Toward a More Appropriate Conceptualization of Research on School Effects: A Three-Level Hierarchical Linear Model
Author(s): Anthony S. Bryk and Stephan W. Raudenbush
Source: *American Journal of Education*, Vol. 97, No. 1 (Nov., 1988), pp. 65-108
Published by: The University of Chicago Press
Stable URL: http://www.jstor.org/stable/1084940
Accessed: 27/05/2009 18:29

Toward a More Appropriate Conceptualization of Research on School Effects: A Three-Level Hierarchical Linear Model

ANTHONY S. BRYK
University of Chicago

STEPHAN W. RAUDENBUSH
Michigan State University

Two aspects of educational research—that children's learning is typically the object of inquiry and that this learning usually occurs in organizational settings of classrooms and schools—locate educational research at the nexus of two of the most troublesome and persistent methodological problems in the social sciences: the measurement of change and the "unit-of-analysis" problems. Over the past several years, a satisfactory solution to these problems has been developed through the use of hierarchical linear models. This article summarizes these developments and shows how these techniques have already been used to illuminate important substantive questions about the effects of school organization and children's language development. In the final section of the article, the authors introduce a three-level hierarchical model that they argue should constitute the basic paradigm for future quantitative research on student learning.

Introduction

The central phenomenon of interest in educational research is the growth in knowledge and skill of individual students. We seek to understand how personal characteristics of students, such as their ability and motivation, and aspects of their individual educational experiences, such as the amount of instruction, influence their academic growth. This learning chiefly takes place in the organizational settings of schools and classrooms. Features of these settings—size, climate,

0036-6773/89/9701-0004$01.00

resources, materials, and the like—can have substantial influences on the learning processes occurring within them. External factors such as district, state, and federal policies can also affect these processes.

In research on instruction, we are concerned about the interaction of students with a teacher around specific curricular materials. These interactions usually occur within a classroom setting and are typically bounded by a single academic year. The general research problem has three foci: to assess the growth of students over the course of the academic year (or segment of a year); to examine the possible effects of student background characteristics and individual educational experiences on this learning; and to determine how the organization of classrooms, the activities of teachers, and policies imposed from outside may influence the distribution of student outcomes within these settings.

Research on school effects has a similar structure. We are still concerned about learning and how this growth is influenced by students' personal backgrounds and educational experiences. The time frame is longer, however, usually spanning several years, and is frequently connected with the organizational division of schools into nursery, elementary, middle, secondary, and postsecondary. The school building is now the principal organizational unit, rather than the classroom, and the effects of features of this organizational unit and its external environment, such as academic policies, disciplinary climate, teacher collegiality, and principal leadership, are all important research interests.

Two aspects of educational research—that children's growth is typically the object of inquiry and that such growth occurs in organizational settings—correspond to two of the most troublesome and persistent methodological problems in the social sciences: the measurement of change and "unit-of-analysis" problems. (For a comprehensive review of the problems in measuring change, see Willett [1988]; on the dif-

---

ANTHONY S. BRYK is an associate professor of education and of the College at the University of Chicago. His research interests include quantitative educational research methods and the social organization of high schools and their effects. STEPHAN W. RAUDENBUSH is an associate professor in the department of educational psychology in the Graduate School of Education at Michigan State University. His specializations include multilevel data analysis and research synthesis. Raudenbush and Bryk, who have been collaborating for the past five years on research on the theory and applications of hierarchical linear models in the behavioral and social sciences, have jointly published a number of technical papers on these developments.

ficulties associated with assessing multilevel effects, see Burstein [1980] or Haney [1980].)

In brief, previous research on individual change has been plagued by inadequacies in conceptualization and design. From a conceptual point of view, a model for the phenomenon under study is an important heuristic for guiding inquiry. Yet researchers studying student learning rarely posit an explicit model of individual growth and how features of schools and classrooms affect this growth (Sorenson and Hallinan 1977). Further, much research on learning is based on student data collected at two time points. Such pretest-posttest designs, however, are generally inadequate for studying individual change (Bryk and Weisberg 1977; Rogosa, Brand, and Zimowski 1982). By implication, this means that they are also inadequate for studying effects on student learning since learning is a process of individual change involving the acquisition of knowledge and skill over time (Sorenson and Hallinan 1977). In principle, research on learning requires multi–time point data and a statistical model that permits an explicit representation of individual growth.

The "unit-of-analysis" problem derives from the fact that much educational research involves propositions about relationships among constructs at different organizational levels. For example, educational sociologists routinely ask questions about the organization of schools and their effects on the individuals within them. Does a smaller class size (a classroom characteristic) foster greater academic learning (an individual student characteristic)? Or what effect does extensive bureaucratic control (a school and district characteristic) have on the sense of efficacy and commitment among teachers (an individual adult characteristic) who work there?

Rarely is the basic organizational structure of schooling, where students are nested within classrooms within schools within districts, and so on, explicitly represented in statistical analyses of educational data. In research on learning, for example, data are routinely analyzed at the student level. It is assumed that educational interventions have a constant effect on all students who are exposed to them, and these effects are invariant across organizational contexts. When effects vary among students and across contexts, however, traditional analytic approaches can produce misleading results even in carefully controlled experimental studies.

Cronbach and Webb (1975), for instance, reanalyzed the data from an important study that had reported an aptitude-by-treatment interaction effect in arithmetic instruction. A key sociological unit, the classroom, was ignored in the original analyses. When Cronbach and Webb decomposed the observed relationship between aptitude and achievement into within- and between-classroom components, they found that most of the interaction effect was between classrooms rather than within them. This reanalysis cast doubt on the original interpretation of the data as evidence of differences among children in learning style. Barr and Dreeben (1983) make a similar claim in examining the extant research on the effects of ability grouping. They argue that ability groups are important organizational units within classrooms, and the failure to represent these units in the analytic models used in ability grouping research has produced seriously flawed results.

Historically, this methodological problem has been posed as a need to choose an appropriate unit of analysis. It has become clear, however, that the "choice of unit" is the wrong question (cf. Rogosa 1978; Burstein 1980). Rather, the real need is for statistical models that provide explicit representation of the multiple organizational levels typically encountered in educational research.

Although the methodological problems of measuring change and unit of analysis have distinct, long-standing, and nonoverlapping literatures, they, in fact, share a common cause—the inadequacy of traditional statistical techniques for the modeling of hierarchy. The mismatch between the hierarchical character of much educational phenomena and traditional statistical methods has plagued research, leading to many spurious inferences (Cronbach 1976). With the recent development of a statistical theory of hierarchical linear models (HLMs), however, the basis for a more appropriate approach now exists.

Over the past several years, a satisfactory solution to measuring change and assessing multilevel effects has been developed using hierarchical linear models. We discuss these applications of hierarchical linear models below and show how these techniques have already been used to illuminate substantive questions about the effects of high school organization on the social distribution of achievement within schools, as well as the effects of maternal speech on children's language development.

Both of these uses of HLMs involve two-level models and are important developments. Nevertheless, as noted above, educational research is typically concerned about situations where problems of measuring change and assessing multilevel effects occur simultaneously. For example, we might seek to study how the organization of reading instruction in elementary schools affects students' learning in these early grades. An adequate approach to questions of this sort requires a three-level model that allows us to combine the features of modeling both individual growth over time and the organizational structure of students nested within different contexts such as classrooms or schools.

The three-level model introduced in this article provides a comprehensive framework for examining the structure of individual learning, for investigating how children's background and individual educational experiences influence the shape of their learning curves, and, further, for exploring how aspects of organizational context may have differential effects on the students within it. The three-level HLM may be viewed as a basic paradigm for future quantitative educational research in that it provides not only tools for analysis but also a general conceptual model for representing explicitly the central features of concern in educational research on student learning. The illustration of a three-level HLM presented in the last section of the article provides a glimpse of the potential that this approach offers for illuminating educational effects on individual development.

Overview of the Hierarchical Linear Model

As noted in the introduction, educational data often have a nested structure of repeated observations on students, students within classrooms within schools, and so on. Each of these levels is formally represented within HLM by its own statistical model. Although hierarchial linear models can represent any number of levels, for simplicity, we begin by considering two-level problems. In our first application, the levels are students nested within schools. In our second application, the levels are repeated observations or time points within individuals.

The statistical theory for HLM has developed out of several streams of methodological work. Significant contributions come from biometric applications of mixed-model ANOVA, from econometrics under the rubric of random coefficient regression models, and, most important, from developments in the statistical theory of covariance component models and Bayesian estimation for linear models. For a comprehensive review of these technical developments see Raudenbush (1988).

In general, the two-level HLM requires specification of two interrelated equations. For clarity of exposition, we describe the basic features of the two-level model in the context of our first application—research on the effects of high school organization on the social distribution of student achievement. In this case, the two-level HLM consists of a *within-school* and a *between-school* model. (The formulation of a two-level HLM for the measurement of change problem is discussed in a subsequent section.)

Formally, in the *within-school* equation, the achievement of student i in school j, Y_{ij}, is represented as a function of student background characteristics, X_{ijk}, and a random error, R_{ij}:

$$Y_{ij} = \beta_{j0} + \beta_{j1}X_{ij1} + \beta_{j2}X_{ij2} + \cdots + \beta_{jK}X_{ijK} + R_{ij} \, . \quad (1)$$

The β_{jk} regression coefficients indicate the strength of associations between student background characteristics and the outcome. For this reason, these regression coefficients indicate how achievement is distributed in school j as a function of measured student characteristics such as race/ethnicity, social class, and academic background. Equation (1) may be viewed as a measurement model of the effects of school j on the students within it. Rather than simply assuming that a school has a constant effect on all of its students, as in conventional school effects analyses, the model allows us to represent different effects for different students through the inclusion of student characteristics, X_{ijk}, in the equation. Formally, we say that the β_{jk} coefficients represent the distributive effects of school j on student achievement. That is, the set of regression coefficients, β_{jk}, summarize how effects are distributed among different types of students within a school. (For a further elaboration of this idea see the next section.)

