns of change as well as individual differences in those patterns. Growth models include two or more level of analyses. A growth trajectory

14. As Raudenbush and Bryk (2002) point out, there is more than one way to disentangle the individual and compositional effects of student background characteristics, with the choice of method depending on whether the analyst wishes to test for random slopes (pp. 139–149). In this example, the conditional individual effect of SES (i.e., expected within-school effects on achievement in public schools with MEANSES equal to zero) is 4.51, and the compositional effect of SES =  $8.11 - 4.51 = 3.6$ .

is estimated for each individual at Level 1 of the multilevel model, and between-individual differences in the change pattern are estimated at Level 2.<sup>15</sup> A multilevel achievement growth model for schools will typically include three levels of analysis (e.g., Lee, Smith, & Croninger, 1997; Seltzer, Choi, & Thum, 2003). A special situation arises when there is a need to estimate teacher or classroom effects in addition to school effects. Typically, students will have been members of more than one classroom in a growth model, which means that they are not strictly nested within classrooms over time. In this scenario, a cross-classified random-effects model can be used to partition the variance in student learning into both classroom and school components (see Raudenbush & Bryk, 2002, chap. 12).

In this section, we discuss two different ways of specifying and estimating achievement growth models: one using the multilevel regression models similar to the ones we discussed above and the other using multilevel latent growth curves. As we did earlier, we discuss these models in relation to the types of research questions about school effectiveness they can be used to address.

#### 13.3.2.1. Multilevel Growth Models

We begin with a Level 1 model for individual growth, where repeated, within-student measurements of achievement are modeled as a function time. The simplest model depicts a linear growth trajectory, although piecewise linear and polynomial terms can be added to examine nonlinear trends if there are sufficient observations (see Raudenbush & Bryk, 2002, chap. 6). A Level 1 linear growth model can be written as follows:

$$\begin{aligned} \text{Level 1 model: } Y_{ij} &= \pi_{0ij} + \pi_{1ij}a_{ij} + e_{ij}, \\ e_{ij} &\sim N(0, \sigma^2), \end{aligned}$$

where  $Y_{ij}$  represents the achievement outcome measure of student  $i$  in school  $j$  at time  $t$ ;  $\pi_{0ij}$  and  $\pi_{1ij}$  represent, respectively, the initial status (when time equals zero) and rate of change for student  $i$  in school  $j$ ;  $a_{ij}$  is a measure of time; and  $e_{ij}$  is a random error term. For the NELS data, we coded time 0, 0.5, and 1 for 1988, 1990, and 1992, respectively. Coding the time variable this way offers two advantages in

15. One of the advantages of this approach is that individuals only have to have a single observation to be included in the analysis (Raudenbush & Bryk, 2002, p. 199).

interpreting the results. First, the intercept can be interpreted as an approximation of student achievement level upon entering high school since the first wave of testing was conducted in the spring of 1988, just before most students entered high school. Second, the slope represents achievement gains during the 4-year period of high school.<sup>16</sup>

*13.3.2.1.1. Do schools make a difference in student learning?* This question is similar to the one addressed earlier, except here we are interested in whether schools make a difference in student learning, not simply student achievement. This question can be addressed with a fully unconditional model with no predictors at Levels 2 and 3:

Level 2 model:

$$\begin{aligned}\pi_{0ij} &= \beta_{00j} + r_{0ij}, r_{0ij} \sim N(0, \tau_{\pi 00}) \\ \pi_{1ij} &= \beta_{10j} + r_{1ij}, r_{1ij} \sim N(0, \tau_{\pi 11}).\end{aligned}$$

Level 3 model:

$$\begin{aligned}\beta_{00j} &= \gamma_{000} + u_{00j}, u_{00j} \sim N(0, \beta_{00}). \\ \beta_{10j} &= \gamma_{100} + u_{10j}, u_{10j} \sim N(0, \beta_{10}).\end{aligned}$$

Note that Level 2 here is equivalent to Level 1 in the multilevel cross-sectional model. In this model, there are two fixed effects: one for initial status or achievement ( $\gamma_{000}$ ) and one for achievement growth ( $\gamma_{100}$ ), with the latter being of primary focus. There are also five random effects, which can be used to partition the variance in both initial achievement and achievement growth into their within- and between-school components.

We can illustrate this technique with the NELS data and math test scores in Grades 8, 10, and 12 as the dependent variables. The estimated parameters from this model are shown in Table 13.2 (column 1). The results indicate that the average math score for students entering high school ( $\hat{\gamma}_{000}$ ) is 45.87 points and that students increase their math scores ( $\hat{\gamma}_{100}$ ) by an average of 8.76 points over 4 years. The estimated values for the variance components can be used to partition the variance in both initial status and learning between students and schools as we did earlier.<sup>17</sup> The results

show that about one quarter of the total variance in both initial achievement and achievement growth occurs at the school level in this sample of data (see Table 13.3).

The proportion of variance in achievement growth at the school level is similar to the proportion we calculated earlier for 10th-grade achievement. In another study using this same data set, we found the proportion varied by subject area—ranging from a low of 20% in reading to a high of 60% in history (Rumberger & Palardy, 2003a). One study of elementary schools in Chicago found that almost 60% of the variance in achievement growth occurred at the school level (Raudenbush & Bryk, 2002, p. 239). In general, these studies suggest schools account for a sizable amount of variance in both student achievement and achievement growth.<sup>18</sup>

To illustrate the usefulness of the growth outcome and to draw comparisons between it and the achievement outcome, we estimate a series of achievement growth models to address the questions we posed above for the achievement models. The results are shown in Table 13.2. Because of space limitations, we will not discuss all of the results of these models, but instead we will point out where the results of these models yield different answers to the set of questions about school effectiveness.

