# A HIERARCHICAL MODEL FOR STUDYING SCHOOL EFFECTS

STEPHEN RAUDENBUSH
*Michigan State University*

ANTHONY S. BRYK
*University of Chicago*

Sociology of Education 1986, Vol. 59 (January):1–17

*When researchers investigate how school policies, practices, or climates affect student outcomes, they use multilevel, hierarchical data. Though methodologists have consistently warned of the formidable inferential problems such data pose for traditional statistical methods, no comprehensive alternative analytic strategy has been available. This paper presents a general statistical methodology for such hierarchically structured data and illustrates its use by reexamining the High School and Beyond data and the controversy over the effectiveness of public and Catholic schools. The model enables the researcher to utilize mean achievement and certain structural parameters that characterize the equity in the social distribution of achievement as multivariate outcomes for each school. Variation in these school-level outcomes is then explained as a function of school characteristics.*

At the heart of much educational research are hypotheses about the influence of policies or practices implemented at the school or district level on processes occurring within schools. Such hypotheses are implicitly multilevel—i.e., key independent variables are typically measured at a higher level of aggregation than the outcome variables of interest. Methodologists have warned that the use of traditional linear models to study multilevel phenomena can produce misleading results (Boyd and Iverson 1979; Haney 1980; Burstein 1980a, 1980b; Cooley, Bond, and Mao 1981). As Cronbach (1976, p. 1) stated,

> The majority of studies of educational effects—whether classroom experiments, or evaluations of programs or surveys—have collected and analyzed data in ways that conceal more than they reveal. The established methods have generated false conclusions in many studies.

Yet despite such forceful warnings, single-level linear-model analyses of school effects abound. For example, there has been a spate of research, generated by the High School and Beyond (HSB) survey, on the differential effectiveness of public and private schools (see Coleman, Hoffer, and Kilgore 1982, and critics such as Page and Keith 1981; McPartland and McDill 1982; and Alexander and Pallas 1983). In the past, analysts clung to single-level models not out of conviction but because of the absence of viable alternatives. However, increasing awareness of the mismatch between multilevel social processes and the traditional statistical models used to study them has spurred a search for multilevel analytic strategies. Moreover, recent advances in statistical theory offer the necessary tools for a more appropriate approach.

In this paper, we present a general statistical model for studying school effects. The model responds to many of the criticisms of single-level methods and in fact incorporates as special cases many of the proposed alternatives. We illustrate its application using the HSB data to reexamine the controversy over the relative effectiveness of public and private schools. Though the results are preliminary, the analysis suggests that the relationship between students' socioeconomic status (SES) and their mathematics achievement varies from school to school, that failure to consider such variation may lead to invalid inferences about the effects of sector (Catholic vs. public), and that the attempt to explain such variation may enrich the conceptualization of future research in this area.

## THE MISMATCH BETWEEN SINGLE-LEVEL MODELS AND MULTILEVEL EFFECTS

Even carefully controlled experimental research can be plagued by the hierarchical character of educational data. For instance, Cronbach and Webb (1975) reanalyzed a published

The research reported here was supported in part by a seed grant from the Spencer Foundation to Harvard University. We also wish to acknowledge Mediax Interactive Technologies for their assistance in developing the computer programs used in this work. Lastly, we are indebted to Richard Murnane for his thoughtful criticisms of an earlier draft. Address all correspondence to Professor Stephen Raudenbush, College of Education, Michigan State University, 460 Erickson Hall, East Lansing, MI 48824-111.

study that found an important aptitude-by-treatment interaction effect for alternative methods of arithmetic instruction. Students were the unit of analysis in the original study. A key sociological unit, the classroom, was ignored. The reanalysis, which decomposed the observed relationship between aptitude and achievement into between- and within-classroom components, showed little evidence of an aptitude-by-treatment interaction *within* classrooms. Further, the interactions that occurred between classrooms were not theoretically interpretable.

Barr and Dreeben (1983) make a similar claim in their examination of the existing research on the effects of ability grouping. Though it may be appropriate to conceive of ability grouping as a school policy, the ability group itself is an important social unit whose effects must be considered in this research. The failure to incorporate this social unit into the analytic models used in ability grouping research has produced seriously flawed inferences.

Similar problems can occur in survey research. For example, Coleman et al. (1982) report that the relationship between social class and achievement is weaker in Catholic schools than in public schools. This finding is the primary basis for their conclusion that "Catholic schools more closely approximate the common-school ideal" (p. 177). The statistical evidence for this conclusion derives from separate student-level regressions within the public and Catholic sectors. The potential effects of the school as a sociological unit were ignored. Willms (1984) and Alexander and Pallas (1985) have recently challenged the validity of the common-school finding on somewhat related grounds.

## SLOPES-AS-OUTCOMES: A PROMISING APPROACH

To detect and explain variations in structural relations across schools—e.g., variations in the effect of SES on achievement—one might use regression slopes as outcomes (see Burstein, Linn, and Capell 1978; Burstein and Miller 1980). Consider, for example, the hypothesis that the weak relationship between social class and achievement within Catholic schools is due, in part, to the greater academic course requirements in these schools. Testing such a hypothesis involves two stages. In the first stage, achievement is regressed on SES within each school. The estimated slopes from this first stage become the outcomes in the second-stage analysis, which attempts to explain variations in slopes on the basis of a range of academic policy variables.

The use of slopes-as-outcomes is appealing because it extends substantially the kinds of questions that school research can examine. It permits researchers to look beyond the effects of school policies on average achievement to explore the effects of these policies on the relationships occurring within groups.

Unfortunately, a number of technical difficulties have inhibited the widespread use of slopes-as-outcomes. First, as a general rule, regression coefficients have considerably greater sampling variability than sample means (Wiley 1970; Burstein and Miller 1980). If the sample within a unit, such as a school, is small, the regression coefficient will be estimated with large sampling error. The resultant unreliability in slopes weakens our power to detect relationships in the second-stage model. This unreliability is exacerbated when the dispersion in the first-stage independent variable is constrained. For example, students tend to be more homogeneous in social class within schools than they are in a true random sample. As a result, the sampling variability of the within-group SES-achievement slope is increased. The analysis can produce negative within-school slope estimates even when the structural parameter is clearly positive (Willms 1984). This is particularly problematic, since outliers caused by sampling variability can exert undue influence on the second-stage results (Burstein et al. 1978).

Second, the sampling precision of the estimated slopes varies across units depending upon the data collection design used within each unit. But ordinary least squares (OLS), the estimation method typically used for the second-stage model, assumes equal variances across cases on the dependent variable. The main effect of this mismatch is weakened efficiency in parameter estimation for the second-stage model. This further limits our ability to detect relationships between slopes-as-outcomes and the group-level variables hypothesized to account for them.

Third, the variability in the estimated slopes consists of two components. We assume that there are real differences across units in the slope parameters, and we refer to such variability as *parameter variance*. It is essential to distinguish between parameter variance and the *sampling variance* discussed above. This distinction becomes especially important when we attempt to interpret the results from the second-stage model. Only parameter variance is potentially explainable. For the reasons mentioned above, we expect that in many applications, much of the observed variance in the slopes is sampling variance. A second-stage model that explains only a small percentage of the observed variance might be discounted when in fact it is explaining a very large portion of what can, in principle, be explained. Unfortunately, the simple slopes-as-outcomes model does not provide a ready means for estimating these separate variance components.

Finally, until very recently, estimation of the slopes-as-outcomes model has been limited to a single independent variable for the within-group regressions (cf. Hanushek 1974). Even if one is interested only in a single within-group relationship, it is often crucial to control for other variables that are confounded with it. To include multiple slopes-as-outcomes in the second stage requires an extension of the model to take into account the special covariance structure that exists among the multiple-regression coefficients estimated within each group. In the absence of such a model, biased parameter estimates and further weakened precision are likely to result.

Recent advances in statistical theory make possible an approach that responds to these various deficiencies in the slopes-as-outcomes model. This new approach provides a flexible statistical tool for studying how variations in policies and practices influence educational processes. It offers a basis for constructing richer definitions of school effectiveness—definitions that include not only mean achievement but, for example, the social distribution of such achievement. Lastly, it enables researchers to recognize situations in which the data lack the necessary power to detect effects that exist.

## OVERVIEW OF THE HIERARCHICAL LINEAR MODEL

The statistical theory behind the proposed approach builds on several developments, including mixed-model ANOVA (cf. Elston and Grizzle 1962), regression with random coefficients (cf. Rao 1972; Swamy 1973; Rosenberg 1973; Dielman 1983), James-Stein estimation (cf. Efron and Morris 1975), covariance component models (Harville 1977), and Bayesian estimation for linear models (Lindley and Smith 1972; Smith 1973; Dempster, Rubin, and Tsutakawa 1981; and Morris 1983). The papers by Lindley and Smith are key contributions: They explicitly lay out a hierarchical structure in which parameters estimated at one level in the model become the outcome variables at the next level. It is this hierarchical parametric structure that permits the modeling of multilevel phenomena encountered in school-effects research. Following the lead of Lindley and Smith, we refer to this as a hierarchical linear model (HLM). In other recent sociological applications, this model has been termed a general multilevel linear model (Mason, Wong, and Entwisle 1984) and a multilevel mixed linear model (Goldstein in press).