A distinctive feature of HLM is that these regression coefficients are presumed to vary across schools, and it is this variation that is of particular interest. Therefore, we formulate a set of *between-school* equations that represent each of the k regression parameters, β_{jk}, as a function of school-level variables, W_{pj}, and a unique residual school effect, U_{jk}:

$$\underbrace{\beta_{jk}}_{\substack{\text{school}\\ \text{effects}\\ \text{in unit } j}} = \underbrace{\gamma_{0k} + \gamma_{1k}W_{1j} + \gamma_{2k}W_{2j} + \cdots + \gamma_{pk}W_{pj}}_{\substack{\text{effects of school characteristics on the}\\ \text{distribution of achievement}\\ \text{within schools}}} + \underbrace{U_{jk}}_{\substack{\text{unique effect}\\ \text{associated with}\\ \text{school } j}} \, . \quad (2)$$

Equation (2) models how features of schools affect the distribution of achievement within them. The γ_{pk} coefficients represent the influence of specific school-level factors, such as size, disciplinary climate, and academic organization, on the distribution of school effects among students. For example, suppose that β_{j1} is the regression coefficient of mathematics achievement on students' social class. The size of this coefficient measures the extent to which initial differences among students' family background are related to subsequent achievement levels within a school. Does the strength of this relationship vary across schools? If so, do certain school characteristics act to attenuate the effects of family background on student outcomes? These are illustrative of the kinds of hypotheses that can be formulated in the *between-school* model.

Statistical Estimation

One obvious difficulty with estimating hierarchical linear models is that the outcome variables in the between-school equations, β_{jk}, are not directly observed. However, they can be estimated using standard regression techniques such as ordinary least squares. But these estimates, $\hat{\beta}_{jk}$, are measured with error, that is,

$$\hat{\beta}_{jk} = \beta_{jk} + e_{jk} . \tag{3}$$

Substituting from equation (3) into equation (2) for β_{jk} yields an equation in which the estimated relation, $\hat{\beta}_{jk}$, varies as a function of measurable school characteristics and an error term equal to $U_{jk} + e_{jk}$.

$$\hat{\beta}_{jk} = \gamma_{0k} + \gamma_{1k}W_{1j} + \gamma_{2k}W_{2j} + \cdots + \gamma_{pk}W_{pj} + U_{jk} + e_{jk} . \tag{4}$$

Equation (4) resembles a conventional linear model except that the structure of the error term is more complex. A consequence of this more complex error term is that neither the γ coefficients nor the covariance structure among the errors can be appropriately estimated with conventional linear model methods. However, recent breakthroughs in statistical theory and computation now make this estimation possible (see Raudenbush [1988] for a full review).

From a technical point of view, estimation of the γ coefficients in equation (4) can be viewed as a generalized or weighted least squares regression problem where the weighting factor involves the covariance structure among the errors, $U_{jk} + e_{jk}$. Maximum likelihood estimation of these covariance structures can be obtained using iterative computational techniques. As a result, efficient estimates for the γ's are also available.

The statistical estimates generated by this procedure have several important properties. First, the precision of the $\hat{\beta}_{jk}$ coefficients will vary across schools because the amount of data available in each school will generally vary. In estimating the γ coefficients, HLM methods weight the contribution of the $\hat{\beta}_{jk}$ from each school proportional to their precision. This optimal weighting procedure minimizes the effects of unreliability in the $\hat{\beta}_{jk}$ on inferences about model parameters.

Second, the estimation procedures are fully multivariate since they take into account the covariation among the β coefficients. To the extent that these parameters covary, estimation will be more precise.

Third, HLM estimation enables the investigator to distinguish between variation in the true school effects, β_{jk}, and the sampling variation that arises because $\hat{\beta}_{jk}$ measures β_{jk} with error. That is, from equation (3),

$$\mathrm{var}(\hat{\beta}_{jk}) = \mathrm{var}(\beta_{jk}) + \mathrm{var}(e_{jk}) \ . \tag{5}$$

Or,

total observed variance = parameter variance + sampling variance .

The HLM applications that follow demonstrate the substantive importance of this variance decomposition. Knowledge of the amount of parameter variability is important in the process of formulating HLMs and in evaluating their results.

Fourth, the method enables estimation of the covariation among the β's, that is, "parameter covariance," which also can be of substantive interest. For example, in applications of HLM to measurement of change problems, estimated parameter variances and covariances provide the basis for a maximum likelihood estimate of the correlation of change with initial status. This correlation, an important parameter in research on change, is routinely misestimated with conventional techniques.

Finally, the method enables improved estimates of the β coefficients in each school. Hierarchical linear models borrow strength from the fact that the estimation of β is being repeated across a number of schools. Although this advantage plays no role in applications presented in this article, it has been used effectively in other published research. For example, Braun et al. (1983) have capitalized on this feature to improve the validity of the prediction equations used by graduate business schools in selecting minority applicants for admissions. This feature of HLM is also of value in studies seeking to identify unusually effective schools (Raudenbush and Bryk 1988).

Research on School Effects and Instruction

A number of conceptual and technical difficulties have plagued the past analysis of multilevel data encountered in school effects research. Among the most commonly encountered difficulties have been aggregation bias, misestimated standard errors, and heterogeneity of regression. Raudenbush and Bryk (1986) explain how HLMs resolve each of these problems.

In brief, aggregation bias can occur when a variable takes on different meanings and therefore may have different effects at different organizational levels. For example, at the student level, social class provides a measure of the intellectual and tangible resources in a child's home environment. At the school level, it is a proxy measure of a school's resources and normative environment. Clearly, the average social class of a school may have an effect on student achievement above and beyond the effect of the individual child's social class. Hierarchical linear modeling resolves the confounding of these two effects by facilitating a decomposition of any observed relationship among variables, such as achievement and social class, into separate between-school and within-school components.

As noted earlier, this decomposition can be critical to correct interpretation of empirical relationships. For example, Barr and Dreeben (1983) found that, although there is a strong association between ability and reading achievement, almost all of this relationship is between reading groups. The relationship within groups is almost zero. This important finding supported their contention that the organization of reading instruction through ability grouping has extensive effects on student learning.

Misestimated standard errors occur in multilevel data when investigators fail to take into account the dependence among the outcome of students who attend the same school. This dependence arises because of the shared experiences among students within a school and because of the ways students are assigned to schools. This same problem occurs in survey research whenever a cluster sample is employed. For example, a sample of 1,000 students drawn purely at random will provide a more precise base for statistical estimates than would a sample of 1,000 students who are drawn through a two-stage procedure of first selecting a set of schools at random and then drawing students at random from within them.

Hierarchical linear modeling resolves this problem by incorporating the unique effects of individual schools into the statistical model for the outcome. The variability in these unique school effects is included in the standard error estimates generated under HLM. In the terminology of survey research, HLM estimates adjust for the intraclass correlation (or related to it the design effect) that results from cluster sampling.

Heterogeneity of regression occurs when the relationships between student characteristics and outcomes vary across schools. Although this phenomenon has often been viewed as a methodological nuisance, the causes of such heterogeneity of regression are in fact of much substantive interest. In research on school organization, for example, Bidwell and Kasarda (1980) and Barr and Dreeben (1983) have argued that school policies produce differential learning opportunities for the students within them. Since these vary across schools, heterogeneity of regression is a likely empirical consequence.

Hierarchical linear modeling addresses this problem by enabling the investigator to estimate a separate multiple regression equation for each school and then to model variation among schools in these regression coefficients as a multivariate outcome to be explained by school-level factors.

A School Effects Application: The Social Distribution of Achievement in Public and Catholic High Schools

The HLM application presented in this section is part of a larger investigation (Bryk, Holland, and Lee, in preparation) of the internal organization of Catholic high schools and the effects associated with this organization. The research was motivated by findings, first reported by Coleman, Hoffer, and Kilgore (1982) and Greeley (1982), which indicated that academic achievement had a more equitable social distribution in Catholic schools than in the public sector. Field research (Bryk et al. 1984) suggested that the academic organization and normative environment of Catholic high schools might play a role in achieving this distribution of outcomes.

The validity of the Coleman et al. claim, however, has been subject to methodological dispute (see, e.g., Willms 1984; also Alexander and Pallas 1983). The statistical evidence offered by Coleman et al. was based on separate student-level regressions in the public and Catholic sectors. It was possible, however, that the observed weaker relationships between family background and student achievement in the Catholic sector was mostly between schools, a phenomenon that would go undetected in their student-level analysis. If true, this would greatly weaken the claim of Coleman et al. that "Catholic schools more closely approxiamte the common-school ideal" (p. 177). Thus, an HLM reanalysis of the data also afforded an important new test of the common-school hypothesis.

Data.—We employed the High School and Beyond (HS&B) data, a general purpose survey of American high schools, in this investigation. In the base year (1980) questionnaires and standardized achievement tests were administered to a stratified random sample of approximately 30,000 sophomores and 30,000 seniors in over 1,000 high schools. The sophomores were resurveyed 2 years later. The analyses described below are based on a composite subsample from both the sophomore and senior cohorts. Only students who completed high school were included. All Catholic high schools ($N = 83$) and a random subsample of public high schools ($N = 94$) were selected for this investigation. The follow-up data on the sophomore cohort (i.e., their responses at senior year) and the baseline data from the senior cohort were combined in order to increase the effective sample size within each school. The student samples per school ranged from 10 to 70, although samples of less than 45 were rare. The outcome variable in these analyses is a standardized mathematics achievement score. Since the mathematics tests employed in 1980 and 1982 were not identical, the results from the sophomore and senior cohort were equated using item response theory in order to make maximum use of the available student information. Because of missing data at the school level, the final sample size was reduced to 160 schools.

The HLM program (Bryk et al. 1986) was used to partition the total variance in mathematics achievement into its within- and between-school components. These variance components were estimated by fitting an HLM model where only a base or constant term is specified for the within-school equation, and the between-school equation just represents the variability among schools in these coefficients:

$$y_{ij} = \beta_{j0} + R_{ij} ,$$

and

$$\beta_{j0} = \mu + U_j . \tag{6}$$

In this HLM, the base coefficient, β_{j0}, is just the mean achievement in school j. Equation (6) is a standard one-way random effects ANOVA model where schools are a random factor with varying numbers of students in each school sample. The pooled within-school variance, var(R_{ij}), which is the variability in student achievement around their respective school means, was estimated as 40.273. The between-school variance, which is the variability among the school means (var(U_j) = var(β_{j0}) in this case), was estimated as 10.525. As a result, the intraschool correlation, which is the ratio of the between-school variance over the sum of the between- and within-school variances, was .207. This means that approximately 21 percent of variance in mathematics achievement is between schools.