For example, consider the following question: *Do the effects of student background characteristics vary among schools?* To address this question, we specify a random-coefficient model similar to the one estimated earlier, where student SES is group-mean centered:

Level 2 model:

$$\begin{aligned}\pi_{0ij} &= \beta_{00j} + \beta_{01j}(\text{SES}_{ij} - \overline{\text{SES}}_{.j}) + r_{0ij}. \\ \pi_{1ij} &= \beta_{10j} + \beta_{11j}(\text{SES}_{ij} - \overline{\text{SES}}_{.j}) + r_{1ij}.\end{aligned}$$

Level 3 model:

$$\begin{aligned}\beta_{00j} &= \gamma_{000} + u_{00j}. \\ \beta_{01j} &= \gamma_{010} + u_{01j}. \\ \beta_{10j} &= \gamma_{100} + u_{10j}. \\ \beta_{11j} &= \gamma_{110} + u_{11j}.\end{aligned}$$

The achievement model estimated earlier (see Table 13.1, column 5) found that the effect of student

16. An alternative scheme is 0, 2, 4, which also sets the intercept as achievement upon entering high school, but now the growth parameter is scaled so that it is interpreted as achievement gains per year.

17. The variance components can also be used to examine the correlation between initial status and growth at both the individual and school levels (see Raudenbush & Bryk, 2002, p. 240). In this example, the correlation at the student level is .34, and the correlation at the school level is .39,

which suggests that students who begin high school with higher math achievement have higher achievement growth rates than lower achieving students.

18. Because of the size and heterogeneity of course offerings (tracking) found in high schools, schools may account for a great proportion of the variance at the elementary level compared to the secondary level.

**Table 13.2** Parameter Estimates for Alternative Multilevel Math Achievement Growth Models

	<i>Null Model</i> (1)	<i>Means-as- Outcomes Model 1</i> (2)	<i>Means-as- Outcomes Model 2</i> (3)	<i>Random- Coefficient Model</i> (4)	<i>One-Way ANCOVA Model 1</i> (5)	<i>One-Way ANCOVA Model 2</i> (6)
<i>Fixed effects</i>						
Model for initial status ( $\pi_{0ij}$ )						
Model for school mean of initial status ( $\beta_{00j}$ )						
INTERCEPT ( $\gamma_{000}$ )	45.87** (0.16)	45.10** (0.15)	45.82** (0.10)	45.87** (0.16)	45.97** (0.11)	45.92** (0.10)
MEANSES ( $\gamma_{001}$ )			7.17** (0.22)			3.60** (0.25)
CATHOLIC ( $\gamma_{002}$ )		2.18** (0.54)	-0.86* (0.39)			-0.87* (0.39)
PRIVATE ( $\gamma_{003}$ )		8.21** (0.62)	0.61 (0.50)			0.33 (0.50)
Model for within-school relationship between SES and initial status ( $\beta_{01j}$ )						
INTERCEPT ( $\gamma_{010}$ )				3.58** (0.10)	4.19** (0.09)	3.59** (0.10)
Model for 4-year learning rate ( $\pi_{1ij}$ )						
Model for school mean of 4-year learning rate ( $\beta_{10j}$ )						
INTERCEPT ( $\gamma_{100}$ )	8.76** (0.08)	8.49** (0.08)	8.66** (0.08)	8.76** (0.18)	8.79** (0.07)	8.69** (0.11)
MEANSES ( $\gamma_{101}$ )			1.65** (0.16)			0.53** (0.25)
CATHOLIC ( $\gamma_{102}$ )		1.84** (0.36)	1.15** (0.34)			1.15** (0.37)
PRIVATE ( $\gamma_{103}$ )		2.15** (0.33)	0.40 (0.37)			0.40 (0.37)
Model for within-school relationship between SES and 4-year learning rate ( $\beta_{01j}$ )						
INTERCEPT ( $\gamma_{110}$ )				1.12** (0.08)	1.28** (0.07)	1.12** (0.08)
<i>Variance components</i>						
Within students (Level 1) ( $\sigma^2$ )	8.18	8.18	8.18	8.19	8.20	8.19
Within school (Level 2)						
Initial status ( $\tau_{\pi 00}$ )	49.81**	49.83**	49.86**	44.04**	44.62**	44.50**
Four-year learning rate ( $\tau_{\pi 11}$ )	13.05**	13.05**	13.04**	12.23**	12.47**	12.46**
Between school (Level 3)						
Initial status ( $\tau_{\beta 00}$ )	19.11**	13.98**	4.66**	19.59**	7.75**	5.03**
SES/initial status ( $\tau_{\beta 01}$ )				0.98*		
Four-year learning rate ( $\tau_{\beta 10}$ )	4.00**	2.91**	2.39**	3.43**	2.57**	2.42*
SES/4-year learning rate ( $\tau_{\beta 11}$ )				0.55		
Proportion school-level variance explained						
Initial status		.27	.76		.59	.74
Four-year learning rate		.27	.40		.36	.40

NOTE: SES = socioeconomic status; PRIVATE = private schools; CATHOLIC = Catholic schools; MEANSES = mean socioeconomic status.  
\* $p < .05$ ; \*\* $p < .01$ .

SES (group-mean centered) on math achievement ( $\gamma_{10} = 4.22, p < .01$ ) varied significantly between schools ( $\tau_{11} = 1.34, p < .01$ ). As a result, several predictors were added to the model, and it was found that the effect of SES on math achievement was

lower in Catholic and private schools—that is, the distribution of achievement appeared to be more equitable in Catholic schools. In the achievement growth model (see Table 13.2, column 4), however, the effect of student SES (group-mean centered) on math

**Table 13.3** Decomposing the Variance in a Linear Math Achievement Growth Model

	<i>Initial Status</i>	<i>Achievement Growth</i>
Student-level variance (%)	49.81	13.05
School-level variance (%)	19.11	4.00
Total student-level and school-level variance (%)	68.92	17.05
Proportion of variance at school level	.28	.24

achievement growth ( $\gamma_{110} = 8.76, p < .01$ ) did not vary significantly between schools ( $\hat{\tau}_{\beta 11} = .055, p \geq .05$ ).<sup>19</sup>

Consequently, the effect of student SES on initial status and achievement growth was fixed, and a one-way ANCOVA model (with SES was grand-mean centered) was estimated:<sup>20</sup>

Level 2 model:

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}(\text{SES}_{ij} - \overline{\text{SES}}_{..}) + r_{0ij}.$$

$$\pi_{1ij} = \beta_{10j} + \beta_{11j}(\text{SES}_{ij} - \overline{\text{SES}}_{..}) + r_{1ij}.$$