In fact, the hierarchical linear model has broad applicability in social and psychological research. It can be used in studies of individual growth (Laird and Ware 1982; Strenio, Weisberg, and Bryk 1983; Bock 1983), in the measurement of change (Bryk and Raudenbush 1984), in studies of contextual effects in cross-national fertility data (Mason et al. 1983), in research on teaching styles (Aitkin, Anderson, and Hinde 1981), and in research synthesis or meta-analysis (Raudenbush and Bryk 1985).

HLM enables us to conceptualize in terms of multiple levels. In the interests of simplicity and clarity of exposition, we consider only two levels in this paper, although the statistical theory permits us to consider more (see Goldstein in press). Further, in our examples, the school is the key sociological unit of interest, and the basic design involves students nested within schools. The methods introduced in this paper can be applied to any grouping variable, including, for example, districts or classrooms.

We first pose a *within-group* model that specifies the relationships among various student-level characteristics, $X_{ijk}$, and outcomes of interest, such as achievement scores, $y_{ij}$. Thus, we estimate a separate regression equation for each school:

$$y_{ij} = \beta_{j0} + \beta_{j1}X_{ij1} + \beta_{j2}X_{ij2} + \ldots + \beta_{jK-1}X_{ijK-1} + R_{ij}, \qquad (1)$$

for $i = 1, \ldots, n_j$ students in school $j$; $j = 1, \ldots, J$ schools; and $k = 0, \ldots, K$-1 independent variables in the within-school model. Here, $y_{ij}$ is the outcome score for student $i$ in school $j$, the $X_{ijk}$ are the values on the student-level characteristics for individual $i$ in school $j$, $R_{ij}$ is random error, and the $\beta_{jk}$ are regression coefficients that characterize the structural relationships within school $j$.

This is a standard linear model with one major exception: The within-school regression coefficients, $\beta_{jk}$, are allowed to vary across schools. We are particularly interested in situations in which the variability in the regression parameters is a function of differences in policies and practices among schools. Therefore, we pose a second, or *between-group*, model. For each of the $K$ regression coefficients in equation (1) we assume that

$$\beta_{jk} = \theta_{0k} + \theta_{1k}Z_{1j} + \theta_{2k}Z_{2j} + \ldots + \theta_{P-1,k}Z_{P-1,j} + U_{jk}, \qquad (2)$$

for $p = 0, \ldots, P$-1 independent variables in the between-group model, where the $\theta_{pk}$ are regression coefficients that capture the effects of school-level variables on the within-school structural relationships ($\beta_{jk}$), $U_{jk}$ is random error in this school-level equation, and the $Z_{pj}$ are values on the school-level variables for unit $j$.

By estimating such an equation for each of the $K$ within-group parameters, we have completed the multivariate representation for the slopes-as-outcomes model. When the independent variables in equation (1) are centered around their respective group means, $\overline{X}_{.jk}$, $\beta_{j0}$ represents the mean achievement level in school $j$. The between-school model for $\beta_{j0}$ affords a direct representation of the effects of school variables on mean achievement. Thus, the between-school model allows a simultaneous investigation of effects on school means and regression coefficients.

## RELATIONSHIP TO OTHER MODELS THAT EXAMINE SCHOOL EFFECTS

The model represented in (1) and (2) reduces to simpler and better-known models under specific conditions. This can be seen by substituting (2) into (1), which yields

$$\begin{aligned} y_{ij} = {} & \theta_{00} + \sum_{k=1} \theta_{0k} X_{ijk} + \sum_{p=1} \theta_{p0} Z_{pj} \\ & + \sum_{k=1} \sum_{p=1} \theta_{pk} X_{ijk} Z_{pj} \\ & + \underbrace{U_{j0} + \sum_{k=1} U_{jk} X_{ijk} + R_{ij}}_{\text{error term}}. \end{aligned} \tag{3}$$

When there are no random effects in the between-group model (i.e., when $U_{jk} = 0$ for all $j, k$), the hierarchical linear model becomes equivalent to an ordinary regression model that includes student-level variables, $X_{ijk}$, school-level variables, $Z_{pj}$, and their interaction terms, $X_{ijk}Z_{pj}$. Valid application of this model, which is commonly used in school-effects research, requires the assumption that all the variation in the within-school parameters, $\beta_{jk}$, has been explained by knowledge of the school-level variables, $Z_{pj}$. When random effects remain (i.e., when one or more of the $U_{jk}$ are not equal to zero), the application of OLS to (3) is inefficient, and the estimated standard errors are too small.

Another model that has received some attention considers the within-school intercepts, $\beta_{j0}$, as random but the regression slopes, $\beta_{jk}$, as fixed (i.e., $U_{jk} = 0$ for all $j$ when $k \geq 1$). The simplest case involves a between-school model, where

$$\begin{aligned} & \beta_{j0} = \gamma_0 + U_{j0}, \text{ and} \\ & \beta_{jp} = \beta_p \text{ for } p = 1, \ldots, P\text{-}1. \end{aligned} \tag{4}$$

Aitkin et al. (1981) used a variant of (4) in their research on teaching styles.

Similarly, the multilevel decomposition suggested by Cronbach (1976) can be accomplished by setting

$$\begin{aligned} \beta_{j0} = {} & \gamma_{00} + \gamma_{10}\overline{X}_{1j} + \gamma_{20}\overline{X}_{2j} \\ & + \ldots + \gamma_{K\text{-}1,j}\overline{X}_{K\text{-}1,j} + U_{j0}, \text{ and} \\ \beta_{jk} = {} & \beta_k \text{ for } k = 1, \ldots, K\text{-}1. \end{aligned} \tag{5}$$

This permits estimation of both the pooled within-group slopes, $\beta_k$, and the corresponding between-group slopes, $\gamma_{k0}$. The hierarchical regression models proposed by Keesling and Wiley (1974) and the covariance component approaches suggested by Wisenbaker and Schmidt (1979) are also specific submodels of (1) and (2). For a comprehensive review of proposed alternative approaches to the analysis of school effects see Burstein (1980a, 1980b). Also, Mason et al. (1983) provide a detailed discussion of the relationship between the general statistical model and a variety of simplified submodels. In the example presented below, we illustrate specification tests that enable the analyst to determine whether any of these simplifications are justifiable.

## EMPIRICAL BAYES: UNIVARIATE RESULTS FOR NORMAL DATA

Estimates of the parameters in HLM can be derived in several ways (see Raudenbush 1984). We are interested in point estimates and confidence intervals for the $\gamma$'s, and since the $\beta$'s are assumed random, we are also interested in the expectations, variances, and covariances among these components. Empirical Bayes methods provide a comprehensive approach to this estimation problem. We illustrate the logic of this approach with a simple case.

Assume that within schools the outcome variable, academic achievement, is a function of a single predictor, a student's social class, plus random error. We can further simplify this illustration by assuming that the data for each school have been centered around the school mean. Thus we have the *within-school* model

$$y_{ij} = \beta_j X_{ij} + R_{ij}, \tag{6}$$

where $y_{ij}$ is the academic achievement of student $i$ in school $j$ and $X_{ij}$ is the social class of the same student.

We assume that the errors, $R_{ij}$, are normally distributed within each school with mean 0 and constant variance $\sigma^2$. Again, $\beta_j$ varies across groups. However, at this point we assume no knowledge of the school-level factors that influence $\beta_j$. Thus, we have the *between-school* model

$$\beta_j = \mu_\beta + U_j. \tag{7}$$

The outcome variable in this model is the within-group regression coefficient, which is a function of a grand mean plus random error. We assume that $U_j$ is normally distributed with mean 0 and variance $\tau$. Under this simple model, $\tau$ represents the parameter variance in $\beta_j$, i.e., the variability in the true SES-achievement slopes across the population of schools. We are interested in estimating the strength of the SES-achievement relationship in each school, $\beta_j$, and the average magnitude of the SES-achievement relationship, $\mu_\beta$, across the population of schools.