A within-school model for the social distribution of achievement.—The social distribution of mathematics achievement in each school was represented by a within-school model that regressed mathematics achievement on minority status (MINORITY), social class (SES), and academic background (ACDBKGD):

$$\text{MATHACH}_{ij} = \beta_{j0} + \beta_{j1}\text{MINORITY} + \beta_{j2}\text{SES} + \beta_{j3}\text{ACDBKGD} + R_{ij} . \qquad (7)$$

Each school's distribution of achievement is now characterized in terms of four parameters: an intercept and three regression coefficients. The SES and ACDBKGD variables were centered around their respective school means. As a result, the four parameters can be interpreted as follows:

β_{j0} = "base" achievement in school j, which is the mean achievement for white students of school average social class and academic background;
β_{j1} = the "minority gap" in school j (i.e., the mean difference between the achievement of white and minority students);
β_{j2} = the differentiating effect of social class in school j (i.e., the degree to which social class differences among students relate to senior year achievement);
β_{j3} = the differentiating effect of academic background in school j (i.e., the degree to which entry differences in the academic background of students eventuate in senior year achievement differences).

Under this model, an effective school in equalizing the distribution of achievement would be characterized by a high base level of achievement, a small minority gap (these coefficients are usually negative), and weak differentiating effects with regard to class and academic background. In these schools the overall achievement is high, and initial differences play a relatively minor role.

The unconditional model.—The first step in an HLM analysis involved fitting a random regression coefficient model. In this "unconditional" model, the regression coefficients, β_{jk}, are allowed to vary randomly across schools, but no school-level predictors are included in the between-school equations. That is, for each regression coefficient in the within-unit model, β_{jk}, the between-school equation was simply

$$\beta_{jk} = \mu_k + U_{jk} \quad \text{for } k = 0, 1, 2, 3 . \qquad (8)$$

Table 1 presents these results. The estimated coefficients for μ_k in equation (8) provide the mean regression equation for the 160 schools.

TABLE 1

Unconditional Model for the Social Distribution of Mathematics Achievement

Means of Fixed Effects	Estimated Coefficient	Standard Error	t-Ratio
Base achievement, μ_0	12.776	.261	48.952
Minority gap, μ_1	−3.053	.317	−9.630
SES differentiation, μ_2	1.197	.112	10.679
Academic differentiation, μ_3	2.490	.097	25.513

Random Parameter	Estimated Parameter Variance	Degrees of Freedom	χ^2	P-Value
Base achievement, β_0	6.595	137	955.8	.000
Minority gap, β_1	1.228	137	162.7	.079
SES differentiation, β_2	.380	137	172.1	.019
Academic differentiation, β_3	.500	137	219.0	.000

Correlations among School-Level Random Effects	β_0	β_1	β_2
Base achievement, β_0			
Minority gap, β_1	+.359		
SES differentiation, β_2	.021	.095	
Academic differentiation, β_3	.161	−.120	.652

	Reliability of School-Level Random Effects
Base achievement, β_0	.694
Minority gap, β_1	.089
SES differentiation, β_2	.175
Academic differentiation, β_3	.332

NOTE.—All estimates for two-level models reported in this article were computed using the HLM program (Bryk et al. 1986).

The average base achievement (for white students) is 12.78 test-score points. The average minority gap is 3.05 (i.e., the estimated MINORITY coefficient is −3.05). The average social class differentiation effect (SES slope) is 1.20, and the average academic background differentiation effect is 2.49 (ACDBKGD slope). All of the mean regression coefficients are statistically significant. These results indicate that minority status, social class, and academic background are related to achievement differences *within* schools.

The unconditional model is particularly valuable because it provides estimates of the *total* parameter variances and covariances among the distributive effects (i.e., β_{jk}). Expressed as correlations (see table 1), they describe the general structure among these school effects. A high base level of achievement is associated with a smaller minority gap (r = .359) and to some extent a greater differentiation with regard to academic background (r = .161). There is also a substantial positive association of schools' social and academic differentiating effects. The estimated correlation between β_{2j} and β_{3j} of .652 suggests that these differentiating effects may share some common structural causes.

By comparing the estimated parameter variance in each regression coefficient, var(β_{jk}), to the total variance in the ordinary least squares estimates (see eq. [5]), HLMs derive an indicator of the reliabilities of the random effects (see Raudenbush and Bryk 1986). These results are also displayed in table 1. The base achievement estimates are quite reliable, .694. As expected, the regression coefficients are less reliable ranging from a low of .089 for minority gap to a high of .332 for academic background differentiating effects. The latter means that much of the observed variability among schools in regression slopes is due to sampling error that, of course, cannot be explained by school-level factors.

Finally, the results of the homogeneity of variance tests provide statistical evidence of significant variation among schools in each of the four random regression coefficients. The probability of the observed variability in these coefficients, under a homogeneity hypothesis, is less than .001 for β_{0j}, β_{1j}, and β_{3j} and less than .02 for β_{2j}. Hence, it appears that schools vary significantly in the degree to which achievement depends on the child's socioeconomic status and academic background. Although the results of the homogeneity test for the minority gap coefficients are only marginally significant (p = .079), these coefficients were treated as random because of previously reported sector effect on minority achievement.

Sector effects.—Having discovered heterogeneity in the social distribution of achievement, the next step is to attempt to explain this heterogeneity as a function of school characteristics. The investigators first examined the "common school" hypothesis claimed by Coleman et al. This involved introducing SECTOR (−1 = public/1 = Catholic) into the between-unit equations for each of the school effect parameters. Average social class (AVSES), minority concentration (HIMNRTY: −1 = <40 percent/1 = 40 percent), and average academic background (AVACBGD) were also included in the model to adjust for possible compositional effects.

In general, this HLM analysis reaffirmed the "common school" hypothesis articulated by Coleman et al. The base mathematics achievement, β_{j0}, was higher in Catholic schools, with the greatest differences occurring in low SES schools. The minority achievement gap, β_{j1}, was smaller in the Catholic sector, with the size of the gap unrelated to minority concentration in either sector. The social class distributive effect, β_{j2}, was stronger in the public sector, and this was particularly true for high SES schools where mathematics achievement is distributed in a very disequalizing fashion. With regard to academic background differentiation, β_{j3}, however, there was no evidence of context, sector, or sector-by-context effects.

Explanatory model.—Next, the investigators sought to identify specific features of schools' academic organization and normative environment that contribute to the social distribution of achievement. The list of school variables considered is summarized in the Appendix. The explanatory variables were grouped into five categories: first, measures of the social and academic context of schools, then three categories representing different aspects of schools' normative environment (perceived teacher quality and interest in students, academic press in the school, and disciplinary climate), and, last, measures of the schools' academic organization.

The first empirical test of the importance of these variables in explaining the "common school" effect is whether they account for parameter variance in the within-school regression coefficients. A more restrictive test involves determining whether the variables actually explain away the "common school" effect. That is, after we have added these variables to the model, does the Catholic advantage still persist? If the more equalized social distribution of achievement in the Catholic sector results from differences in academic organization and normative environment, we would expect the sector effects to disappear once these characteristics are introduced into the model. Finally, the most rigorous test of the explanation involves acceptance of the homogeneity of residual variance hypotheses. That is, after modeling each β_{jk} as a function of some school-level variables, is there evidence of residual parameter variation in the β_{jk} that remains unaccounted? A nonsignificant test statistic would be consistent with the hypothesis that the remaining variance in β_{jk} is just sampling error.

Table 2 presents the results of the final fitted model. The base achievement differences between sectors disappear once we take into account the number of math courses taken (AVMTHEMP), the average amount of homework done by students (AVHMEWRK), and principals' reports about problems with staff (STFPBLM). The direction of effects

TABLE 2

Explanatory Model: Social Distribution of Mathematics Achievement

Fixed Effects	Estimated Coefficient	Standard Error	t-Ratio
Base achievement, β_0:			
BASE	8.431	.902	9.343
AVSES	2.656	.541	4.910
SECTOR	.143	.264	.542
AVMTHEMP	1.489	.335	4.444
AVHMEWRK	.311	.137	2.254
STFPBLM	−.245	.121	−2.029
Minority gap, β_1:			
BASE	−3.108	.323	−9.610
SECTOR	−.144	.441	−.327
DISCLIM	−.971	.430	−2.258
STFPBLM	.165	.176	.938
SES differentiation, β_2:			
BASE	2.421	1.392	1.738
AVSES	−.052	.332	−.157
SECXSES	−.562	.328	−1.713
SECTOR	−.113	.178	−.633
SDMTHEMP	.839	.419	2.003
AUTHRTY	−.958	.428	−2.237

SIZE (in 100's)	.060	.027	2.176
STFPBLM	−.194	.067	−2.882
Academic differentiation, β_3:			
BASE	5.129	1.197	4.283
SECTOR	.547	.129	4.237
SDMTHEMP	.987	.353	2.794
AUTHRTY	−1.699	.378	−4.488
ATTACAD	.178	.068	2.624
STFPBLM	−.096	.056	−1.709

Random Parameter	Estimated Parameter Variance	Degrees of Freedom	χ^2	*P*-Value
Base achievement, β_0	2.380	132	626.3	.000
Minority gap, β_1	.637	134	146.3	.221
SES differentiation, β_2	.153	130	145.5	.167
Academic differentiation, β_3	.334	131	190.5	.000

	PERCENTAGE OF VARIANCE EXPLAINED							
	Base Achievement		Minority Gap		*SES* Differentiation		Academic Differentiation	
MODEL	var(β_0)	$\%R^2$	var(β_1)	$\%R^2$	var(β_2)	$\%R^2$	var(β_3)	$\%R^2$
Unconditional	6.595	. . .	1.228	. . .	.380	. . .	.500	. . .
Explanatory model	1.911	71.0	.346	71.8	.153	59.7	.407	18.3

for these school variables are predictable. Greater math course work, more homework, and fewer staff problems are all associated with higher levels of mathematics achievement. All of these characteristics are more prevalent in the Catholic sector (Bryk et al. 1984).

The sector effect on the minority gap disappears once we take into account disciplinary climate (DISCLIM). The minority gap is largest in schools where there is a high incidence of disciplinary problems. The smaller minority gap in Catholic schools appears linked to the fact that these schools have a more orderly and less disruptive environment.