Level 3 model:

$$\beta_{00j} = \gamma_{000} + u_{00j}.$$

$$\beta_{01j} = \gamma_{010}.$$

$$\beta_{10j} = \gamma_{100} + u_{10j}.$$

$$\beta_{11j} = \gamma_{110}.$$

The results (see Table 13.2, column 5) show that not only do differences in student SES explain a large proportion of variance between schools in initial achievement (.59), but these differences also explain a substantial proportion of the variance between schools in achievement growth (.36). Nonetheless, even after

controlling for student SES, significant variation in student achievement growth remains. This model answers a similar yet more important question that the earlier model could not address: *What difference does the school a child goes to make in the child's learning (as opposed to achievement)?*

*13.3.2.1.2. How effective are different kinds of schools?* To address this question, we estimated a second ANCOVA model with the same set of predictors as in the earlier achievement model. Because student SES is grand-mean centered in this model, the model estimates the effects of the school-level predictors on the adjusted school mean—in this case, the expected achievement growth for a student with average SES. As a result, the coefficient for school SES provides a direct estimate of the contextual or compositional effect of student SES. In this example, the individual ( $\gamma_{110}$ ) and contextual ( $\gamma_{101}$ ) effects of student SES are both significant—a one standard deviation increase in student SES increases 4-year learning rates by .91 ( $1.12 \times .81$ ) units, or about 4 months of learning over a 4-year period, and a one standard deviation increase in school SES increases learning rates by .28 ( $.53 \times .51$ ), or about 1 month of learning over a 4-year period. After controlling for school SES, the results also show that learning rates in math are not significantly higher in private schools than in public schools ( $\gamma_{103} = .40, p \geq .05$ ), but they are significantly higher in Catholic schools than in public schools ( $\gamma_{102} = 1.15, p < .05$ ): 8.69 in public schools versus 9.84 ( $8.69 + 1.15$ ) in Catholic schools.

The preceding question on school effectiveness focused on differences between different kinds of schools. To more thoroughly address this question, an investigator needs to develop a more comprehensive model that more adequately controls for a variety of differences in student background characteristics (e.g., prior achievement, aspirations, school experiences) and a variety of other school inputs that schools typically have little control over (e.g., teachers, textbooks, facilities, location), as suggested by the earlier conceptual framework (see, e.g., Rumberger & Palardy, 2003a). Yet one additional important question remains to be addressed: *Why are some schools more effective than others?* If schools are to be improved, it is important not just to more accurately identify effective and ineffective schools but also to determine why some schools are more effective than others. By identifying the factors that contribute to school effectiveness, it may be possible to use the information to improve existing schools. Based on the framework presented earlier, this involves identifying the factors

19. This result is based on a single-parameter hypothesis in which  $p = .05$ , the threshold of statistical significance. Investigators one can also use a multiparameter test that tests for significant differences in the entire array of variances and covariances between two separate models (Bryk & Raudenbush, 2002, pp. 63–65). In this case, the results of the multiparameter test confirmed the results of the single-parameter test—that is, a model with a fixed SES/achievement growth term was not significantly different from a model with a random SES/achievement growth term. Similar models and procedures can be used to examine differences in achievement growth rates between students within the same school (see Seltzer, Choi, & Thum, 2003).

20. Because we focused on achievement growth, we fixed the effect of student SES on initial status, even though its variance was significantly different from zero. As in the case of achievement models that we discussed earlier, the grand-mean centered student-level predictors produce biased estimators of the within-school effects of student characteristics when those characteristics affect educational outcomes at both the individual and school levels. The estimates in Table 13.2, column 6 suggest that is the case in this example, especially for initial status.

that mediate the relationship between school inputs and school outcomes as well as explain variance in mean student achievement, which we refer to as process variables.

We illustrate this by estimating two additional ANCOVA models. First, we estimate a model that includes a single school-level predictor, MEANSES, because it represents a school input that has been shown to strongly affect math learning. The second model adds two school process variables that teachers and other school personnel have at least partial control over: MEANNAEP, which measures the mean number of college-prep courses (as designated by the National Assessment of Educational Progress [NAEP]) students complete during high school, and MEANHW, the mean amount of time students spend on homework per week during the 10th grade (see the appendix for means and standard deviations for these measures). The first model estimates the total compositional effects of student SES (without additional school-level predictors), and the other can be used to see if the two school process variables mediate the relationship between student composition and achievement growth as well as affect student learning.

Table 13.4 displays the estimates for the school-level predictors in standardized form so that the relative magnitude of effects of these factors can be compared. Results from the first model show that the compositional effect of student SES (MEANSES) is highly significant. Results from the second model show that adding the two process variables reduces the effects of MEANSES to the point that it actually has a negative impact on math learning ( $\hat{\gamma}_{101} = -0.139$ ,  $p < .05$ ). That is, the compositional effects of mean SES is reversed after controlling for the average number of college-prep courses that students take in the school and by the average amount of homework that students do—what some investigators have found other studies and have labeled *academic press* (Lee & Smith, 1999; Phillips, 1997). Moreover, MEANNAEP and MEANHW both have significant positive effects on the mean rate of math learning at schools ( $\hat{\gamma}_{102} = 0.324$ ,  $p < .01$ ;  $\hat{\gamma}_{103} = 0.276$ ,  $p < .01$ ). Notice that although we have concluded that MEANNAEP and MEANHW mediate the effect of MEANSES on mean math learning, we have not examined the exact nature of that relationship. Multilevel regression models are not suited for estimating this type of indirect effects. To address this question, we introduce another class of models: multilevel latent growth curves (MLGC), an extension of the latent growth curve (LGC) in the structural equation modeling (SEM) literature.