We can estimate the $\beta_j$ in equation (6) for each school using OLS. The resultant estimate, $\hat{\beta}_j$, contains some sampling error, $e_j$:

$$\hat{\beta}_j = \beta_j + e_j. \tag{8}$$

Under OLS theory, the errors, $e_j$, are normally distributed with mean 0 and variance $v_j$, where

$$v_j = \text{var}(\hat{\beta}_j | \beta_j) = \frac{\sigma^2}{\Sigma X^2}. \tag{9}$$

When we combine equations (7) and (8) we get

$$\hat{\beta}_j = \mu_\beta + U_j + e_j. \tag{10}$$

Assuming that the $e_j$ and $U_j$ are statistically independent, it follows that $\hat{\beta}_j$ is normally distributed with a mean $\mu_\beta$ and observed variance $v_j + \tau$:

$$\begin{aligned}\text{var}(\hat{\beta}_j) &= \text{var}(\hat{\beta}_j | \beta_j) + \text{var}(\beta_j) \\ &= v_j + \tau.\end{aligned} \tag{11}$$

This expression is particularly important in slopes-as-outcomes research because it allows us to partition the observed variance in the estimated regression coefficients, $\text{var}(\hat{\beta}_j)$, into two components: sampling variance ($v_j$) and parameter variance ($\tau$).

Under empirical Bayes, assuming $v_j$ and $\tau$ are known, the minimum mean squared error point estimators for $\beta_j$ and $\mu_\beta$ are

$$\beta_j^* = w_j\hat{\beta}_j + (1 - w_j)\, \mu_\beta^*, \tag{12}$$

$$\mu_\beta^* = \frac{\Sigma w_j \hat{\beta}_j}{\Sigma w_j}, \tag{13}$$

where

$$w_j = \frac{\tau}{\tau + v_j}. \tag{14}$$

It also follows directly that both $\beta_j^*$ and $\mu_\beta^*$ are normally distributed with known variance.

The $w_j$ can be viewed as reliability coefficients, analogous to those of classical measurement theory. They are the ratio of the true parameter variance, $\tau$, to the observed variance, $\tau + v_j$, and they indicate the reliability of $\hat{\beta}_j$ as an estimator of school $j$'s slope.

The empirical Bayes estimator for the mean slope in the population of schools, $\mu_\beta^*$, is the average of the within-school OLS slopes, where each slope is weighted according to its reliability. The empirical Bayes estimator of the individual school slopes, $\beta_j^*$, is a weighted combination of (1) $\hat{\beta}_j$, the OLS slopes derived for each school from its student data, and (2) $\mu_\beta^*$, the estimated mean slope for the population of schools. The weight accorded $\hat{\beta}_j$ is the reliability of that group's estimate. If $\hat{\beta}_j$ is not very reliable, $\beta_j^*$ will be based primarily on the estimated mean slope, $\mu_\beta^*$.

We immediately see one of the benefits of the empirical Bayes approach. Outlier values for $\beta_j^*$ resulting from a large sampling error in $\hat{\beta}_j$ are controlled. Empirical Bayes "shrinks" these outliers in toward the population mean. It is for this reason that these methods are sometimes referred to as statistical shrinking. Efron and Morris (1975) and Morris (1983) have shown that in general, $\beta_j^*$ will have smaller mean squared error than $\hat{\beta}_j$. The improvement in some applications has been quite dramatic.

### MULTIVARIATE EXTENSION

The empirical Bayes results for the univariate case presented above extend in a natural fashion to the full multivariate model represented by equations (1) and (2). Instead of a single $\beta_j$ for each school, equation (1) contains a $K \times 1$ vector of coefficients, $\boldsymbol{\beta}_j$, which captures the structural relationships within school $j$. The within-school model, analogous to (1), is

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{R}_j, \tag{15}$$

where $\mathbf{y}_j$ is an $n_j \times 1$ vector on the outcome variable, $\mathbf{X}_j$ is an $n_j \times K$ matrix of predictor variables, and $\mathbf{R}_j$ is an $n_j \times 1$ vector of random errors.

As before, OLS yields an estimate for $\boldsymbol{\beta}_j$, in which the estimates $\hat{\boldsymbol{\beta}}_j$ will have sampling errors with dispersion:

$$\text{var}(\hat{\boldsymbol{\beta}}_j | \boldsymbol{\beta}_j) = \sigma^2(\mathbf{X}_j'\mathbf{X}_j)^{-1} = \mathbf{V}_j, \tag{16}$$

where $\mathbf{V}_j$ is a $K \times K$ covariance matrix whose diagonal elements, $v_{jkk}$, contain the sampling variances for each $\hat{\boldsymbol{\beta}}_{jk}$ respectively.

The extension for the between-group model is somewhat more complex because it is doubly multivariate: Multiple school-level factors, $Z_{pj}$, are explaining multiple within-school structural relationships, $\boldsymbol{\beta}_j$. When we rewrite equation (2) in matrix form, we get

$$\boldsymbol{\beta}_j = \mathbf{Z}_j\boldsymbol{\theta} + \mathbf{U}_j, \tag{17}$$

where $\mathbf{Z}_j$ is a matrix containing the school-level data and $\boldsymbol{\theta}$ is a vector of coefficients that specify the effects of the school factors on the various within-school relationships, $\boldsymbol{\beta}_j$. We assume that the errors, $\mathbf{U}_j$, are normally distributed with mean $\mathbf{0}$ and dispersion matrix $\mathbf{T}_Z$. Analogous to $\tau$, $\mathbf{T}_Z$ is a measure of parameter dispersion. Now, however, it is the residual parameter variance-covariance in $\boldsymbol{\beta}_j$ after accounting for the effects of measured school-level factors, $\mathbf{Z}_j$.

As in (11), the total dispersion in the $\hat{\boldsymbol{\beta}}_j$ estimates can be partitioned into sampling dispersion and a residual parameter dispersion:

$$\begin{aligned}\text{var}(\hat{\boldsymbol{\beta}}_j|\mathbf{Z}) &= \text{var}(\hat{\boldsymbol{\beta}}_j|\boldsymbol{\beta}_j) + \text{var}(\boldsymbol{\beta}_j|\mathbf{Z}) \\ &= \mathbf{V}_j + \mathbf{T}_Z. \end{aligned} \tag{18}$$

Empirical Bayes theory yields the following estimates for the model parameters:

$$\boldsymbol{\beta}_j^* = \mathbf{W}_j\hat{\boldsymbol{\beta}}_j + (\mathbf{I} - \mathbf{W}_j)\,\mathbf{Z}_j\boldsymbol{\theta}^*, \tag{19}$$

where

$$\mathbf{W}_j = \mathbf{T}_Z(\mathbf{V}_j + \mathbf{T}_Z)^{-1} \tag{20}$$

and

$$\begin{aligned}\boldsymbol{\theta}^* &= \{\Sigma\mathbf{Z}_j'[\text{var}(\hat{\boldsymbol{\beta}}_j|\mathbf{Z})]^{-1}\mathbf{Z}_j\}^{-1} \\ &\times \{\sum_j \mathbf{Z}_j'[\text{var}(\hat{\boldsymbol{\beta}}_j|\mathbf{Z})]^{-1}\hat{\boldsymbol{\beta}}_j\}. \end{aligned} \tag{21}$$

Further, under the assumption that $\sigma^2$ and $\mathbf{T}_Z$ are known, both $\boldsymbol{\beta}_j^*$ and $\boldsymbol{\theta}^*$ are multivariate normal with known dispersion.

The empirical Bayes estimates for the within-school regression coefficients, $\boldsymbol{\beta}_j^*$, are again a weighted composite. The first component is an estimate based only on the individual student data within school $j$. The second component, $\mathbf{Z}\boldsymbol{\theta}^*$, is an estimate based on the school-level data, $\mathbf{Z}$. The weight matrix $\mathbf{W}_j$ resembles a multivariate reliability coefficient, analogous to the $w_j$ discussed earlier. Thus, the regression coefficients estimated by OLS within any school are weighted according to their "reliability." However, if these estimates are not very reliable, empirical Bayes will rely more heavily on the school-level results. As in the univariate case, the latter provide mean estimates for the within-school structural parameters, but these estimates are now conditional on the specific values for the school-level variables, $\mathbf{Z}$.

Empirical Bayes also yields estimates for the $\boldsymbol{\theta}$, the parameters that specify the effects of school factors on within-school structural relationships. Equation (21) is recognizable as the generalized least squares estimate of $\boldsymbol{\theta}$, where each $\hat{\boldsymbol{\beta}}_j$ is weighted inversely proportional to the observed dispersion in $\hat{\boldsymbol{\beta}}_j$, $\text{var}(\hat{\boldsymbol{\beta}}_j|\mathbf{Z})$. Thus, each school's $\hat{\boldsymbol{\beta}}_j$ is weighted proportional to its precision. Further, the formula is fully multivariate; i.e., it takes into account the covariances among the elements of $\hat{\boldsymbol{\beta}}_j$. To the extent that such covariances exist, the $\boldsymbol{\theta}^*$ estimates will be more precise. This is another advantage over previous slopes-as-outcomes research, which does not capitalize on these associations.

In essence, the empirical Bayes estimates borrow their strength from whatever relationships happen to exist in the data. As a result, HLM enables a more precise measurement of the structural parameters for each school, thereby mitigating the problem of imprecision in slopes, which has plagued previous slopes-as-outcomes research.