Table 2 provides strong evidence that the academic organization of high schools plays a central role in converting initial differences in social class and academic background into differences in academic achievement. Greater diversity in math course taking (SDMTHEMP) and larger school size (SIZE) are both associated with a more disequalizing distribution of achievement in schools along class and academic background lines. Schools where students perceive discipline to be fair and effective (AUTHRTY) are less differentiating environments. Positive student attitudes toward academic success (ATTACAD), however, are associated with a more academically differentiated distribution of outcomes. The overall sector effect on SES differentiation disappeared, and the sector-by-school SES interaction effect is no longer statistically significant. A small (in size) but significant sector effect has appeared for academic background differentiation. Catholic schools are somewhat more differentiating with regard to academic background than we would expect given their favorable organizational characteristics (smaller size, less differentiation in course taking, fewer staff problems and more adult authority).

The pattern of effects for staff problems (STFPBLM) across the four school-effect indicators merits comment. Schools with a high incidence of staff problems are in fact equalizing in the distribution of academic achievement—everyone tends to do poorly. The base level is reduced. The achievement among white and black students is more similar. The achievement of socially and academically advantaged children is more like that of their disadvantaged counterparts. If we interpret principals' reports about staff problems as an indicator of the vitality of the social relations among the adults within a school, these HLM results indicate that all children suffer when the internal organization of a school begins to break down.

The last panel in table 2 compares the residual parameter variances, from the explanatory model to the total parameter variances estimated in the unconditional model. The proportion reduction in these parameter variances, percent of R^2, can be interpreted as an indicator of the power of the explanatory model. The fitted model accounts for a substantial percentage of the variance in base achievement (71.0 percent), in the minority gap (71.8 percent), and in the differentiating effects of SES (59.7 percent). School differences on academic differentiation, however, are not as well explained (18.3 percent).

There is evidence of significant residual variation in base achievement and academic differentiation. The homogeneity hypothesis for the minority gap (p = .221) and social differentiation, however, are now sustained (p = .167).

In sum, this HLM analysis reaffirms the "common school" hypothesis of Coleman et al. More important from the perspective of educational theory and practice, however, is that it also provides empirical support for the contention that the academic organization and normative environment of high schools have a substantial impact on the the social distribution of achievement within them. Sector differences appear readily explainable. There are fewer curricular choices in the Catholic sector with schools taking a more proactive stance toward the issue of what students should be taught. These schools are safe, orderly environments conducive to learning. A relatively small school size also facilitates these outcomes. For a further discussion of this HLM application, see Lee and Bryk (1988).

Research on Individual Change and Correlates of Change

Finding adequate measures of individual change and valid techniques for research on change are problems that have long perplexed behavioral scientists. Many concerns cataloged by Harris in 1963 still trouble quantitative studies of growth. These methodological problems have led to a bewildering array of well-intentioned but misdirected suggestions about the design and analysis of research on change. Recent efforts (Rogosa et al. 1982; Rogosa and Willett 1985), however, have gone a long way toward dispelling these misconceptions. Further, as Willett (1988) notes, with the development of HLM techniques, the problems that have plagued past research in this area are now largely resolved.

Formulating HLM for Research on Individual Growth

We assume that y_{it}, the observed status of individual i at time t, is a function of a systematic growth trajectory or growth curve plus random error. It is convenient to assume that systematic growth over time can be represented as a polynomial of degree K. Thus, the within-unit model is

$$y_{it} = \pi_{0i} + \pi_{1i}a_{it} + \pi_{2i}a_{it}^{2} + \cdots + \pi_{Ki}a_{it}^{K} + e_{it} \qquad (9)$$

for $i = 1, \ldots, n$ subjects, each of whom is observed on T_i occasions, where

a_{it} = the age of the subject i at time t, and
π_{ki} = the growth trajacetory parameters for subject i associated with the polynomial of degree K, i.e., $k = 0, \ldots, K$.

Typically, we assume that errors, e_{it}, are normally distributed with a mean of zero and constant variance. The error structure, however, can take on a variety of forms, including the possibility of autocorrelations (Ware 1985; Goldstein 1986; Bryk and Raudenbush 1987).

An important feature of equation (9) is the assumption that the growth parameters vary across individuals. We formulate a *between-person* model to represent this variation. Specifically, for each of the individual growth parameters,

$$\pi_{ki} = \beta_{k0} + \beta_{k1}X_{k1i} + \beta_{k2}X_{k2i} + \cdots + \beta_{kp}X_{kpi} + U_{ki}\ , \qquad (10)$$

where

X_{kpi} = either measured characteristics of the individual's background (e.g., sex or social class) or of the instructional setting (e.g., type of curriculum employed, amount of instruction or some experimental treatment);
β_{kp} = the effect of X_p on the kth growth parameter; and
U_{ki} = the random effects with full covariance matrix, T, dimensioned $(K + 1 \times K + 1)$.

Bryk and Raudenbush (1987) demonstrate how this model can be applied in research on change to describe the structure of the mean growth trajectory, to estimate the extent and character of individual variation around mean growth, to assess the reliability of measures for studying both status and change, to estimate the correlation between subjects' entry status and rate of growth, to estimate correlates of both status and change, to assess the adequacy of between-subject models by estimating the reduction in unexplained parameter variance, and to improve estimates for each individual's growth trajectory and prediction about future individual growth.

A Growth Application: A Study of Children's Early Language Development

We illustrate some of the benefits of an HLM formulation for research on individual change with data from a recent study on children's language development (Huttenlocher et al. 1988). It is hypothesized in the language development literature that the acquisition of words depends both on exposure to appropriate speech from which a child can learn and on innate differences in ability (or capacity) to learn from such exposure. It is widely assumed that differences in capacity are largely responsible for observed differences in children's vocabulary development. The empirical support for this assumption, however, is weak. Heritability studies have found that parent scores on standardized vocabulary tests account for 10 to 20 percent of the variance in children's scores on the same tests. Exposure studies have not fared much better. Newport, Gleitman, and Gleitman (1977) reported negligible effects of maternal speech on children's vocabulary in a study of 1- and 2-year-olds. Clearly, most of the individual variation in vocabulary acquisition remains unexplained.

One of the key strengths of the Huttenlocher et al. study, however, was that it employed a multi–time point design that permitted HLM growth modeling. Prior research, such as the Newport et al. study, has used the more traditional two–time point design.

We report here on an HLM analysis of one sample from the Huttenlocher et al. study consisting of 11 middle-class mother-child pairs (six boys and five girls). Beginning when the child was 14 months of age, the mother-child pairs were observed for 5 hours every 2 months during their typical daily activities. Sessions were tape-recorded and subsequently coded to produce estimates of child's vocabulary size at each age and amount of maternal speech at 16 months.

The observed data and estimated growth trajectories for three cases are depicted in figure 1. The upward curvature in these trajectories indicates that the rate of new word acquisition was increasing over time. This feature was characteristic of the growth trajectories for all 11 children.

Thus, inspection of the data suggested fitting a quadratic individual growth model. Formally, we represent the vocabulary size for child i at age t as

$$\text{VOCABSIZE}_i(t) = \pi_{0i} + \pi_{1i} \cdot [\text{AGE}_i(t) - L] + \pi_{2i} \cdot [\text{AGE}_i(t) - L]^2 + e_i(t) \,. \quad (11)$$

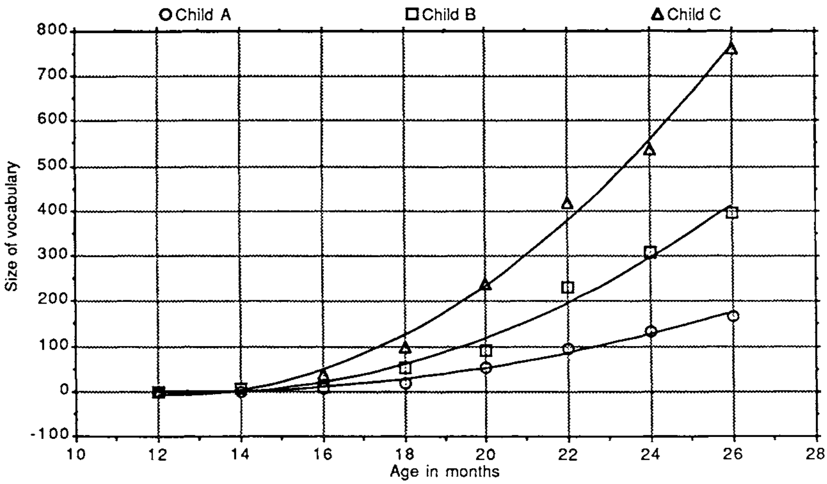

FIG. 1.—A sample of individual vocabulary growth trajectories (Huttenlocher et al. 1988 study). Squares, circles, and triangles represent the actual observations. The smooth curves result from fitting a separate quadratic polynomial to each child's vocabulary data.

Each of the growth parameters in equation (11) has substantive meaning. The intercept, π_{0i}, represents the status of child i at time L. The linear component, π_{1i}, is the instantaneous growth rate for child i at time L, and π_{2i} captures the curvature or acceleration in each child's growth trajectory. While acceleration is a characteristic of the entire trajectory, the status and instantaneous rate parameters depend on the particular choice of the location parameter L. For example, if L is set at 20 months, then π_{0i} and π_{1i} represent the status and instantaneous rate for child i at that particular time point.

To represent the variability in the growth parameters, we pose a second, or between-child model, in which the outcome variables are status, instantaneous rate, and acceleration. The simplest form of this model is

$$\pi_{ki} = \mu_k + U_{ki} \quad \text{for } k = 0, 1, 2, \tag{12}$$

where μ_0 is mean status at age L, μ_1 is mean instantaneous rate of growth at age L, and μ_2 is mean acceleration. The U_{ki} are random effects that represent the deviations in the growth parameters for child i from the respective means. It is assumed that the three random effects are normally distributed with mean 0 and variance-covariance, T.

Equation (12) is another example of an unconditional model. The growth parameters are viewed as random among individuals, but the sources of variation are unknown. The results for this model with L set at 20 months, the median time point in the data set, are presented in table 3.