**Table 13.4** Standardized Parameter Estimates of School-Level Predictors in Model for School Mean of 4-Year Math Learning Rate ( $\beta_{10j}$ )

	Composition Model (1)	Process Model (2)
MEANSES ( $\gamma_{101}$ )	0.229**	-0.139*
MEANNAEP ( $\gamma_{102}$ )		0.324**
MEANHW ( $\gamma_{103}$ )		0.276**

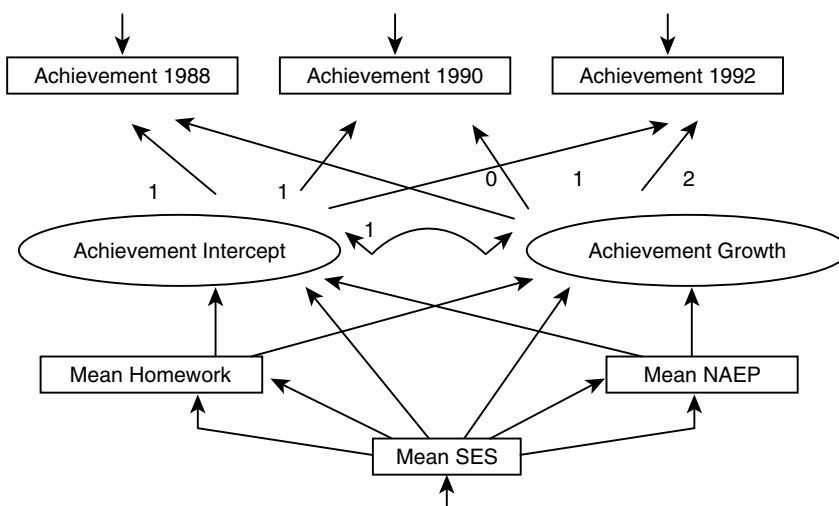
NOTE: MEANSES = mean socioeconomic status; MEANNAEP = mean number of NAEP (college-prep) courses students complete during high school; MEANHW = mean amount of time students spend on homework per week during the 10th grade. This model can be estimated as a three-level multilevel regression model or a two-level multilevel latent growth curve model. The models include student SES (grand-mean centered) in the student-level model (fixed) and a school intercept model with the same three school-level predictors, although those parameter estimates are not shown.

\* $p < .05$ ; \*\* $p < .01$ .

### 13.3.2.2. Multilevel Latent Growth Curves

SEM is widely used among social scientists because of its flexibility for modeling covariance structure in both measurement and structural models. Multilevel SEM has evolved over the past few decades (Muthén, 1989, 1991) but has not received much attention from educational researchers until recently, although Kaplan and his associates have written on its usefulness to the study of school effects (Kaplan & Elliott, 1997; Kaplan & Kreisman, 2000). LGCs (see McArdle & Epstein, 1987; Meredith & Tisak, 1990) are a special class of SEMs designed for modeling between-person change in an outcome over time. LGCs are highly similar to regression-based individual growth trajectories in function, although these two methods evolved independently. Like other single-level SEMs, LGCs have limited applications in the study of school effects because they do not include a school level of analysis.

LGCs have only recently been formulated to analyze multilevel data (Muthén, 1997), resulting in a model that is especially appropriate for the study of school effects. Much like three-level hierarchical linear model (HLM) growth models, this method can accommodate individual growth trajectories, as well as within- and between-school analyses when at least three waves of longitudinal data are available for the student achievement outcome. The appeal of MLGC compared with the multilevel regression growth model is precisely the appeal of SEM over regression models—that is, additional flexibility in specifying covariance relationships, which can result in a more compelling model of school

**Figure 13.2** School-Level Path Diagram of MLGC With Indirect Effects

effects. Additional modeling options include the estimation of latent variables from multiple observed variables, of measurement error on observed variables, of complex measurement error structures, and of group comparison models. One MLGC advantage that stands out in particular is the ability to estimate direct, indirect, and total effects between variables. Although MLGCs are in many ways ideally suited for the study of school effects, they have rarely been applied to this field of study (Palardy, 2003). In this section, we illustrate how this method can be used to estimate direct, indirect, and total effects.<sup>21</sup>

*13.3.2.2.1. What are the magnitudes of the indirect effects of mean SES on mean math learning, flowing through mean NAEP and mean homework?* Recall that this question evolved from our multilevel regression growth model in which we determined that two process variables, MEANNAEP and MEANHW, mediated the effects of MEANSES on mean student achievement growth in math. We now examine the magnitudes and significance levels of those indirect effects.

Figure 13.2 shows the path diagram for the school-level MLGC with the indirect effects of MEANSES flowing through MEANHW and MEANNAEP. Note

that this model is highly similar to the “process” model for which results are displayed in Table 13.4. The same assumptions hold in this model. Here the math achievement intercept and growth factors are estimated by fixing path loadings to a linear arrangement with the intercept centered on Time 1 (1988), but the interpretation of these parameters, as well as their values and standard errors, is equivalent to the intercept and growth parameters of the multilevel regression growth model. Other than its multilevel nature, this model is like other LGCs. Table 13.5 shows the standardized coefficient estimates for the direct and indirect effects of MEANSES. The results show that MEANHW and MEANNAEP are significant mediators of MEANSES on mean achievement growth. Students attending higher SES schools took more college-prep courses (0.590,  $p < .01$ ), which resulted in more learning (0.191,  $p < .01$ ). Similarly, students attending higher SES schools did more homework (0.628,  $p < .01$ ), which resulted in a greater learning rate (0.173,  $p < .01$ ). The total effect of mean SES on student learning is the sum of its direct effect and indirect effects (0.225,  $p < .01$ ). Note that the total effect of mean SES in Table 13.5 is equal to the direct effect of mean SES in the compositional model, with no other covariates shown in Table 13.4. Note that multilevel regression software can estimate some forms of indirect effects.<sup>22</sup>

21. MLGCs can be used to address some additional questions about school effectiveness, including whether parameter estimates vary among different samples of schools and alternative specifications for measurement models for independent and dependent variables. See Palardy (2003) for some examples.

22. For example, HLM software can estimate indirect effects that flow through variables with random effects (see Raudenbush & Bryk, 2002, pp. 356–360).

**Table 13.5** Indirect Effects of Mean SES on Math Achievement Growth Mediated by Mean NAEP and Mean Homework

	<i>Mediating Process Variable</i>	<i>Effect on Mediator</i>	<i>Effect on Growth</i>
Mean SES (direct effect)			-0.139*
	NAEP courses	0.590**	0.191**
	Homework	0.628**	0.173**
Total effects			0.225**

NOTE: Standardized coefficients. Significance levels of indirect effects were computed using the Sobel method. SES = socioeconomic status; NAEP = National Assessment of Educational Progress.