## THE PRACTICAL PROBLEM: APPLYING EMPIRICAL BAYES WHEN $\mathbf{T}_Z$ AND $\mathbf{V}_j$ ARE UNKNOWN

We have assumed so far that the variances and covariances, $\mathbf{T}_Z$ and $\mathbf{V}_j$, are known. Though this assumption is helpful in illuminating the logic of the model, it is seldom tenable in practice. In fact, the problems associated with estimating these dispersion matrices have limited the applications of hierarchical linear models since Lindley and Smith's (1972) seminal work.

The development of the EM algorithm by Dempster, Laird, and Rubin (1977) affords a theoretically satisfactory and computationally manageable approach to covariance component estimation in hierarchical linear models. It has been successfully used in a broad range of applications (see, for example, Dempster et al. 1981; Strenio et al. 1983; and Mason et al. 1983), and we use it in the example that follows. Under fairly general conditions, the EM algorithm produces maximum likelihood estimates for variance components such as $\mathbf{T}_Z$ and $\mathbf{V}_j$. These estimates have several desirable properties: They are asymptotically unbiased, consistent, efficient, and asymptotically normally distributed. Their normal distribution provides the basis for large-sample statistical inference with HLM. When the EM estimates are substituted into equation (21), the resulting $\boldsymbol{\theta}^*$ are also maximum likelihood estimates with known asymptotic distributions. The logic of EM variance components estimation with HLM, using the simple univariate model presented in equations (6) and (7), is illustrated in the appendix. This logic extends readily to facilitate estimation in the multivariate case.

Recently, other options for estimating $\mathbf{T}_Z$ and $\mathbf{V}_j$ have begun to appear. Goldstein (in press) details an iterative generalized least squares procedure that also produces maximum likelihood estimates when used with multilevel models such as those specified above. Longford (1985) considers the application of the Fisher scoring method to this problem. Rachman-Moore and Wolfe (1984) explore the use of robust estimation techniques with a constrained version of our general model.

## AN ILLUSTRATIVE EXAMPLE

We use HSB data to illustrate the logic of HLM and to show how it allows researchers to pose and test a variety of hypotheses about school effects. In particular, we demonstrate how we might use HLM to reconsider the question of the relative effectiveness of public and private schools. Coleman et al.'s assertion that Catholic schools are more effective than public schools has been the focus of unusually intense controversy. Critics have challenged the adequacy of the cognitive tests used to evaluate school effects (Heyns and Hilton 1982; Goldberger and Cain 1982), the practical importance and magnitude of the effects (Alexander and Pallas 1983; Willms 1984), and especially the adequacy of the control for selectivity bias (Bryk 1981; Murnane 1981; Goldberger and Cain 1982; Alexander and Pallas 1983; Noell 1982; see Coleman et al. 1982 and Kilgore 1983 for replies). Since the release of the first HSB follow-up tapes, the analyses and debates have continued (Hoffer, Greeley, and Coleman 1985; Alexander and Pallas 1985; Willms 1985).

We focus our attention on the common-school hypothesis first advanced by Coleman et al. (1982). In brief, they concluded that Catholic schools had higher overall academic achievement than public schools and that this Catholic-school advantage was most pronounced for lower-SES students. This is equivalent to a sector-by-SES interaction, in which the SES-achievement relationship is weaker in Catholic schools than in public schools.

*Data*

The analyses presented below use the baseline data from the HSB survey. The analysis sample consisted of 10,231 students from 82 Catholic schools (the entire Catholic sector) and 94 public schools. The student samples per school ranged from 10 to 70, though samples of less than 45 students were rare. The outcome variable for these illustrations is the standardized mathematics achievement score.

*Model I: Examining Variability Among Schools*

Our first model tests the hypothesis that the SES-achievement relationship varies across schools. This model permits us to estimate both the mean equation for the regression of achievement on SES pooled within schools and the extent of the variability among the individual school regression equations. We can also test whether the observed variability across schools in mean achievement levels and in the strength of the SES-achievement relationship actually represents more than sampling error.

The *within-school* model is

$$y_{ij} = \beta_{j0} + \beta_{j1}(X_{ij1} - \overline{X}_{.j1}) + R_{ij},$$

where $y_{ij}$ is mathematics achievement for student $i$ in school $j$, $\beta_{j0}$ is the mean mathematics achievement for school $j$, $\beta_{j1}$ is the SES-achievement relationship in school $j$, $X_{ij1}$ is the SES of student $i$ in school $j$ (a composite variable with a mean of $-0.18$ and a standard deviation of 0.68), $\overline{X}_{.j1}$ is the mean SES for school $j$, and $R_{ij}$ is the error of estimate for student $i$ in school $j$.

The *between-school* model is

$$\beta_{j0} = \theta_{00} + U_{j0},$$
$$\beta_{j1} = \theta_{10} + U_{j1},$$

where $\theta_{00}$ is the grand mean for mathematics achievement across all schools, $\theta_{10}$ is the mean slope for the SES-achievement relationship pooled within all schools, $U_{j0}$ is the effect of school $j$ on the mean mathematics achievement, and $U_{j1}$ is the effect of school $j$ on the SES-achievement relationship. The full results for this model are reported in Table 1.

The estimated mean within-school regression equation is

$$\hat{y}_{ij} = \hat{\theta}_{00} + \hat{\theta}_{10}(X_{ij1} - \overline{X}_{.j1})$$
$$= 100.74 + 4.52(X_{ij1} - \overline{X}_{.j1}).$$

We can test the hypothesis that $\theta_{10} = 0$ (i.e., that the mean SES-achievement slope is zero). Under the null hypothesis, $\theta_{10}/[\text{s.e. } (\theta_{10})]$ has an asymptotic $z$ distribution. We see from Table 1 that the observed coefficient 4.52 could hardly have occurred by chance alone ($z = 15.56$). We are led to conclude that, on average, there is a positive relationship between SES and mathematics achievement within schools. The relationship is modest in size, equivalent to a correlation of approximately .20.

Table 1 also reports the estimated parameter variance for $\beta_{j0}$ and $\beta_{j1}$. We can pose hypotheses about these parameters as well. Specifically, we can examine whether $\text{var}(\beta_{j0}) = 0$ (i.e., whether the observed differences among schools in mean achievement level could have occurred by chance alone) and $\text{var}(\beta_{j1}) = 0$ (i.e., whether the observed differences among schools in the SES-achievement relationship could have occurred by chance alone).

Recall from (16) that the $v_{jkk}$ are the sampling variances associated with $\beta_{jk}$. Under the hypotheses posed above, the test statistics

$$\sum_j \left(\frac{1}{v_{j00}}\right)(\hat{\beta}_{j0} - \theta_{00}^*)^2, \text{ and}$$
$$\sum_j \left(\frac{1}{v_{j11}}\right)(\hat{\beta}_{j1} - \theta_{10}^*)^2 \qquad (22)$$

have asymptotic $\chi^2$ distributions with $J$-1 = 175 degrees of freedom (Hedges 1982). We see from Table 1 that both of these hypotheses are highly implausible ($\chi^2 = 2{,}234.09$, $p < .001$; and

Table 1. HLM Results for Model 1.

| | Effect | S.E. | $z$ statistic | Estimate | $\chi^2$ | $df$ |
|---|---|---|---|---|---|---|
| Fixed effects | | | | | | |
| Average within-school equation | | | | | | |
| Mean achievement ($\hat{\theta}_{00}$) | 100.74 | .573 | 175.81 | | | |
| Mean SES-achievement slope ($\hat{\theta}_{10}$) | 4.52 | .290 | 15.56 | | | |
| Random effects | | | | | | |
| Variation among school means in achievement | | | | | | |
| Parameter variance [var($\beta_{j0}$)] | | | | 53.25 | 2,234.09* | 175 |
| Total observed variance [var($\hat{\beta}_{j0}$)] | | | | 58.66 | | |
| Variation among school SES-achievement slopes | | | | | | |
| Parameter variance [var($\beta_{j1}$)] | | | | 5.25 | 270.46* | 175 |
| Total observed variance [var($\hat{\beta}_{j1}$)] | | | | 15.11 | | |
| Model statistics | | | | | | |
| Maximum likelihood estimate for $\sigma^2$ | | | | 247.03 | | |

* Significant at the .001 level.

$\chi^2 = 270.46$, $p < .001$). It seems reasonable to conclude that schools vary in their mean achievement levels and that the relationship between mathematics achievement and SES is truly different across schools. To use the $\chi^2$ statistic, we must assume that the estimated $v_{jkk}$ are equal to their parameter values. This assumption is plausible when the total sample size, $\Sigma n_j$, is large, as it is here ($\Sigma n_j = 10{,}231$).