The t statistics provide strong statistical evidence that each of the parameters in the mean growth trajectory is different from zero. The diagonal elements of T estimate the total parameter variance for each growth parameter. Hierarchical linear modeling provides a large sample chi-square test of the homogeneity hypothesis, that is, $H_0{:}\tau_{kk} = 0$.[1] In each case, the chi-square value was found to be highly significant. Thus, we have statistical evidence of real differences across children in individual growth. Further, each growth parameter is measured with high reliability ranging from .984 and .989 for status and instantaneous growth rate to .774 for acceleration.

The question now becomes, to what extent can differences in the amount of mothers' speech account for the observed differences in children's growth parameters? To examine this question, the investigators introduced the amount of maternal speech as a potential explanatory variable for each of the individual growth parameters. Specifically,

TABLE 3

Hierarchical Linear Model Estimates for the Unconditional Model of Vocabulary Development

Fixed Effects	Estimated Coefficient	Standard Error	t-Ratio
Mean status at 20 months, μ_0	152.36	29.74	5.12
Mean growth rate at 20 months, μ_1	37.67	5.87	6.42
Mean acceleration, μ_2	2.22	.28	7.78

Random Parameter	Estimated Parameter Variance	Degrees of Freedom	χ^2	P-Value
Status, π_0	9,575.29	8	636.5	.000
Growth rate, π_1	375.99	8	876.0	.000
Acceleration, π_2	.69	8	42.1	.000

$$\pi_{0i} = \text{status}_i(L) \quad = \beta_{00} + \beta_{01} \cdot \log(\text{MOMSPEAK})_i + U_{0i},$$
$$\pi_{1i} = \text{rate}_i(L) \quad = \beta_{10} + \beta_{11} \cdot \log(\text{MOMSPEAK})_i + U_{1i}, \qquad (13)$$
$$\pi_{2i} = \text{acceleration}_i = \beta_{20} + \beta_{21} \cdot \log(\text{MOMSPEAK})_i + U_{2i},$$

where log(MOMSPEAK) is a natural logarithm of the amount of maternal speech at 16 months for child i, and β_{01}, β_{11}, and β_{21} relate maternal speech to status at 20 months, instantaneous rate at 20 months, and acceleration, respectively. Because of positive skewness in the distribution of maternal speech, the natural logarithm of maternal speech was used as the independent variable. The U_i are residual random effects—individual differences in growth parameters not accounted for by maternal speech.

Table 4 displays the results of this HLM analysis. The amount of maternal speech, log(MOMSPEAK), is related to each of the individual growth parameters. Maternal speech is associated with status and instantaneous rate at 20 months; it is also associated with acceleration (the speed with which rate of growth increases).

TABLE 4

Hierarchical Linear Model Estimates of the Effects of Maternal Speech on Vocabulary Development

Fixed Effects	Estimated Coefficient	Standard Error	t-Ratio
Effect of log(MOMSPEAK) on status at 20 months, β_{01}	92.02	39.06	2.36
Effect of log(MOMSPEAK) on rate at 20 months, β_{11}	19.70	7.29	2.70
Effect of log(MOMSPEAK) on acceleration, β_{21}	.87	.38	2.30

Random Parameter	Estimated Parameter Variance	Degrees of Freedom	χ^2	P-Value
Status, π_0	6,528.12	8	387.9	.000
Growth rate, π_1	228.35	8	484.0	.000
Acceleration, π_2	.42	8	25.5	.003

The coefficients relating maternal speech to status and instantaneous growth rates can be interpreted as follows: a 10 percent increase in the raw frequency of maternal speech implies an increase in status at 20 months of approximately nine words (10 percent · 92.02) and an increase in rate of growth at 20 months of about two words per month (10 percent · 19.70). Although the coefficient relating maternal speech to acceleration is numerically small (.87), it is substantively significant. The raw frequency of mothers' speech over the 5-hour observation period ranges from approximately 1,000 words to 11,000 words in this sample of middle-class mothers. The estimate of β_{21} indicates that this difference in the amount of maternal speech produces an acceleration difference of approximately four words per month for each month.

The estimated residual parameter variances in table 4 are substantially smaller than the unconditional estimates from table 3. Maternal speech accounts for nearly 32 percent of the total parameter variance in status and 39 percent of the total parameter variance in instantaneous rates at 20 months, and approximately 39 percent of the total parameter variance in acceleration.

By changing the value of the location estimation parameter, *L,* we can examine the relationship of maternal language use to status and instantaneous rates at any time point of interest. These results can be determined either analytically or the HLM model (eqs. [11] and [13]) can be refit for the desired value of *L*. The estimates for β_{01} and β_{11} along with their 2 standard error (SE) bands are displayed in figure 2*a* and *b* for values of *L* of 16, 18, 20, 22, 24, and 26 months.

As early as 16 months of age, we have statistical evidence that maternal speech is significantly related to the rate of new word acquisition. The 2 SE band does not include zero. The strength of this association increases over age as we would expect, given the significant association of log(MOMSPEAK) with acceleration as noted above. A 10 percent increase in the raw frequency of maternal speech implies approximately a one-word-per-month increase in rate of new word acquisition at 16 months (i.e., 10 percent · 12.76), while at 26 months, the same difference in raw frequency of maternal speech implies a three-word-per-month increase in rate (i.e., 10 percent · 30.11).

Further, the amount of maternal speech explains over time an increasing proportion of the variance in instantaneous growth rates. The log(MOMSPEAK) accounts for approximately 30.9 percent of the variance in growth rates at 16 months. By 26 months, the amount of variance in instantaneous growth rate explained by log(MOMSPEAK) increases to 42.1 percent.

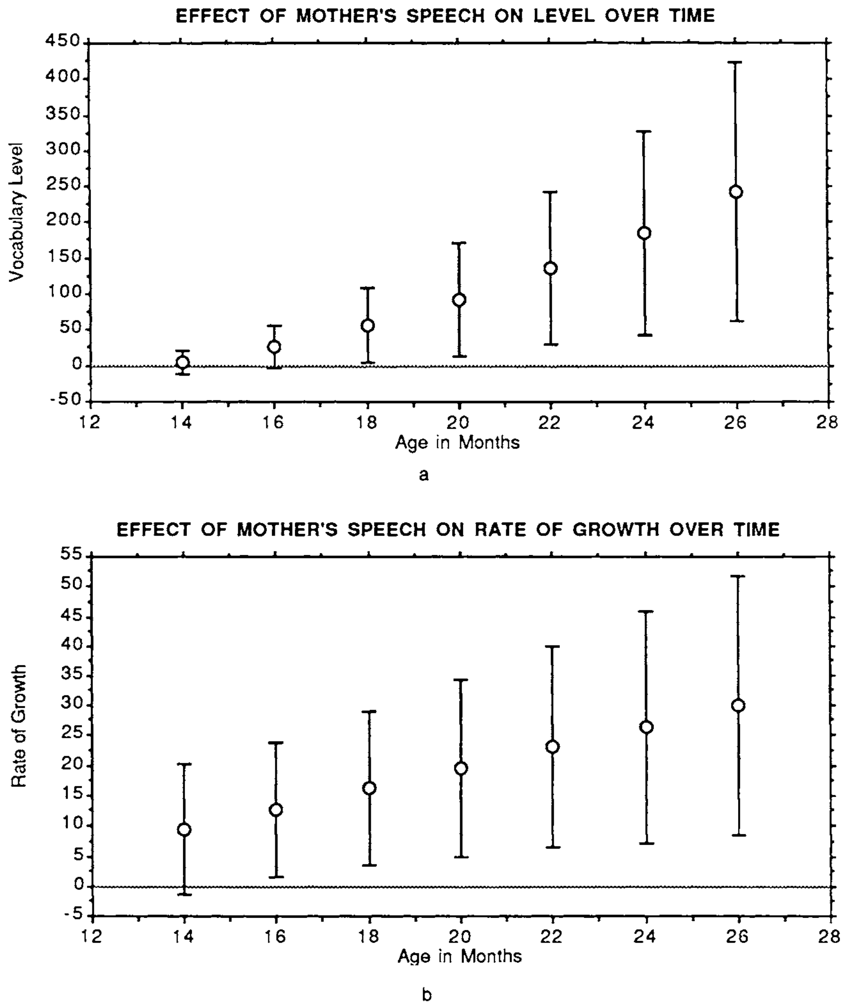

FIG. 2.—*a*, The effects of mother's speech on status differences, β_{01}, at different ages (values of L). *b*, The effects of mother's speech on growth rate differences, β_{11}, at different ages (values of L).

The relationship of log(MOMSPEAK) to children's vocabulary knowledge at time L (i.e., status) follows a similar pattern, but with an important difference—the effects are not detectable as early nor are they quite as strong. While growth-rate effects were detectable at 16 months, it is not until 18 months that the 2 SE confidence band for status effects, β_{01}, does not include zero. This pattern is also reflected in the proportion of variance accounted for by maternal speech. At 16 months, log(MOMSPEAK) explains 21.0 percent of the variance in status. By 26 months, this increases to 38.3 percent. The latter are somewhat weaker than those estimated for the log(MOMSPEAK) effects on growth rates.

These results suggest that maternal speech has a substantial impact on children's early language acquisition. The rate of acquiring new words depends strongly on the amount of exposure to language in the home. As these rate differences persist over time, they eventuate in substantial status differences among children in vocabulary size.

It is instructive to compare the results summarized above with those that would have been obtained had a more conventional two–time point design been employed. Following the basic strategy used by Newport et al. (1977), we fit a linear model where child's vocabulary size at 26 months of age (i.e. status) was regressed on the amount of maternal speech at 16 months while controlling for initial status differences among children, also at 16 months of age. In essence we are discarding the other time points and treating this as conventional two-time-point design with a statistical control for initial differences.

Not surprisingly, we found a highly significant relationship between vocabulary size at these two time points. Initial differences in vocabulary size account for 70 percent of the variability in status differences at 26 months of age. These results would be consistent with an explanation that early vocabulary acquisition is strongly influenced by the individual ability differences reflected in the 16-month scores. The coefficient for log(MOMSPEAK) was not, however, significant at the .05 level, and the incremental variance explained was less than 10 percent.

This reanalysis of the Huttenlocher et al. (1988) data clearly demonstrates the difficulty associated with drawing valid inferences about correlates of growth from conventional analyses of two time-point designs. Had this been the only analysis, Huttenlocher et al. would have concluded, as did Newport et al. (1977), that the effects of maternal speech on vocabulary development were negligible.