\* $p < .05$ ; \*\* $p < .01$ .

### 13.3.3. Categorical Outcome Models

Most school effectiveness studies have focused on student achievement and other student outcomes that can be estimated with linear models in which the random effects are normally distributed. But some student outcomes cannot be estimated with such models. In particular, student outcomes such as dropout rates are binary, taking on one value if the outcome is present and another value if the outcome is not (e.g.,  $Y = 1$  if the student is a dropout,  $Y = 0$  otherwise). As a result, the random effect can also only take on two values and hence is not normally distributed. Other outcomes can involve several discrete conditions, such as attending a 4-year college, a 2-year college, or no college.<sup>23</sup>

Discrete outcomes require a different type of model from the standard multilevel or hierarchical linear models we have discussed up until this time. These models are known as hierarchical generalized linear models (HGLMs), or simply generalized linear models (see Raudenbush & Bryk, 2002, chap. 10). These methods can be used to estimate a wide range of models using multilevel data, including nonlinear models with random effects that are not normally distributed. In fact, hierarchical linear models simply represent a specific and simple type of generalized linear model.

Estimating generalized linear models requires several additional steps from those we have discussed so far. First, the researcher has to specify a Level 1 sampling model. In the linear case, the sampling model is simply a normal distribution with a mean,  $\mu_{ij}$ , and a variance,  $\sigma$ .<sup>24</sup> Second, the researcher has to specify a link function that transforms the

expected value,  $\Phi_{ij}$ , into a predicted value that can be estimated with a linear model. In the linear case, this link function is simply the value 1 because no transformation is required. Finally, the researcher specifies a linear structural model to estimate the transformed expected value.

We can illustrate this process for the case of school dropouts. For binary student outcomes, such as dropout, the Level 1 sampling model is Bernoulli:

$$\text{Prob}(Y_{ij} = 1|\beta_j) = \Phi_{ij},$$

where  $\Phi_{ij}$  represents the probability of student  $i$  in school  $j$  dropping out of school. The Level 1 link function is a log odds ratio:

$$\eta_{ij} = \log[\Phi_{ij}/(1 - \Phi_{ij})],$$

which has a range of  $-u$  to  $+u$  and takes on the value of 0 when the probability of an outcome equals .5 and the odds of success are even [ $.5/(1 - .5) = 1$ ]. The log odds ratio can be converted to a probability through the following equation:

$$\Phi_{ij} = 1/[1 + \exp\{-\eta_{ij}\}].$$

The Level 1 structural model is similar to the previous Level 1 models. In the case of a null or unconditional model, it is simply

$$\eta_{ij} = \log[\Phi_{ij}/(1 - \Phi_{ij})] = \beta_{0j},$$

and the Level 2 model is exactly as in the linear case:

$$\beta_{0j} = \gamma_{00} + u_{0j}.$$

Conditional models can be constructed by adding Level 1 and Level 2 predictors. And as in the linear case, the analyst must also decide whether the Level 1 predictors should be centered and whether they should be fixed or random at Level 2.

We can illustrate the use of nonlinear models with the NELS data. We first estimated an unconditional model with no Level 1 or Level 2 predictors. The HLM program that we used actually produces two sets of estimates for the fixed effects. The first is a unit-specific estimate, which corresponds to the estimated log odds with a random effect of zero. The second is a population estimate, which provides a better estimate of the true population mean. The second estimate is needed because the nonlinear transformation of probability into log odds means that a symmetrical distribution of log odds results in an asymmetrical distribution of

23. For other examples, see Raudenbush and Bryk (2002, chap. 10).

24. To generate accurate school-level composition measures, we restricted the sample to respondents who had a valid school ID in 1990, had valid

test scores in 1988 and 1990, and attended a high school with at least five students.

probabilities that is positively skewed and thus has a higher mean value than the mean of the log-odds distribution.<sup>25</sup> The difference in these two estimates for dropout rates is shown as follows:

Unit-specific estimated mean: 6.49%  
 Population average estimated mean: 6.94%  
 Sample mean: 6.81%

As the figures show, the population average dropout rate is higher than the unit-specific rate and closer to the sample mean. The two sets of estimates differ not only in the values they produce but also in their assumptions about the underlying distribution of random effects and in the type of questions they can be used to address. In general, unit-specific estimates are more useful for analyzing differences in the effects of Level 1 and Level 2 predictors across Level 2 units, whereas population-average estimates are more useful for estimating average probabilities for the population as a whole.

We next estimated the same model we did earlier with one student-level predictor, SES, and three school-level predictors: MEANSES, CATHOLIC, and PRIVATE. Both SES and MEANSES were centered on the grand mean, which affects the value and interpretation of the intercept term. The unit-specific estimated parameters are shown in Table 13.6. The parameter estimate for the student-level predictor is  $-.868$ . A student with average SES attending a typical public school with average MEANSES would have a predicted log-odds dropout rate of  $-2.843$ , corresponding to a predicted probability of  $1/(1 + \exp\{2.843\}) = .055$ . A student with an SES one unit higher than average attending a typical public school with average MEANSES would have a predicted log-odds dropout rate of  $-2.843 - .868 = -3.711$ , corresponding to a predicted probability of  $1/(1 + \exp\{3.711\}) = .022$ . The average SES of the school, MEANSES, would also affect the odds of dropping out, even after controlling for the individual effects of SES, something referred to as the contextual or compositional effect of SES (which we discuss below). A student with average SES attending a school with a MEANSES one unit higher than average (about two standard deviations, as shown in the appendix) would have a predicted log-odds dropout rate of  $-2.843 - .295 = -3.138$ , corresponding to a predicted probability of  $1/(1 + \exp\{3.138\}) = .042$ . Because of the nonlinear relationship between the log odds and probability, an additional one-unit increase

**Table 13.6** Estimated Parameters for Dropout Models

	<i>Null Model</i>	<i>School Model</i>
<i>Fixed effects</i>		
Model for school mean dropout rate ( $\beta_0$ )		
INTERCEPT ( $\gamma_{00}$ )	$-2.667^{**}$	$-2.843^{**}$
MEANSES ( $\gamma_{01}$ )		$-0.295^{**}$
CATHOLIC ( $\gamma_{02}$ )		$-1.358^{**}$
PRIVATE ( $\gamma_{03}$ )		$-0.913^{**}$
Model for SES dropout slope ( $\beta_1$ )		
INTERCEPT ( $\gamma_{10}$ )		$-0.868^{**}$
<i>Variance components</i>		
Between school ( $\tau_{00}$ )	$.455^{**}$	.207
Proportion school-level variance explained		.545
Reliability	.292	.154

NOTE: SES = socioeconomic status; PRIVATE = private schools; CATHOLIC = Catholic schools; MEANSES = mean socioeconomic status.