Finally, it is instructive to compare the estimated parameter variances, var($\beta_{jk}$), to their observed variances, var($\hat{\beta}_{jk}$). Of the total observed variability among the school means, about 90 percent is parameter variance. However, only about 35 percent of the observed slope variability is parameter variance. The remaining 65 percent is sampling variance. This result illustrates the difficulties frequently encountered in much of the previous slopes-as-outcomes research. Only parameter variance is potentially explainable by school-level factors. Without decomposing the observed variance into its parameter and sampling components, the analyst would be unaware that only about 35 percent of the observed variance in slopes can be explained. Stated somewhat differently, the traditional measure for the adequacy of a regression model, $R^2$, cannot exceed .35. Analysts who fail to distinguish between these components of variance might conclude that a model for slopes-as-outcomes is inadequate even if it accounts for most of the explainable variance.

### *Model 2: Identifying Variability Among Schools as a Function of Sector*

Having concluded from our first model that mean achievement and the SES-achievement relationship differ across schools, we now try to account for these differences. We ask, To what extent are the observed differences a function of sector? Model 2 represents a first approach to this question. We pose the same *within-school* model as above:

$$y_{ij} = \beta_{j0} + \beta_{j1}(X_{ij1} - \overline{X}_{.j1}) + R_{ij}.$$

But we now extend the *between-school* model to include sector:

$$\beta_{j0} = \theta_{00} + \theta_{01}Z_{j1} + U_{j0},$$
$$\beta_{j1} = \theta_{10} + \theta_{11}Z_{j1} + U_{j1},$$

where $Z_{j1}$ is sector (1 = Catholic, 0 = public), $\theta_{00}$ is the mean mathematics achievement in the public sector, $\theta_{10}$ is the mean slope for the SES-achievement relationship in the public sector, $\theta_{01}$ is the Catholic-school effect on mean mathematics achievement, $\theta_{11}$ is the Catholic-school effect on the SES-achievement slope, $U_{j0}$ is the effect of school $j$ on the mean mathematics achievement after accounting for the sector effects, and $U_{j1}$ is the effect of school $j$ on the SES-achievement relationship after accounting for the sector effects.

The HLM results are reported in Table 2. Substituting the between-school equation into the within-school equation for $\beta_{j0}$ and $\beta_{j1}$, respectively, yields the following estimated mean within-school regression equation:

$$\hat{y}_{ij} = [98.37 + 5.06(Z_{j1})] + \{[6.23 - 3.86(Z_{j1})]\,(X_{ij1} - \overline{X}_{.j1})\}. \quad (23)$$

This is analogous to a conventional regression analysis where $Z_{j1}$ is the design variable and $X_{ij1}$ is a covariate whose slope is assumed to be different for the two sectors. However, the estimated coefficients are not identical: HLM assumes that the $\beta$'s are random, whereas in regression, the parameters are assumed constant.

Table 2. HLM Results for Model 2

| | Effect | S.E. | $z$ statistic | Estimate | $\chi^2$ | $df$ |
|---|---|---|---|---|---|---|
| Fixed effects | | | | | | |
| Average public-school equation | | | | | | |
| Mean achievement ($\hat{\theta}_{00}$) | 98.37 | .74 | 132.93 | | | |
| Mean SES-achievement slope ($\hat{\theta}_{10}$) | 6.23 | .33 | 18.88 | | | |
| Effects of between-school variables | | | | | | |
| Catholic-sector effects on | | | | | | |
| mean achievement ($\hat{\theta}_{01}$) | 5.06 | 1.09 | 4.64 | | | |
| SES-achievement slope ($\hat{\theta}_{11}$) | −3.86 | .49 | −7.87 | | | |
| Random effects | | | | | | |
| Variation among school means in achievement | | | | | | |
| Residual parameter variance [var($\beta_{j0}\|Z_1$)] | | | | 47.36 | 1,988.79** | 174 |
| Total observed variance [var($\hat{\beta}_{j0}$)] | | | | 58.66 | | |
| Variation among school SES-achievement slopes | | | | | | |
| Residual parameter variance [var($\beta_{j1}\|Z_1$)] | | | | 1.49 | 197.48* | 174 |
| Total observed variance [var($\hat{\beta}_{j1}$)] | | | | 15.11 | | |
| Model statistics | | | | | | |
| Maximum likelihood estimate for $\sigma^2$ | | | | 246.95 | | |
| Percent of observed variance explained ($R^2$) | | | | | | |
| School means ($\hat{\beta}_{j0}$) | | | | 10.2 | | |
| SES-achievement slopes ($\hat{\beta}_{j1}$) | | | | 25.0 | | |
| Percent of parameter variance explained ($R^{2*}$) | | | | | | |
| School means ($\beta_{j0}$) | | | | 11.3 | | |
| SES-achievement slopes ($\beta_{j1}$) | | | | 71.6 | | |

* Significant at the .08 level.
** Significant at the .001 level.

Under this model, the estimated Catholic-school effect on average achievement is 5.06 points ($\theta_{00}^* = 5.06$, $z = 4.65$, $p < .001$). This represents an effect of slightly less than ⅓ standard deviation. The relationship of SES to mathematics achievement is significantly smaller in Catholic schools than in public schools ($\theta_{11}^* = 3.86$, $z = -7.78$, $p < .001$).

The parameter variance estimates, var($\beta_{j0}|Z_1$) and var($\beta_{j1}|Z_1$), are now conditional variances. They measure the amount of variability remaining among the school means and slopes after sector is taken into account. By comparing these to the unconditional parameter variance estimates from model 1, we have a natural estimator of the proportion of parameter variance that is explained by sector. For each $\beta_{jk}$ where $k = 0, 1$,

$$R^{2*} = 1 - \frac{\text{var}(\beta_{jk}) - \text{var}(\beta_{jk}|Z_1)}{\text{var}(\beta_{jk})}. \quad (24)$$

Though sector accounts for only 11.3 percent of the parameter variance among school means, it accounts for 71.6 percent of the variance in SES-achievement slopes.

We can also compare these conditional parameter variances to the observed variances for $\beta_{jk}$ from Table 1. Sector accounts for 10.2 percent of the observed variance among school means and 25.0 percent of the observed variance among slopes. The latter is considerably smaller than $R^{2*}$ because of the substantial sampling error in the observed SES-achievement slopes.

Finally, we can pose null hypotheses about the conditional variances too. We hypothesize that var($\beta_{j0}|Z_1$)$=0$ (i.e., after taking sector into account, the remaining differences among schools in mean achievement level could have occurred by chance alone) and var($\beta_{j1}|Z_1$)$=0$ (i.e., after taking sector into account, the remaining differences among schools in the SES-achievement relationship could have occurred by chance alone).

Under these hypotheses, the statistics

$$\sum_j \left(\frac{1}{v_{jkk}}\right)[\hat{\beta}_{jk} - \theta_{k0}^* - \theta_{k1}^*(Z_{j1})]^2 \quad \text{for } k = 0, 1 \quad (25)$$

have an asymptotic $\chi^2$ distribution with $J$-2 degrees of freedom. From Table 2 it is apparent that the hypothesis is highly implausible for school means ($\chi^2 = 1{,}989.79$, $df = 174$, $p < .001$). The parameter variance among school means has been reduced by taking sector into account, but more variability remains to be explained. For the SES-achievement slopes, the hypothesis that there is no residual variability among the $\beta_{j1}$ is more plausible ($\chi^2$

= 197.48, $df$ = 174, $p < .08$). This result, however, does not preclude the possibility that other school factors may account for additional variance in the SES-achievement slopes.

*Model 3: Accounting for the Sector Effects*

The results presented in Table 2 indicate a relationship between sector and both the average achievement and the SES-achievement slope. However, at this point it is premature to conclude that such relationships are causal. Rather, these findings may result from misspecification of either the within-school or the between-school models. Our purpose now is to illustrate how to approach this problem using HLM. The results reported below should be viewed as preliminary. A more detailed investigation of this issue will be reported later in a separate paper.

The most obvious alternative hypothesis is that differential selection of students into Catholic and public schools accounts for the observed sector effects. For example, suppose that low-SES students who attend Catholic schools have a better academic background than their public-school counterparts. Then, the within-school model would have to be expanded to include academic background; otherwise, estimates for $\beta_{j1}$ would be biased, which would lead to an overestimate of the sector effects in the between-school model.

Misspecification problems can also occur in the between-school model. Suppose, for example, that the SES-achievement relationship is weaker in the Northeast than in the rest of the country. Since Catholic schools are disproportionately represented in this region, failure to take region into account in the between-school model would again lead to a biased estimate of the sector effect.

Finally, even if sector effects are real, it is important to identify the specific features of Catholic schools—including characteristics of their student bodies and school policies—that might account for their higher mean achievement and flatter slopes. To detect such school factors, we must expand the between-school model.