The theoretical significance of this empirical evidence linking maternal speech and children's vocabulary development and its implications for the heredity-environment controversy are discussed by Huttenlocher et al. (1988). The arguments here are somewhat complex, and the Huttenlocher et al. conclusions are likely to generate debate. Of primary concern to the purposes of the present article, however, is that the methodological approach used by Huttenlocher et al. enabled the researchers to demonstrate what previous studies have failed to achieve—maternal language use *is* strongly related to the vocabulary growth in young children.

This HLM application also makes a very important research design point. Analyses that employ measures of status at only two time points are incapable of capturing the structure of individual growth or of detecting how personal and environmental factors influence growth characteristics such as rate and acceleration (Bryk and Raudenbush 1987; Rogosa et al. 1982; Goldstein 1986). Analysis of multi–time point data with hierarchical linear models, however, offers a powerful alternative in investigating such phenomena.

Toward a Three-Level Model for Research on School and Instructional Effects

As we stated at the outset of this article, the fundamental phenomenon of interest in educational research is the growth of the individual student learner within the organizational context of classrooms and schools. We have already demonstrated how HLM can be applied separately to research on individual growth and to studies of school effects. We now illustrate how these two research problems can be joined in a three-level HLM that holds much promise for illuminating the basic phenomena of education. This extension is straightforward and simply involves combining the individual growth model and school-effects models illustrated above.

A Simple Application: The Sustaining Effects Study Data

The Sustaining Effects Study (SES) was designed primarily as a longitudinal examination of the effects of compensatory education. However, extensive data were also collected on students in regular educational programs and during the summer as well (Carter 1984). Our analysis is based on a small subsample of these data: 618 students in 86 schools who were in grade 1 during the base year and who were followed longitudinally through the third grade. We used mathematics and reading achievement data from five time points: spring for grade 1, and fall and spring for grades 2 and 3. Test scores were in a vertical scale score metric.

In general, achievement gains during the academic year were considerably greater than those during the summer hiatus. The literature on student gains and losses over the summer is controversial (see, e.g., Pelavin and David 1977; Heyns 1978; Hammond and Frechtling 1979). In addition to conflicting evidence about the magnitude of such effects, questions have also been raised about whether apparent losses during the summer represent a real reduction in learning rates or are just measurement artifacts that result from test scaling. The analyses described below present some new information on this issue.

This three-level HLM application focuses on the relationships of school poverty concentration and child poverty on student learning in reading and mathematics. (For a full investigation of the SES data on this question, see Myers [1986].) We posed the following individual growth model:

$$y_{ijt} = \pi_{0ij} + \pi_{1ij} \cdot t + \pi_{2ij} \cdot \mathrm{SUMDROP}_{ij} + e_{ijt} . \quad (14)$$

The outcome variable y_{ijt} is the mathematics (or reading) achievement of each student i in school j at time t. The time metric, t, was set to zero for the spring grade 1, one for fall grade 2, two for spring grade 2, three for fall grade 3, and four for spring grade 3. In this metric, π_0 represents the initial achievement status of each student at spring grade 1. The π_1 parameter is the linear growth rate (or learning rate) over the 2-year period. The dummy variable SUMDROP equals one at time points 1 and 3 takes on values of zero elsewhere. The associated parameter, π_2, captures the reduction in learning rate that occurs during the summer months as compared to the academic year, that is $\pi_1 + \pi_2 =$ summer learning rate.

For each of the individual growth parameters in π, we posed a student-level model:

$$\pi_{kij} = \beta_{k0j} + \beta_{k1j} \cdot \mathrm{CHILDPOV}_{ij} + \beta_{k2j} \cdot \mathrm{MOMED}_{ij}$$

$$+ R_{kij} \quad \text{for } k = 0, 1, 2 . \quad (15)$$

CHILDPOV is a dummy variable indicating that the child lives in a family whose household income is below the poverty line. The variable MOMED is the level of education of the child's mother, measured in years. In general, the errors, R_{kij}, are correlated among the i students within school j. The latter is a natural result of both the social processes that assign students to schools and the shared experiences students have after they arrive. For simplicity, we assume that this covariance structure, T_2, which is of dimension 3×3 is homogeneous across the J schools.

Equation (15) models the social distribution of initial status, π_{0ij}, learning rates, π_{1ij}, and summer drop-off effects, π_{2ij}, in school j as a function of child poverty and maternal education. The structure of the distributive effects in school j are captured in the β coefficients.

The variation among schools in these distributive effects are in turn modeled as a function of school poverty concentration. For each β coefficient in equation (15), we pose a school-level model:

$$\beta_{mj} = \gamma_{m0} + \gamma_{m1} \cdot \text{SCHLPOV}_j + U_{mj} , \tag{16}$$

where there are $m = 1, \ldots, M$ distributive effects to be modeled. In general, the errors, U_{mj}, are likely to be correlated since the distributive effects are likely to share common causes. Thus, we have another covariance matrix, T_1, of dimension $M \times M$ to estimate.

Fully unconditional model.—Tables 5 and 6 present results for an unconditional student and school model. For each of the k individual student growth parameters,

$$\pi_{kij} = \beta_{kj} + R_{kij} , \tag{17}$$

and, for each of the school mean growth parameters, β_{kj},

$$\beta_{kj} = \mu_k + U_{1j} . \tag{18}$$

This model provides estimates of the mean growth trajectories in mathematics and reading for the entire sample of students and schools. The mean initial achievement level is 404.28 for mathematics and 418.78 for reading. The learning rates over the 2-year period are similar for mathematics and reading, 28.39 and 23.35, respectively. The summer drop-off effects, however are noticeably different. For

TABLE 5

Sustaining Effects: Unconditional Student and School Models (Fixed Effects)

Individual Growth Model	Estimated Coefficient	Standard Error	t-Ratio
Mathematics:			
Initial status, μ_0	404.28	1.57	257.43
Learning rate, μ_1	28.39	.34	83.32
Summer drop-off, μ_2	−28.22	.95	−29.66
Reading:			
Initial status, μ_0	418.78	1.50	278.97
Learning rate, μ_1	23.35	.33	69.44
Summer drop-off, μ_2	−13.80	.80	−17.13

TABLE 6

Reliabilities for Individual Growth Parameters (π), School Mean Growth Parameters (β), and Percentage of Total Variance between Schools

Parameter	Mathematics	Reading
Reliability of individual growth parameters:		
Initial status, π_0	.648	.725
Learning rate, π_1	.051	.300
Summer drop-off, π_2	.003	.014
Reliability of school mean growth parameters:		
Initial status, β_0	.277	.537
Learning rate, β_1	.441	.436
Summer drop-off, β_2	.196	.035
Percentage of total parameter variance between schools:*		
Initial status	14.4	31.4
Learning rate	82.6	43.9
Summer drop-off	95.5	42.5

* $\{\mathrm{var}(\beta_k)/[\mathrm{var}(\beta_k) + \mathrm{var}(\pi_k)]\}$

mathematics, the estimated summer drop-off of −28.22 means that the learning rate during the summer period (28.39 − 28.22 = 0.17) is virtually zero. For reading, the summer learning rate is 9.55 (i.e., 23.35 − 13.80). These results are consistent with the contention that the observed summer drop-off is really an educational effect rather than a measurement artifact. Reading and verbal skills in general are learned both in school and at home, while mathematics learning tends to be restricted to the school context (Murnane 1975). Since students are not normally in school during the summer, the near-zero learning rate in mathematics is not surprising.

The unconditional model also provides valuable information about the reliabilities of the estimated individual growth trajectories within schools, the reliability of the estimated mean growth trajectories between schools, and the partitioning of the total variability into its within- and between-school components (see table 6). The differences in the reliabilities of the learning rates in reading and mathematics are particularly interesting. Although the reliabilities of the school mean learning rates, $\hat{\beta}_1$, are comparable (.441 in mathematics and .436 in reading), the reliability of the student (within-school) learning rates are quite different: in reading it is .300, whereas in mathematics it is only .051. These differences reflect the fact that there is more variability among student growth rates within schools in reading ($\text{var}(\pi_1) = 21.62$) than in mathematics ($\text{var}(\pi_1) = 3.80$; see bottom panel of table 7).

The observed differences in the within-school variability in mathematics and reading learning rates can be explained in terms of two phenomena reported in the instructional literature. First, since differences in home environment have a greater impact on reading than mathematics learning, greater variability in reading learning rates within schools is to be expected. In essence, there is an extra source of variability in individual reading learning rates which accrues from effects of home environmental differences. Second, the organization of classrooms typically employed in the teaching of reading and mathematics is also quite different. As Barr and Dreeben (1983) have shown, ability grouping is the common mode for organizing reading instruction, and much of the variation in reading achievement results from the differentiation among groups within classrooms. In mathematics, however, the whole-class method of instruction is more common (Stodolsky 1988). As a result, greater homogeneity of mathematics learning rates would be expected since there is less structural differentiation in students' experiences in this content area.

Perhaps the most dramatic results from the entire analysis are contained in the last panel of table 6 that displays the partitioning of variability into its within- and between-school components. For initial status, 14.4 percent of the variance in mathematics and 31.4 percent of the reading variance is between schools. In general, these results are consistent with reports in the school effects literature going back to *Equality of Educational Opportunity* (Coleman et al. 1966)—between 20 to 30 percent of the variance in achievement is between schools. However, the results for learning rates, particularly in mathematics, are startling indeed. Over 80 percent of the variance in mathematics learning is between schools! These results constitute powerful evidence of school effects that have gone undetected in past research. As we would expect, for the reasons outlined above, the between-school variance in reading learning rates is somewhat less, 43.9 percent, although still substantial.