\*\* $p < .01$ .

in MEANSES (a 100% increase) would only lower the predicted probability to .031 (a 27% decrease).

### 13.4. IDENTIFYING EFFECTIVE SCHOOLS

Although many school effectiveness studies attempt to identify school-level factors that predict student outcomes based on a sample of schools, some analysts are also interested in identifying individual schools that are particularly effective. That is, even after controlling for a given set of predictors, each school may have a mean student achievement that is above or below the mean predicted from the model. Schools whose mean achievement is above the level predicted by the model can be considered effective, whereas schools whose mean achievement is below the level predicted can be considered ineffective schools.

The unique contribution of each school to its effectiveness is captured in the school-level random effect or error term. Consider the following simple, two-level achievement model:

$$\text{Level 1 model: } Y_{ij} = \beta_{0j} + r_{ij}.$$

$$\text{Level 2 model: } \beta_{0j} = \gamma_{00} + \mu_{0j}.$$

The dependent variable in the Level 2 model,  $\beta_{0j}$ , which represents the average achievement of each school, is composed of a fixed effect,  $\gamma_{00}$ , and a random effect,  $\mu_{0j}$ . Hierarchical analysis produces an empirical Bayes estimator for the random effect

25. The two estimates can be quite similar when the fixed effect is close to zero (which translates to a probability of .5) or when the random effect is close to zero. For a more complete discussion, see Raudenbush and Bryk (2002, pp. 297–304).

**Table 13.7** Models and Estimated Random Effects From Two Dropout Models

	<i>Unconditional Model</i>	<i>Conditional Model</i>
Level 1 model	$\log[\Phi_{ij}/(1 - \Phi_{ij})] = \beta_{0j}$	$\log[\Phi_{ij}/(1 - \Phi_{ij})] = \beta_{0j}$
Level 2 model	$\beta_{0j} = \gamma_{00} + u_{0j}$	$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{MEANSES}_j + u_{0j}$
Estimated random effect ( $u_{0j}$ )		
Mean, standard deviation	0, .365	0, .179
Minimum, maximum	-.759, 1.466	-.409, .710

**Table 13.8** Estimated Fixed and Random Effects for Two Schools

<i>Case</i>	<i>n</i>	<i>MEANSES</i>	<i>Unconditional Model</i>			<i>Conditional Model</i>		
			<i>Fixed Effect</i>	<i>Random Effect</i>	<i>Dropout Rate</i>	<i>Fixed Effect</i>	<i>Random Effect</i>	<i>Dropout Rate</i>
81	15	-.332	-2.667	-.326	4.78	-2.348	-.219	7.13
393	16	.785	-2.667	-.342	4.70	-3.812	-.066	2.03

NOTE: Fixed and random effects are log odds. Dropout rate =  $1/\{1 + \exp[-(\text{Fixed effect} + \text{Random effect})]\}$ .

that provides a better and more stable estimate of the unique school effect than other methods (e.g., ordinary least squares [OLS] estimates) by taking into account group membership and the within-school sample size (Raudenbush & Bryk, 2002, p. 154). More accurate estimates can be obtained by adding school-level variables to the Level 2 model, which provides conditional shrinkage estimates of the random effects (Raudenbush & Bryk, 2002, pp. 90–94). In achievement models, schools with positive random effects have higher than predicted achievement rates and should be considered effective, whereas in dropout models (as we illustrate below), schools with negative random effects have lower than predicted dropout rates and should be considered effective.

To illustrate how this technique can be used to identify effective schools, we can compare the empirical Bayes estimates for the Level 2 random effects from two simple dropout models, one unconditional and one conditional. The two models and descriptive statistics for the empirical Bayes estimates of the Level 2 random effects are shown in Table 13.7.

As the descriptive statistics show, the estimated random effects in the conditional model have a much narrower range and hence smaller standard deviation than the unconditional estimates.

Moreover, the conditional model provides a better way to identify effective schools. Consider the two schools shown in Table 13.8. Based on the unconditional model, both schools are equally effective—their unique or random log-odds dropout rate are both about one third of a logit less than the fixed or expected rate—and hence both schools have similar estimated

dropout rates that are considerably smaller than the average dropout rate of 6.49 for the entire sample of schools. Estimates from the conditional model tell another story, however. School 393 has a much higher average SES than School 81, so its expected log-odds dropout rate is much higher. Yet the unique contribution to its dropout rate—that is, its random effect—is not very large, and hence the school is not particularly effective. In contrast, School 81 has a much higher expected dropout rate (i.e., lower log-odds dropout rate) because its average SES is much lower, yet its estimated dropout rate is actually lower than the expected rate. Hence, School 81 should be considered more effective than School 393, even though it has a higher dropout rate.

### 13.5. SUMMARY AND FUTURE DIRECTIONS

The need for useful and methodologically sound school effectiveness studies has never been greater. Fortunately, the development of large-scale, comprehensive, longitudinal studies of student development has coincided with the development of new and powerful statistical techniques for analyzing the data for these studies. The result has been the continued growth of more sophisticated and comprehensive school effectiveness studies.