In sum, HLM offers two methods for adjusting the sector effects: (1) controlling the student-level differences and (2) controlling for school-level differences. We illustrate this approach by assuming that the amount of homework is a confounding variable in the within-school model and that the SES composition of the school and its interaction with sector are confounding variables in the between-school model. Thus, we now have the following *within-school* model:

$$y_{ij} = \beta_{j0} + \beta_{j1}\,(X_{ij1} - \overline{X}_{.j1}) + \beta_{j2}\,(X_{ij2} - \overline{X}_{.j2}) + R_{ij}, \qquad (26)$$

where $y_{ij}$ is mathematics achievement for student $i$ in school $j$, $\beta_{j0}$ is the mean mathematics achievement in school $j$, $\beta_{j1}$ and $\beta_{j2}$ are the SES-achievement and homework-achievement relationships in school $j$, $X_{ij1}$ is the SES of student $i$ in school $j$, $X_{ij2}$ is the amount of time spent on homework (hours per week) by student $i$ in school $j$, $X_{.j1}$ and $X_{.j2}$ are the mean SES and mean hours of homework in school $j$, and $R_{ij}$ is the error of estimate for student $i$ in school $j$.

The *between-school* model is

$$\begin{aligned}
\beta_{j0} &= \theta_{00} + \theta_{01}Z_{j1} + \theta_{02}(Z_{j2} - \overline{Z}_{.2}) + \theta_{03}Z_{j1}(Z_{j2} - \overline{Z}_{.2}) + U_{j0},\\
\beta_{j1} &= \theta_{01} + \theta_{11}Z_{j1} + \theta_{12}(Z_{j2} - \overline{Z}_{.2}) + \theta_{13}Z_{j1}(Z_{j2} - \overline{Z}_{.2}) + U_{j1}, \qquad (27)\\
\beta_{j2} &= \theta_{20} + \theta_{21}Z_{j1} + \theta_{22}(Z_{j2} - \overline{Z}_{.2}) + \theta_{23}Z_{j1}(Z_{j2} - \overline{Z}_{.2}) + U_{j2},
\end{aligned}$$

where $Z_{j1}$ is sector (1 = Catholic, 0 = public); $Z_{j2}$ is school SES, i.e., the average SES of students in school $j$ ($\overline{Z}_{.2}$ = 0.18; s.d. $[Z_{j2}]$ = .41); $\theta_{00}$ is the mean mathematics achievement in the public sector; $\theta_{10}$ and $\theta_{20}$ are the mean slopes in the public sector for the SES-achievement and homework-achievement relationships; $\theta_{01}$ is the Catholic-school effect on mean mathematics achievement; $\theta_{11}$ and $\theta_{21}$ are the Catholic-school effects on the SES-achievement and homework-achievement slopes; $\theta_{02}$ is the effect of school SES on mean achievement; $\theta_{12}$ and $\theta_{22}$ are the effects of school SES on the SES-achievement and homework-achievement slopes; $\theta_{03}$, $\theta_{13}$, and $\theta_{23}$ are the effects of the sector-by-school-SES interactions on mean achievement, the SES-achievement slope, and the homework-achievement slope; and $U_{j0}$, $U_{j1}$, and $U_{j2}$ are the remaining random effects associated with school $j$.

The full results for this model are reported in Table 3. Some of the salient findings are as follows:

1. School SES is strongly related to mean achievement ($\theta_{02}^*$ = 15.09, $z$ = 11.47, $p <$ .001). This is equivalent to a standardized regression coefficient of .82, indicating that an increase of one standard deviation in school SES predicts an increase of approximately 0.82 standard deviation in mathematics achievement. After adjusting for school SES, the estimated Catholic-school effect on mean achievement is reduced from 5.06 points to 0.43 ($\theta_{01}^*$ = 0.43, $z$ = 0.48, n.s.). Thus, differences in social-class composition between public and Catholic schools appear

Table 3. HLM Results for Model 3

| | Effect | S.E. | $z$ statistic | Estimate | $\chi^2$ | $df$ |
|---|---|---|---|---|---|---|
| Fixed effects | | | | | | |
| Predicted public-school equation[a] | | | | | | |
| Mean achievement ($\hat{\theta}_{00}$) | 101.08 | .65 | 158.88 | | | |
| Mean SES-achievement slope ($\hat{\theta}_{10}$) | 6.14 | .45 | 14.40 | | | |
| Mean homework-achievement slope ($\hat{\theta}_{20}$) | 1.01 | .09 | 11.44 | | | |
| Effects of between-school variables | | | | | | |
| Catholic-sector effects on | | | | | | |
| mean achievement ($\hat{\theta}_{01}$) | .43 | .83 | .52 | | | |
| SES-achievement slope ($\hat{\theta}_{11}$) | −4.24 | .57 | 7.44 | | | |
| homework-achievement slope ($\hat{\theta}_{21}$) | −.36 | .11 | 3.27 | | | |
| School SES effects on | | | | | | |
| mean achievement ($\hat{\theta}_{02}$) | 15.09 | 1.32 | 11.43 | | | |
| SES-achievement slope ($\hat{\theta}_{12}$) | 2.32 | .90 | 2.58 | | | |
| homework-achievement slope ($\hat{\theta}_{22}$) | .11 | .19 | .58 | | | |
| School SES × sector effects on | | | | | | |
| mean achievement ($\hat{\theta}_{03}$) | −2.50 | 1.85 | 1.35 | | | |
| SES-achievement slope ($\hat{\theta}_{13}$) | .41 | 1.28 | .32 | | | |
| homework-achievement slope ($\hat{\theta}_{23}$) | −.31 | .25 | 1.24 | | | |
| Random effects | | | | | | |
| Variation among school means in achievement | | | | | | |
| Residual parameter variance [$\text{var}(\beta_{j0} \mid Z_1\ Z_2\ Z_3)$] | | | | 18.01 | 909.50* | 172 |
| Total observed variance [$\text{var}(\hat{\beta}_{j0})$] | | | | 58.66 | | |
| Variation among school SES-achievement slopes | | | | | | |
| Residual parameter variance [$\text{var}(\beta_{j1} \mid Z_1\ Z_2\ Z_3)$] | | | | .88 | 166.82 | 172 |
| Total observed variance [$\text{var}(\hat{\beta}_{j1})$] | | | | 15.11 | | |
| Variation among school homework-achievement slopes | | | | | | |
| Residual parameter variance [$\text{var}(\beta_{j2} \mid Z_1\ Z_2\ Z_3)$] | | | | .109 | 160.21 | 172 |
| Total observed variance [$\text{var}(\hat{\beta}_{j2})$] | | | | .649 | | |
| Model statistics | | | | | | |
| Maximum likelihood estimate for $\sigma^2$ | | | | 236.51 | | |
| Percent of observed variance explained ($R^2$) | | | | | | |
| School means ($\hat{\beta}_{j0}$) | | | | 60.0 | | |
| SES-achievement slopes ($\hat{\beta}_{j1}$) | | | | 29.0 | | |
| Homework-achievement slopes ($\hat{\beta}_{j2}$) | | | | 4.2 | | |
| Percent of parameter variance explained ($R^{2*}$) | | | | | | |
| School means ($\beta_{j0}$) | | | | 66.2 | | |
| SES-achievement slopes ($\beta_{j1}$) | | | | 83.2 | | |
| Homework-achievement slopes ($\beta_{j2}$) | | | | 19.9 | | |

NOTE: In a separate analysis, we estimated $\text{var}(\beta_{j2}) = .136$.

[a] This is the within-school equation predicted for public schools with a school SES equal to the sample average of −0.18. In the HSB sample that we used in these analyses, the average school SES in the public sector was −0.33.

* Significant at the .001 level.

to account for the observed mean mathematics achievement differences between the two sectors.

2. School SES also affects the strength of the SES-achievement relationship within schools. The estimated effect ($\theta_{12}^* = 2.32$, $z = 2.57$, $p < .01$) indicates that a student's social class has a stronger effect on individual achievement in higher social-class schools than in schools with a larger proportion of less advantaged students. Stated somewhat differently, higher social-class schools are less egalitarian than poorer schools.
3. The Catholic-school effect on the SES-achievement relationship persists after taking homework, school SES, and the sector-by-school-SES interaction into account. In fact, Catholic schools appear to be slightly more egalitarian under this

model ($\theta_{11}^* = -4.24$) than under model 2 ($\theta_{11}^* = -3.86$). Thus, while model 3 explains away the Catholic-school advantage on mean mathematics achievement, it does not account for the Catholic-school effect on the SES-achievement relationship.

4. The interactions of sector with school SES have no apparent effect on mean achievement or on the within-school relationships. The $z$ statistics for $\theta_{03}^*$, $\theta_{13}^*$, and $\theta_{23}^*$ fail to approach significance. This suggests that we delete these interaction terms from subsequent models.
5. Model 3 accounts for 66.2 percent of the parameter variance in mean mathematics achievement and 83.2 percent of the parameter variance in the SES-achievement slopes. The $\chi^2$ test of the hypothesis that the residual variances in $\beta_{j0}$ and $\beta_{j1}$ are 0 yields test statistics of 909.5 ($p < .001$) and 166.8 (n.s.), respectively. Thus, it is possible to retain the null hypothesis of heterogeneity of regression: After including the effect of school SES, no significant variation among slopes remains to be explained. This hypothesis should be retained tentatively, since the test results do not rule out the possibility that residual parameter variance is greater than 0.