We also note from table 6 the very low reliabilities of the individual summer drop-off estimates, .003 and .014 for mathematics and reading, respectively. To some unknown degree, this low reliability

TABLE 7

Sustaining Effects: Mother's Education and Child Poverty Included in Student Model, Unconditional School Model

Fixed Effects	Estimated Coefficient	Standard Error	t-Ratio
Mathematics growth model coefficient:			
Initial status, π_0:			
Base	376.668	6.298	59.79
MOMED	2.540	.500	5.07
CHILDPOV	−12.920	4.488	−2.87
Learning rate, π_1:			
Base	27.013	1.476	18.29
MOMED	.136	.117	1.15
CHILDPOV	−1.233	1.046	−1.17
Summer drop-off, π_2:			
Base	−28.196	.951	−29.63
Reading growth model coefficient:			
Initial status, π_0:			
Base	376.043	6.117	61.47
MOMED	3.878	.486	7.97
CHILDPOV	−16.331	4.368	−3.73
Learning rate, π_1:			
Base	21.580	1.394	15.48
MOMED	.217	.110	1.96
CHILDPOV	−4.929	.991	−4.97
Summer drop-off, π_2:			
Base	−13.770	.807	−17.06

	Variance Components and Percentage Reduction in Variance			
	Initial Status		Linear Growth	
Model	var(π_0)	R^2	var(π_1)	%R^2
Mathematics:				
Unconditional Student model	1,016.22	. . .	3.80	. . .
with CHILDPOV, MOMED	924.99	9.0	4.04	0*
Reading:				
Unconditional Student model	1,030.68	. . .	21.62	. . .
with CHILDPOV, MOMED	944.50	8.4	16.69	22.8

* When most of the observed variability in a random effect is sampling error, as in this case, the estimates of parameter variability are imprecise and can vary somewhat depending on other model specifications. The observed increase in this case has no substantive meaning.

is a function of an important missing piece of information in our SES data file. Some compensatory education students included in the SES were enrolled in summer school programs for 1 or more years. This is potentially an important source of variability that is unfortuntely relegated to sampling error since we lack student-level information on it.[2]

Modeling the effects of mother's education and child poverty.—Table 7 presents results from a model where MOMED and CHILDPOV have been introduced into the student-level models for individual status, π_0, and learning rates, π_1. Because of the low reliability of the estimates of the individual summer drop-off effects, no student-level modeling of π_2 was attempted. The school-level equation remains unconditional. Mother's education and child poverty are significantly related to initial status differences in both reading and mathematics. While mother's education (t-ratio = 1.96) and child poverty (t-ratio = −4.97) are related to reading learning rates, these relations do not exist for mathematics learning. So we have some further evidence supporting the proposition that the acquisition of reading ability is subject to greater home influence than is the acquisition of mathematics. Intepretation of this evidence requires caution, however, as the reliability of the individual mathematics learning rates is low ($\hat{\rho}$ = .051).

The percentage reduction in variances in individual status, π_0, and learning rates, π_1, is reported in the bottom panel of table 7. Mother's education and child poverty account for a similar proportion of the variance in children's initial status in reading and mathematics. In terms of learning rates, the effects are quite different, with a substantial percentage of variance explained in reading (22.8 percent), but none in mathematics. This is consistent with the overall pattern of results that reading acquisition depends to a greater extent on home environment.

Modeling the effects of poverty concentration.—The next fitted model examined the overall effects of school poverty concentration. We specified an unconditional student-level model and included SCHLPOV in the school-level equation for status, learning rates, and summer drop-off effects. This model represents the differences in mean growth trajectories between schools as a function of school poverty concentration. At this point, no controls have been introduced for child poverty or mother's education level.

The results, presented in table 8, indicate that school poverty concentration is negatively related to status differences in both reading and mathematics. It is only related to learning rate differences in reading, however. These results parallel the findings reported above for child poverty. The lack of relationship of school poverty to mathematics learning rates is particularly interesting because, unlike individual learning rates in mathematics, there is sufficient reliability now ($\hat{\rho}$ = .441) to detect effects if present.

Of greater significance is the fact that, for mathematics, school poverty concentration is *positively* associated with SUMDROP. This means that the summer drop-off in mathematics in high poverty schools is less than in their low poverty counterparts. Although this may at first appear anomalous, there is a simple explanation. High-poverty-concentration schools are much more likely to receive compensatory education funds and as a result to offer summer educational programs to disadvantaged youngsters. The absence of similar effects in reading is explainable in terms of the larger component that home environment plays in the acquisition of these skills. Whatever positive effects in reading may be accruing to disadvantaged youngsters as a result of summer programs are offset by the positive home effects accruing to their more advantaged peers. That is, home environment and summer school effects are confounded. As a result, it is difficult to detect the impact of summer programs on reading. Since less mathematics learning occurs at home, however, the confounding is minimized, and the school effects become easier to detect.

The estimates of the residual variance components from this analysis (bottom panel of table 8) indicate that a substantial portion of the variance in school mean status is related to SCHLPOV (64.1 percent and 46.7 percent for mathematics and reading, respectively). For school-level rates, the comparable figures are 0.2 percent and 26.0 percent, respectively.

Composite model including child poverty, mother's education, and school poverty concentration.—Our first attempt at a composite model included the effects of SCHLPOV on all of the β coefficients from the student-level models for initial status, π_0, learning rates, π_1, and summer drop-off as previously estimated in table 8. Several of these terms proved insignificant. They were eliminated and the model was refit. The final results are presented in table 9. The variables MOMED and SCHLPOV are both significantly related to initial mathematics and reading achievement (i.e., π_0). After controlling for school poverty, however, the effects of child poverty on initial reading and mathematics achievement is no longer significant. Both SCHLPOV and CHILDPOV are related to reading learning rates. Neither of the student background variables nor the school poverty measure is significantly related to mathematics learning rates. The positive effects of SCHLPOV on summer drop-off in mathematics, however, persists.

TABLE 8

Sustaining Effects: Unconditional Student Model, School Poverty Concentration in School Model

Fixed Effects	Estimated Coefficient	Standard Error	t-Ratio
Mathematics growth model coefficient:			
Initial status, π_0:			
Base	415.379	2.214	187.59
SCHLPOV	−.565	.080	−6.99
Learning rate, π_1:			
Base	28.690	.488	58.74
SCHLPOV	−.014	.018	−.81
Summer drop-off, π_2:			
Base	−31.039	1.364	−22.76
SCHLPOV	.144	.050	2.90
Reading growth model coefficient:			
Initial status, π_0:			
Base	433.320	2.154	201.13
SCHLPOV	−.737	.078	−9.42

Learning rate, π_1:			
Base	25.460	.481	52.98
SCHLPOV	−.107	.018	−6.09
Summer drop-off, π_2:			
Base	−13.512	1.154	−11.70
SCHLPOV	−.015	.042	−.35

	Variance Components and Percentage Reduction in Variance					
	Initial Status		Learning Rate		Summer Drop-off	
Model	var(β_0)	%R^2	var(β_1)	%R^2	var(β_2)	%R^2
Mathematics:						
Unconditional School model	170.46	. . .	17.95	. . .	41.42	. . .
School Model with SCHLPOV	61.11	64.1	17.92	.2	33.85	18.3
Reading:						
Unconditional School model	470.92	. . .	16.89	. . .	4.34	. . .
School Model with SCHLPOV	250.97	46.7	12.50	26.0	4.54	0*

* See note to table 7.

TABLE 9

Sustaining Effects: Full Model Including Child Poverty, Mother's Education, and School Poverty Concentration

Fixed Effects	Student-Level Model	School-Level Model	Estimated Coefficient	Standard Error	t-Ratio
Mathematics growth model coefficient:					
Initial status, π_0	base	base	387.146	6.774	57.15
	. . .	SCHLPOV	−.395	.091	−4.31
	MOMED	base	2.206	.507	4.35
	CHILDPOV	base	−4.997	4.947	−1.01
Learning rate, π_1	base	base	26.918	1.589	16.94
	. . .	SCHLPOV	.007	.020	.32
	MOMED	base	.136	.119	1.11
	CHILDPOV	base	−1.429	1.164	−1.23
Summer drop-off, π_2	base	base	−31.036	1.362	−22.77
		SCHLPOV	.147	.049	2.92
Reading growth model coefficient:					
Initial status, π_0	base	base	390.707	6.605	59.14
	. . .	SCHLPOV	−.514	.089	−5.80
	MOMED	base	3.337	.495	6.73
	CHILDPOV	base	−4.389	4.835	−.91
Learning rate, π_1	base	base	23.471	1.505	15.59
	. . .	SCHLPOV	−.066	.020	−3.34
	MOMED	base	.148	.113	1.31
	CHILDPOV	base	−3.403	1.110	−3.09
Summer drop-off, π_2	base	base	−13.504	1.155	−11.69
	. . .	SCHLPOV	−.014	.042	−.32

Discussion

On balance, both the results presented here and our discussion of their implications are admittedly speculative because of the limited SES data used in these analyses. Nevertheless, these results should serve to illustrate the illuminative power that the three-level model offers for research on school and instructional effects. Use of this model with data from multi–time point designs, with good measures of individual growth, student's educational experiences, and organizational context, holds much promise for substantially extending our understanding of the effects of teaching and learning.

This first application of a three-level model has also been instructive from a methodological point of view. In combining the two models of growth and school effects, the three-level structure produces much more than the sum of its two parts since the parameters from the two models join in a multiplicative fashion. As a result, the data are summarized with a very dense web of empirical information. Structural effects at the individual and school level can be disentangled, and the variance-covariance can be partitioned into within- and between-school components. Such partitioning provides valuable information about the potential sources of variation that would be useful in formulating future models. For example, conventional analyses would not have discovered that 80 percent of the variation in mathematics learning rates is between schools in the SES data.

The variance partitioning also provides evidence about the reliability of the data for measuring status, learning rates, and summer drop-off effects at both the individual student and school mean level. These reliability estimates are very useful in interpreting the results from the structural analyses. For example, under the summer drop-off hypothesis, we expected that the drop-off should be greater for the more disadvantaged youth. A failure to find such relationships would tend to support the competing explanation, advanced in the literature, that the observed summer drop-off is really just a measurement artifact. Although the direction of the estimated effects of family background on the summer drop-off was consistent with our hypothesis, the related test statistics were not significant. The standard interpretation of this result would be that the hypothesized summer drop-off effects were nonexistent. The estimates of the reliability of the individual summer drop-off effects, approximately 0.01, indicate, however, that these data have little power to detect relationships between student background and a summer drop-off in learning even if these effects were quite large. As a result, rather than having evidence supporting a measurement artifact explanation, we have here an instance where the proposed statistical test was incapable of distinguishing between the competing explanations.