Several important challenges remain, however. One is to develop even more comprehensive studies. Although earlier studies were particularly useful for identifying student, family, and school factors related

to student achievement over time, they were not particularly well suited for studying teacher and classroom effects. In part, this was due to the nature of the sampling frame that was used, in which relatively small samples of students were selected within schools. Future designs should sample intact classrooms and develop better measures of classroom practices to focus on teacher and classroom effects (Mullens & Gayler, 1999).

looseness1 Another challenge is to encourage researchers to develop and use more comprehensive conceptual frameworks for their studies of school effectiveness. For example, although economists routinely examine resource variables in their studies of school effectiveness, sociologists and educational researchers frequently do not. Conversely, economists often ignore important process variables in their models, such as school climate. To the extent that models are misspecified at any level of analysis, the resulting estimates can be biased and the conclusions faulty (Raudenbush & Byrk, 2002, chap. 9). A similar argument can be made regarding outcome measures: School effectiveness studies focus

predominately on student achievement as measured by test scores, thereby ignoring outcomes, such as dropout or attrition, that can be influenced by different factors and could lead to different conclusions about effective schools (Rumberger & Palardy, 2003a).

The final challenge is to encourage better use of the growing advances in statistical modeling techniques in school effectiveness studies. Although statistical advances in multilevel and structural equation modeling have been quite rapid, these advances are slow to find their way into mainstream school effectiveness studies. Although there is always a lag between the initial development of new statistical techniques and their widespread use in the field, as the techniques become more sophisticated, that lag could increase. This may be particularly problematic for existing scholars who were most likely trained in earlier techniques and who will require a sort of in-service training to learn the new approaches. Fortunately, many professional associations, such as the American Educational Research Association and the American Sociological Association, sponsor such training sessions in conjunction with their national meetings each year.

**Appendix:** Variable Descriptive Statistics and Labels for NELS Data

<i>Variable Name</i>	<i>M</i>	<i>SD</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Description and (NELS:88 Variables)</i>
<b>Measurement variables (n = 39, 241)</b>					
Math	50.31	10.28	23.34	80.67	Math IRT theta score (BY2XRTH, F12XRTH, F2XRTH)
Time	0.46	0.40	0.00	1.00	Time (0 = 8th grade; 0.5 = 10th grade; 1 = 12th grade)
<b>Student variables (n = 14, 199)</b>					
Math, Grade 10	51.11	9.88	24.87	72.90	Math IRT theta score (F12XRTH)
SES	0.04	0.81	-2.95	2.75	10th-grade SES composite (F1SES)
Transfer	0.06	0.24	0.00	1.00	Transferred schools between 10th and 12th grades (F2F1SCFG = 1)
Dropout	0.07	0.25	0.00	1.00	Dropped out of school (F2DOSTAT = 3, 4, 5)
<b>School variables (n = 912)</b>					
Mean SES	0.01	0.52	-1.33	1.54	Mean SES of students (F1SES)
Catholic	0.07	0.25	0.00	1.00	(G10CTRL1 = 2)
Private	0.08	0.27	0.00	1.00	(G10CTRL1 = 3-5)
Homework time	4.61	2.05	1.06	14.00	Mean number of hours spent on homework per week (F1S36A2)
NAEP composite	13.76	2.27	6.00	27.74	Number of NAEP units in math, science, English, and social science earned in high school (F2ra11.C + a12.C + geo.C, tri.C + pre.C + cal.C + bio.C + che.C + phy.C + soc.C + his.C)

NOTE: NELS = National Education Longitudinal Study; SES = socioeconomic status; IRT = item response theory; NAEP = National Assessment of Educational Progress.

## REFERENCES

- Alexander, K. L., & Eckland, B. K. (1975). Basic attainment processes. *Sociology of Education*, 48, 457–495.
- Alexander, K. L., & Pallas, A. (1985). School sector and cognitive performance: When is a little a little? *Sociology of Education*, 58, 115–128.
- Barr, R., & Dreeben, R. (1983). *How schools work*. Chicago: University of Chicago Press.
- Bryk, A. S., Lee, V. E., & Holland, P. B. (1993). *Catholic schools and the common good*. Cambridge, MA: Harvard University Press.
- Bryk, A. S., & Thum, Y. M. (1989). The effects of high school organization on dropping out: An exploratory investigation. *American Educational Research Journal*, 26, 353–383.
- Carroll, D. (1996). *National Education Longitudinal Study (NELS:88/94): Methodology report*. Washington, DC: Government Printing Office.
- Chubb, J. E., & Moe, T. M. (1990). *Politics, markets, and America's schools*. Washington, DC: Brookings Institution.
- Coleman, J. S. (1990). *Equality and achievement in education*. San Francisco: Westview.
- Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, F., Weinfeld, F., et al. (1966). *Equality of educational opportunity*. Washington, DC: Government Printing Office.
- Coleman, J. S., & Hoffer, T. (1987). *Public and private high schools: The impact of communities*. New York: Basic Books.
- Coleman, J. S., Hoffer, T., & Kilgore, S. B. (1982). *High school achievement: Public, Catholic, and private schools compared*. New York: Basic Books.
- Croninger, R. G., & Lee, V. E. (2001). Social capital and dropping out of high school: Benefits to at-risk students of teachers' support and guidance. *Teachers College Record*, 103, 548–581.
- Gamoran, A. (1992). Social factors in education. In M. C. Alkin (Ed.), *Encyclopedia of educational research* (pp. 1222–1229). New York: Macmillan.
- Gamoran, A. (1996). Student achievement in public magnet, public comprehensive, and private city high schools. *Educational Evaluation and Policy Analysis*, 18, 1–18.
- Goldhaber, D. D., & Brewer, D. J. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources*, 33, 505–523.
- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24, 1141–1177.
- Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher*, 18, 45–62.
- Hauser, R. M., & Featherman, D. L. (1977). *The process of stratification: Trends and analysis*. New York: Academic Press.
- Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher*, 23, 5–14.
- Jencks, C. S., & Brown, M. D. (1975). Effects of high schools on their students. *Harvard Educational Review*, 45, 273–324.
- Jencks, C., & Mayer, S. E. (1990). The social consequences of growing up in a poor neighborhood. In L. Lynn Jr. & M. G. H. McGeary (Eds.), *Inner-city poverty in the United States* (pp. 111–186). Washington, DC: National Academy Press.
- Johnson, M. K., Crosnoe, R., & Elder, G. H., Jr. (2001). Students' attachment and academic engagement: The role of race and ethnicity. *Sociology of Education*, 74, 318–340.
- Kahlenberg, R. D. (2001). *All together now: Creating middle-class schools through public school choice*. Washington, DC: Brookings Institution.
- Kaplan, D., & Elliott, P. R. (1997). A model-based approach to validating education indicators using multilevel structural equation modeling. *Journal of Educational and Behavioral Statistics*, 22, 323–347.
- Kaplan, D., & Kreisman, M. B. (2000). On the validation of indicators of mathematics education using TIMSS: An application of multilevel covariance structure modeling. *International Journal of Educational Policy, Research, and Practice*, 1, 217–242.
- Lee, V. E., & Burkam, D. T. (2003). Dropping out of high school: The role of school organization and structure. *American Educational Research Journal*, 40, 353–393.
- Lee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, 62, 172–192.
- Lee, V. E., & Smith, J. B. (1993). Effects of school restructuring on the achievement and engagement of middle-grade students. *Sociology of Education*, 66, 164–187.
- Lee, V. E., & Smith, J. B. (1995). Effects of high school restructuring and size on gains in achievement and engagement for early secondary school students. *Sociology of Education*, 68, 241–279.
- Lee, V. E., & Smith, J. B. (1997). High school size: Which works best for whom? *Educational Evaluation and Policy Analysis*, 19, 205–227.
- Lee, V. E., & Smith, J. B. (1999). Social support and achievement for young adolescents in Chicago: The role of school academic press. *American Educational Research Journal*, 36, 907–945.
- Lee, V. E., Smith, J. B., & Croninger, R. G. (1997). How high school organization influences the equitable distribution of learning in mathematics and science. *Sociology of Education*, 70, 128–150.
- Levin, H. M. (1994). Production functions in education. In T. Husen & T. N. Postlethwaite (Eds.), *International encyclopedia of education* (pp. 4059–4069). New York: Pergamon.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: John Wiley.
- McArdle, J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58, 110–133.
- McNeal, R. B. (1997). High school dropouts: A closer examination of school effects. *Social Science Quarterly*, 78, 209–222.
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107–122.
- Mosteller, F., & Moynihan, D. P. (Eds.). (1972). *On equality of educational opportunity*. New York: Random House.
- Mullens, J. E., & Gayler, K. (1999). *Measuring classroom instructional processes: Using survey and case study field-test results to improve item construction* (Working Paper No. 1999–08). Washington, DC: National Center for Education Statistics.