In general, as we introduce variables in the between-school model, the equation for $\beta_{j0}$ takes on special interest. Specifically, if we center the numerical variables in equation (26) around their respective school means, the intercept, $\beta_{j0}$, captures the mean achievement level in school $j$. Thus, the *between-group* equation for $\beta_{j0}$ represents the regression of mean achievement on school-level factors and is equivalent to using the school as the unit of analysis.

Further, as noted previously, the *within-group* model (equation [26]) represents the pooled within-school regression of achievement on student-level characteristics. Thus, by estimating an HLM model in which the specification of $\beta_{j0}$ in the between-school equation parallels the specification of the within-school equation, the analyst can decompose effects into their within- and between-school components, as advocated by Cronbach (1976).

For example, under model 3, the SES-by-sector interaction is partitioned into within- and between-school effects. The $\theta_{02}$ parameter captures the *between-school* relationship of school SES to mean achievement in the public sector. For the Catholic sector, the *between-school* estimate is $\theta_{02} + \theta_{03}$. The pooled *within-school* estimates of the SES-achievement relationship in each sector can be derived from the corresponding parameters in the equation for $\beta_{j1}$. We display these results in Figures 1 and 2.

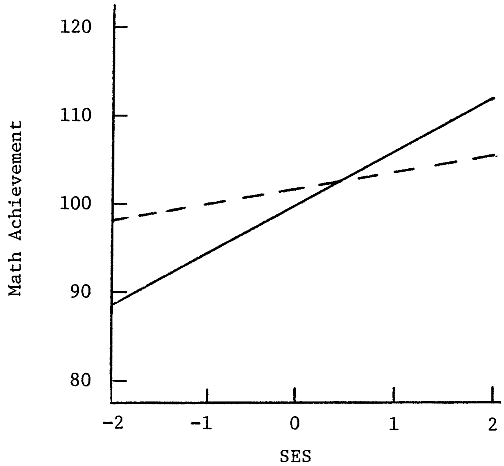


Figure 1. Average effect of SES on math achievement within public schools (solid line) and Catholic schools (broken line), controlling for the effect of homework within schools and mean SES between schools. (Intercepts and slopes for public and Catholic schools are compiled from results in Table 3. The value of school mean SES is assumed to be −0.182, the sample average. The scale for SES [horizontal axis] is in standard deviation units.)

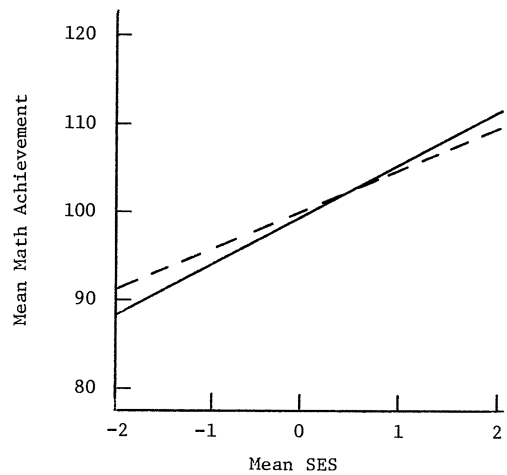


Figure 2. Public-school mean math achievement (solid line) and Catholic-school mean math achievement (broken line) as a function of mean SES. (These between-school slopes are not significantly different for the two sectors. The scale for mean SES [horizontal axis] is in standard deviation units.)

It is apparent from Figure 1 that the relationship between social class and achievement is weaker (i.e., the slope is flatter) *within* the typical Catholic school than it is within the typical public school. (Again, these within-school differences are based on a model that controls for homework at the within-school level and for school SES and the sector-by-school-SES interaction at the between-school level.) This result supports the contention that Catholic schools are more egalitarian than public schools. It is also important to recall, however, that the sector differences in mean achievement virtually disappeared under this model. This is reflected in Figure 1. The two lines cross at a point just slightly above the average SES for the sample. Thus, we have an apparent disordinal interaction: Lower-SES students fare better in Catholic schools, and higher-SES students fare better in public schools.

Figure 2 illustrates the SES-achievement relationships for the public and Catholic sectors based on the between-school model—i.e., when the school is the unit of analysis. A similar disordinal interaction is manifest, but its size is sufficiently small to be discounted as statistically unreliable.

Again, we caution against overinterpretation of these results. Our analyses are preliminary and have been presented here primarily to demonstrate the value of HLM in research on school effects. Clearly, we need to undertake analyses that introduce more extensive controls at both the student and school levels. Should our results be sustained upon further analysis, however, they would tend to support a school-effects explanation. That is, our preliminary findings indicate that the difference between sectors in the strength of the SES-achievement relationship is primarily a within-school phenomenon. Had we found the reverse—i.e., if the difference between sectors in the strength of the SES-achievement relationship were restricted to the between-school model—it would be much more difficult to posit a school-effects explanation. Rather, a selection explanation, in which the difference in the SES-achievement relationship is a result of differences between sectors in the process by which individuals are assigned to schools, would be much more plausible.

## DISCUSSION

### *Usefulness of HLM in School-Effects Research*

By its very nature, school-effects research requires exploration of hierarchical data—a search for statistical associations between school factors on the one hand and student-level variables on the other. Yet analysts typically treat such data as if it were single level, leading almost inevitably to serious inferential problems. This paper has proposed the use of the hierarchical linear model in research on school effects and has illustrated its application. HLM is a powerful tool that permits a separation of within-school from between-school phenomena and allows simultaneous consideration of the effects of school factors not only on school means but also on structural relationships within schools.

One of the key points in the examples presented above is the importance of distinguishing between parameter variance and sampling variance in attempting to account for within-school relationships. Failure to make this distinction can lead to the mistaken conclusion that relationships between school characteristics and slopes are weak. For example, Willms (1984), using a slopes-as-outcomes approach to reconsider the common-school hypothesis, concluded that sector differences in the SES-achievement slopes are trivial in comparison to the within-sector variation in slopes. However, since the sampling error is included in the within-sector variability, Willms has set a very conservative standard for proclaiming an effect. The results we obtained after decomposing the observed slope variance into its parameter and sampling components cast doubt on the validity of Willms's conclusion.

In our examples, we emphasized the estimation and interpretation of the $\boldsymbol{\theta}$ coefficients, since these parameters are likely to be of primary interest in research on school effects. First, we use the $\boldsymbol{\theta}$ coefficients to estimate the mean within-school regression equation. Second, we use them to estimate the effects of school factors on mean school achievement and on within-school structural relationships (i.e. slopes-as-outcomes).

The $\mathbf{T}$ dispersion matrix is also of considerable substantive interest. Its diagonal elements provide estimates of the variability across schools in mean achievement and in regression slopes. The covariance terms in the $\mathbf{T}$ matrix can also be important, because they represent the covariances among the $\beta_{jk}$ coefficients. In general, these covariances allow us to explore the relationships among the within-school structural parameters, which become the dependent variables in the between-school equations. For example, in model 1, the covariance of mean achievement within schools, $\beta_{j0}$, and the SES-achievement slope within schools, $\beta_{j1}$, are represented in the off-diagonal element, $\mathbf{T}_{01}$: i.e., $\mathbf{T}_{01} = \text{cov}(\beta_{j0}, \beta_{j1})$. When this parameter is examined, the investigator can ask, Is the SES-achievement relationship stronger in schools that have a higher mean achievement level?

Similarly, the empirical Bayes estimates for the within-school structural parameters, $\beta_{jk}^*$, can be useful in school-effects research. Since these coefficients are known to have smaller mean squared error than the OLS estimates based only on the individual student data from school $j$, the $\beta_{jk}^*$ provide a better descriptive characterization of each school. This could be particularly useful in research that seeks to identify unusually effective schools.

*Model Assumptions*

HLM requires two key assumptions: (1) the normal distribution of the outcomes, $y_{ij}$, and (2) the normal distribution of the structural within-school parameters, $\beta_{jk}$. Assumption (1) will typically be met when standardized tests are used as the outcome measure, since such tests are routinely designed to yield near-normal distributions. The validity of this assumption can be assessed by looking at histograms (or normal probability plots) for $y$ within each school. Assumption (2) has a time-honored tradition, since it is typically made in random-effects ANOVA models. However, its validity in any application is more difficult to assess, since the random effects are not directly observed.

There has been little empirical work on the consequences of violating normal distribution assumptions in HLM, but we suspect that problems are most likely to occur in estimates of the model variances, $\sigma^2$ and $\mathbf{T}$, and in hypothesis-testing applications. This suggests that one should be very cautious in making substantive inferences on the basis of these statistics. When the estimated differences are large (or the probability levels very small, e.g. $p < .001$), statistical inference regarding these distributional assumptions should generally be robust. However, when the estimated differences are marginal (e.g., when $p = .05$), care is required.