In discussing the design of observational studies over 30 years ago, R. A. Fisher advised researchers to make their theories complex. Rather than viewing design of observational studies as formulating a "single best test of a proposition," Fisher's advice directs us toward developing construct validity webs. That is, if the stated proposition is true, what interrelated set of empirical results should follow? The three-level HLM generates a high density of empirical results useful in formulating and testing such construct validity webs. By comparing structural effects and variance-covariance component estimates at different levels, by contrasting the structural effects with the estimated variance-covariance component, there is much to be learned in any substantive application. This dense web of empirical evidence provides a more extensive basis for framing explanations, and the logical process for evaluating such explanations becomes more complex. When such explanations are successfully crafted, they will surely stand with more force and greater clarity than we are accustomed to in educational research settings.

Appendix
Description of HS&B Variables

Student-Level Predictors

ACADBKGD:	a factor composite of HS&B variables that indicate if the respondent has taken remedial math and/or English, expects to attend college, has been read to before starting school, and has ever repeated a grade
MINORITY:	a dummy variable (1 = black or hispanic/ 0 = others)
SES:	the HS&B standardized composite

School-Level Predictors

I. The social and academic context of schools:

AVSES:	average SES of students within the school
HIMNRTY:	enrollment in excess of 40 percent minority (black and/or hispanic)
AVACBKGD:	a student-level factor composed of remedial placement at entry into high school, reported college aspirations at grade 8, left back in elementary school, and early reading in the home (averaged to the school level)

II. Perceived teacher quality and interest in students:

STFPBLM: principal's report about staff absenteeism and lack of commitment and motivation

PCDQLTCH: a factor composite of student report about the percentage of their teachers who enjoy their work, make clear presentations, work students hard, treat students with respect, are witty and humorous, don't talk over students' heads, are patient and understanding, return work properly, and are interested in students outside of class (school-level average)

III. Discipline climate of the school:

DISCLIM: a composite index based on (i) a factor score from students' reports about the incidence of students talking back to teachers, refusal to obey instructions, attacks on teachers and fights with each other (school-level average); and (ii) the school average of student reports about their own discipline problems in school, suspension, probation, and cutting class.

SAFE: percentage of students who feel safe in the school environment

AUTHRTY: students' ratings of the fairness and effectiveness of discipline within the school (averaged to school level)

IV. Academic press of the school:

AVHMEWRK: hours per week students spend on homework (school average)

ATTACAD: factor composite based on student attitudes toward getting good grades and interest in academics (averaged to school level)

V. Curricular structure:

AVMTHEMP: average number of advanced mathematics courses taken by students (a school measure of the emphasis on academic course work)

SDMTHEMP: standard deviation in the number of advanced mathematics courses taken by students (a school measure of differentiation in academic coursework)

AVACDPGM: percentage of students in the academic program.

Notes

The research here has been supported by grants from the Spencer Foundation and the Benton Center for Curriculum and Instruction at the University of Chicago. We wish to acknowledge the gracious assistance of David Myers in making a small subset of his Sustaining Effects data files available to us for this analysis. We also wish to thank Michael Seltzer for his thoughtful comments on an earlier draft of this article.

1. All HLM significance testing is based on large sample theory. The probability values associated with the test statistics reported in this section should be cautiously interpreted because the sample size in this study is small ($N = 11$). They are best viewed as statistical indicators rather than precise probability statements.

2. Hierarchical linear modeling employs the EM algorithm in estimating variance-covariance components. When a variance is close to zero, as in this case, the convergence of the algorithm is poor, and there is little assurance that the number estimated is actually the likelihood maximum. It is clear that the parameter variance in individual summer drop-off effects was very small, although exactly how small is uncertain. Whatever the precise number, there was clearly little reliability to detect effects, and for this reason, the structural analyses alluded to in this paragraph were not discussed in the presentation of results.

References

Alexander, K. L., and A. M. Pallas. "Private Schools and Public Policy: New Evidence on Cognitive Achievement in Public and Private Schools." *Sociology of Education* 56 (1983): 170–82.

Barr, R., and R. Dreeben. *How Schools Work.* Chicago: University of Chicago Press, 1983.

Bidwell, C. E., and J. Kasarda. "Conceptualizing and Measuring the Effects of School and Schooling." *American Journal of Education* 88 (1980): 401–30.

Braun, H. I., D. H. Jones, D. B. Rubin, and D. T. Thayer. "Empirical Bayes Estimation of Coefficients in the General Linear Model with Data of Deficient Rank." *Psychometriks* 48, no. 2 (1983): 171–81.

Bryk, A. S., P. Holland, and V. Lee. "A Vision of Community: The Contemporary Catholic High School." Manuscript in preparation, Department of Education, University of Chicago.

Bryk, A. S., P. Holland, V. Lee, and R. Carriedo. *Effective Catholic Schools: An Exploration.* Washington, D.C.: National Catholic Educational Association, 1984.

Bryk, A. S., and S. W. Raudenbush. "Application of Hierarchical Linear Models to Assessing Change." *Psychological Bulletin* 101, no. 1 (1987): 147–58. (*a*)

Bryk, A. S., S. W. Raudenbush, R. Congdon, and M. Seltzer. *An Introduction to HLM: Computer Program and User's Guide.* Chicago: University of Chicago, Department of Education, 1986.

Bryk, A. S., and H. I. Weisberg. "Use of the Nonequivalent Control Group Design When Subjects Are Growing." *Psychological Bulletin* 84 (1977): 950–62.

Burstein, L. "The Analysis of Multi-Level Data in Educational Research and Evaluation." *Review of Research in Education* 8 (1980): 158–233.

Carter, L. F. "The Sustaining Effects Study of Compensatory and Elementary Education." *Educational Researcher* 13, no. 7 (1984): 4–13.

Coleman, J. S., E. Q. Campbell, C. F. Hobson, J. McPartland, A. M. Mood, F. D. Weinfeld, and R. L. York. "Equality of Educational Opportunity." Washington, D.C.: Department of Health, Education, and Welfare, 1966.

Coleman, J. S., T. Hoffer, and S. B. Kilgore. *High School Achievement: Public, Catholic, and Other Private Schools Compared.* New York: Basic, 1982.

Cronbach, L. J. "Research on Classrooms and Schools: Formulation of Questions, Design and Analysis." Occasional papers in the Stanford Evaluation Consortium, 1976.

Cronbach, L. J., and N. Webb. "Between- and Within-Class Effects in a Reported Aptitude-by-Treatment Interaction: Reanalysis of a Study by G. L. Anderson." *Journal of Educational Psychology* 6 (1975): 717–24.

Fisher, R. A. *Statistical Methods and Scientific Inference.* New York: Hafner, 1956.

Goldstein, H. "Multi-Level Mixed Linear Model Analysis Using Iterative Generalized Squares." *Biometrika,* no. 1 (1986): 43–56.

Greeley, A. *Minority Students in Catholic Schools.* New Brunswick, N.J.: Transaction, 1982.

Hammond, P. A., and J. A. Frechtling. "Twelve, Nine and Three-Month Achievement Gains of Low and Average Achieving Elementary School Students." Paper presented at annual American Educational Research Association meeting, San Francisco, 1979.

Haney, W. "Units and Levels of Analysis in Large-Scale Evaluation." *New Directions for Methodology of Social and Behavioral Sciences* 6 (1980): 1–15.

Harris, C. W., *Problems in the Measurement of Change.* Madison: University of Wisconsin Press, 1963.

Heyns, B. *Summer Learning and the Effects of Schooling.* New York: Academic Press, 1978.

Huttenlocher, J. E., W. Haight, A. S. Bryk, and M. Seltzer. "Parental Speech and Early Vocabulary Development." Unpublished manuscript, Department of Education, University of Chicago, 1988.

Lee, V., and A. S. Bryk. "The Effects of High School Academic Organization on the Social Distribution of Achievement." Paper presented at annual American Educational Research Association meeting, San Francisco, 1986, and revised, 1988.

Murnane, R. *The Impact of School Resources of Inner-City School Children.* Cambridge, Mass.: Ballinger, 1975.

Myers, D. E. "The Relationship between School Poverty Concentration and Students' Reading and Math Achievement and Learning." Technical Report, Decision Resources Corporation, Washington, D.C., 1986.

Newport, E. L., H. Gleitman, and L. R. Gleitman. "Mother, I'd Rather Do It Myself: Some Effects and Non-effects of Maternal Speech Style." In *Talking to Children: Language Input and Acquisition,* edited by C. E. Snow and C. A. Ferguson. New York: Cambridge University Press, 1977.

Pelavin, S. H., and J. L. David. "Evaluating Long-Term Achievement: An Analysis of Longitudinal Data from Compensatory Education Programs." Report no. EPRC 4537-15, Stanford Research Institute, Menlo Park, Calif., 1977.

Raudenbush, S. W. "Applications of Hierarchical Linear Models in Educational Research: A Review." *Journal of Educational Statistics* (1988), in press.

Raudenbush, S. W., and A. S. Bryk. "A Hierarchical Model for Studying School Effects." *Sociology of Education* 59 (1986): 1–17.

Raudenbush, S. W., and A. S. Bryk. "Quantitative Models for Estimating Teacher and School Effectiveness." In *Multiple Learning Analysis of Educational Data,* edited by D. Bock. New York: Academic Press.

Rogosa, D. R. "Politics, Process, and Pyramids." *Journal of Educational Statistics* 3 (1978): 79–86.
Rogosa, D. R., D. Brand, and M. Zimowski. "A Growth Curve Approach to the Measurement of Change." *Psychological Bulletin* 90 (1982): 726–48.
Rogosa, D. R., and J. B. Willett. "Understanding Correleates of Change by Modeling Individual Differences in Growth." *Psychometrika* 90 (1985): 726–48.
Sorsenson, A., and M. Hallinan. "A Reconceptualization of School Effects." *Sociology of Education* 50 (1977): 273–89.
Stodolsky, S. S. *The Subject Matters: Classroom Activity in Math and Social Studies.* Chicago: University of Chicago Press, 1988.
Ware, J. H. "Linear Models for the Analysis of Longitudinal Studies." *Journal of the American Statistical Association* 80 (1985): 95–101.
Willett, J. "Questions and Answers in the Measurement of Change." *Review of Research in Education* (1988), in press.
Willms, J. D. "School Effectiveness within the Public and Private Sectors: An Evaluation." *Evaluation Review* 8 (1984): 113–35.