- Murnane, R. J. (1981). Interpreting the evidence on school effectiveness. *Teachers College Record*, 83, 19–35.
- Muthén, B. (1989). Latent variable modeling in heterogeneous populations: Presidential address to the Psychometric Society. *Psychometrika*, 54, 557–585.
- Muthén, B. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, 28, 338–354.
- Muthén, B. (1997). Latent variable modeling with longitudinal and multilevel data. In A. Raftery (Ed.), *Sociological methodology* (pp. 453–480). Boston: Blackwell.
- Palardy, G. J. (2003). *A comparison of hierarchical linear and multilevel structural equation growth models and their application in school effectiveness research*. Unpublished doctoral dissertation, University of California, Santa Barbara.
- Phillips, M. (1997). What makes schools effective? A comparison of the relationships of communitarian climate and academic climate to mathematics achievement and attendance during middle school. *American Educational Research Journal*, 34, 633–662.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, 20, 307–335.
- Rumberger, R. W. (1995). Dropping out of middle school: A multilevel analysis of students and schools. *American Educational Research Journal*, 32, 583–625.
- Rumberger, R. W., & Palardy, G. J. (2003a). *Does segregation (still) matter? The impact of student composition on academic achievement in high school*. Revised paper originally presented at the annual meeting of the American Educational Research Association, April 10–14, 2001, Seattle, WA.
- Rumberger, R. W., & Palardy, G. J. (2003b). *Test scores, dropout rates, and transfer rates as alternative measures of school performance*. Revised paper originally presented at the annual meeting of the American Educational Research Association, April 1–5, 2002, New Orleans, LA.
- Rumberger, R. W., & Thomas, S. L. (2000). The distribution of dropout and turnover rates among urban and suburban high schools. *Sociology of Education*, 73, 39–67.
- Rumberger, R. W., & Willms, J. D. (1992). The impact of racial and ethnic segregation on the achievement gap in California high schools. *Educational Evaluation and Policy Analysis*, 14, 377–396.
- Seltzer, M., Choi, K., & Thum, Y. M. (2003). Examining relationships between where students start and how rapidly they progress: Using new developments in growth modeling to gain insight into the distribution of achievement within schools. *Educational Evaluation and Policy Analysis*, 25, 263–286.
- Shavelson, R., McDonnell, L., Oakes, J., & Carey, N. (1987). *Indicator systems for monitoring mathematics and science education*. Santa Monica, CA: RAND.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.
- Summers, A. A., & Wolfe, B. L. (1977). Do schools make a difference? *American Economic Review*, 67, 639–652.
- Tagiuri, R. (1968). The concept of organizational climate. In R. Tagiuri & G. H. Litwin (Eds.), *Organizational climate: Exploration of a concept* (pp. 1–32). Boston: Harvard University, Division of Research, Graduate School of Business Administration.
- U.S. Department of Education, National Center for Education Statistics. (2003). *The condition of education, 2003* (NCES 2003-67). Washington, DC: Government Printing Office.
- Willms, J. D. (1985). Catholic-school effects on academic achievement: New evidence from the High School and Beyond follow-up study. *Sociology of Education*, 59, 98–114.
- Willms, J. D. (1992). *Monitoring school performance: A guide for educators*. Washington, DC: Falmer.
- Willms, J. D., & Raudenbush, S. W. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement*, 26, 209–232.
- Witte, J. F. (1992). Private school versus public school achievement: Are there findings that should affect the educational choice debate? *Economics of Education Review*, 11, 371–394.
- Witte, J. F., & Walsh, D. J. (1990). A systematic test of the effective schools models. *Educational Evaluation and Policy Analysis*, 12, 188–212.