The empirical Bayes estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are likely to be much more robust to violation of distributional assumptions, because both estimators can be defended on other grounds. Intuitively, the $\boldsymbol{\beta}^*$ estimators are eminently sensible. The most reasonable way to handle two estimators of an unknown parameter is to weight each proportional to its reliability. This is exactly the $\boldsymbol{\beta}^*$ estimator. The $\boldsymbol{\theta}^*$ is just a generalized least squares estimator whose appropriateness can be defended without resort to distributional assumptions.

Thus, the estimation problems encountered in HLM are similar to those encountered in typical least squares applications. Point estimators require minimal statistical assumptions. The computation of standard errors and hypothesis testing lean more heavily on distributional assumptions. Although the assumption of normal distributions is convenient, since the empirical Bayes results are well established, other distributions can be assumed for both the data and the prior density. The main analytic problem is the derivation of estimates for the variances through EM or some other means.

On balance, it is clear that by requiring distributional assumptions in HLM, we have greatly expanded the range of substantive questions that can be empirically examined. Analysts are no longer bound by very restrictive linear models, which assume that within-school structural relationships are identical across schools. Essentially, HLM requires more statistical assumptions in exchange for fewer substantive assumptions.

On the substantive side, the key assumptions are contained in two statistical requirements:

1. The factors included in the within-school error term, $R_{ij}$, are independent of the variables represented in this model, $X_{ijk}$.
2. The factors included in the between-school error term, $U_j$, are independent of the variables represented in this model, $Z_{pj}$.

These assumptions are identical to those in ordinary regression analysis. Both the within- and between-school model must be appropriately specified—i.e., all confounding variables must be identified and included in each equation. Moreover, misspecification of the within-school model may bias not only the within-school slopes but also the estimated between-school slopes, $\boldsymbol{\theta}$. Similarly, we must assume that the $X$ and $Z$ variables are measured without error. This latter assumption is problematic in many situations. Goldstein (in press) offers one approach that merits further consideration.

How best to estimate $\mathbf{V}_j$ and $\mathbf{T}$ remains an open question. Nevertheless, the availability of at least one workable procedure with desirable analytic properties, the EM algorithm, has already greatly facilitated developments in multilevel modeling. An important strength of the EM approach is its ability to handle unequal sample sizes within units. However, computational procedures for HLM using the EM algorithm are just now becoming available.[1] The

[1] Computing with HLM involves the use of generalized least squares with some external procedure for estimating the covariance components that form the weight matrix. To date, the use of the EM algorithm to estimate the covariance components has required original FORTRAN programming. A program that performs HLM with the EM algorithm, which cur-

limited experience to date suggests that the rate of convergence may be very slow, and more investigation of the behavior of EM with increasingly complex models is needed.

One might reasonably ask whether simpler, more accessible methods will do the same job nearly as well. In fact, HLM reduces to OLS with student-by-school interaction terms if the homogeneity hypotheses tested by (22) and (25) are maintained. When only the intercepts vary significantly, as in model 3 in our illustration, noniterative approaches for variance estimation can be used (cf. Hanushek 1974). In fact, a worthy research goal is to develop models that can explain away the slope parameter variance (Cooley et al. 1981). Of course, the ability to identify such models requires the technology of HLM, which allows us to estimate slope parameter variances and covariances and monitor their reduction as we develop better models.

*Concluding Comments*

Research on school effects has been plagued by both methodological and conceptual problems. In our view, the two are closely related. The available analytic models tend to limit conceptualization to what can be empirically tested through such models. There is a natural hesitancy to form a conceptualization when it remains unclear how to test the fruits of that conceptualization. Thus, HLM is a promising development; it greatly expands the range of methods for investigating schools and thereby extends conceptualization.

Like any analytic approach, HLM has its limitations. Currently, the model provides representation for only two levels, though the necessary statistical theory for expansion to multiple levels exists. Further, there are certain classes of school effects that HLM cannot represent. For example, school factors, rather than mean achievement or slopes, could influence the variability within groups. Such an effect, however, cannot be directly represented within the hierarchical model presented here.

Initially, at least, we expect researchers to encounter difficulty in interpreting the results of an HLM analysis, which are considerably more complex than results of an ordinary linear model. The ensuing technical discussions, however, should not deflect us from primary concerns. Ultimately, Cooley et al.'s (1981) recommendation that we engage in more substantive discussions about the causal models we assume in conducting research on schools remains paramount. By facilitating the explicit modeling of processes that occur both within and between various levels of school organization, HLM analyses can enrich such discussions and advance research on school effects.

rently runs on an AMDAHL mainframe, is available from William Mason, Department of Sociology, University of Michigan. The program we used is written in FORTRAN 77 and currently runs on a Hewlett-Packard 9000 minicomputer. It is available from Anthony Bryk, School of Education, University of Chicago. Harvey Goldstein at the University of London Institute of Education is also developing a general program that performs his Iterative Generalized Least Squares estimation procedure on HLM models.

REFERENCES

Aitkin, M., D. Anderson, and J. Hinde. 1981. "Statistical Modeling of Data on Teaching Styles." *Journal of the Royal Statistical Society, Series A* 144, pt. 4:419–61.

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

———. 1985. "School Sector and Cognitive Performance: When is a Little a Little?" *Sociology of Education* 58:115–28.

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

Bock, R.D. 1983. "The Discrete Bayesian." Pp. 103–15 in *Principles of Modern Psychological Measurement: A Festschrift for Frederick M. Lord*, edited by H. Wainer and S. Messick. Hillsdale, NJ: Lawrence Erlebaum.

Boyd, L.H., and G.R. Iverson. 1979. *Contextual Analysis: Concepts and Statistical Techniques.* Belmont, CA: Wadsworth.

Bryk, A.S. 1981. "Disciplined Inquiry or Policy Argument?" *Harvard Educational Review* 51:497–509.

Bryk, A.S., and S.W. Raudenbush. 1984. "Applying the Linear Hierarchical Model to Measurement of Change Problems." Unpublished manuscript, University of Chicago.

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

———. 1980b. "The Role of Levels of Analysis in the Specification of Educational Effects." Pp. 119–90 in *The Analysis of Educational Productivity, Vol. I: Issues in Microanalysis,* edited by R. Dreeben and J.A. Thomas. Cambridge, MA: Ballinger.

Burstein, L., R.L. Linn, and I. Capell. 1978. "Analyzing Multi-level Data in the Presence of Heterogeneous Within-Class Regressions." *Journal of Educational Statistics* 4:347–89.

Burstein, L., and M.D. Miller. 1980. "Regression-based Analysis of Multi-level Educational Data." *New Directions for Methodology of Social and Behavioral Sciences* 6:194–211.

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

Cooley, W.W., L. Bond, and B. Mao. 1981.

"Analyzing Multi-level Data." Pp. 64–83 in *Educational Evaluation Methodology,* edited by R.A. Berk. Baltimore: Johns Hopkins University Press.

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

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

Dempster, A.P., N.M. Laird, and D.B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm" (with discussion). *Journal of the Royal Statistical Society, Series B* 39:1–8.

Dempster, A.P., D.B. Rubin, and R.D. Tsutakawa. 1981. "Estimation in Covariance Components Models." *Journal of the American Statistical Association* 76:341–53.

Dielman, T.E. 1983. "Pooled Cross-sectional and Time-Series Data: A Survey of Current Statistical Methodology." *American Statistician* 37:111–22.

Efron, B., and C. Morris. 1975. "Data Analysis Using Stein's Estimator and its Generalizations." *Journal of the American Statistical Association* 74:311–19.

Elston, R.C., and J.E. Grizzle. 1962. "Estimation of Time-Response Curves and their Confidence Bands." *Biometrics* 18:148–59.

Goldberger, A.S., and G.G. Cain. 1982. "The Causal Analysis of Cognitive Outcomes in the Coleman, Hoffer, and Kilgore Report." *Sociology of Education* 55:103–22.

Goldstein, H. In press. "Multilevel Mixed Linear Model Analysis Using Iterative Generalized Least Squares." *Biometrika*.

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

Hanushek, E. 1974. "Efficient Estimators for Regressing Regression Coefficients." *American Statistician* 28:66–67.

Harville, D.A. 1977. "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." *Journal of the American Statistical Association* 72:320–40.

Hedges, L.V. 1982. "Fitting Continuous Models to Effect Size Data." *Journal of Educational Statistics* 7:245–270.

Heyns, B., and T.L. Hilton. 1982. "The Cognitive Tests for High School and Beyond: An Assessment." *Sociology of Education* 55:89–102.

Hoffer, T., A.M. Greeley, and J.S. Coleman. 1985. "Achievement Growth in Public and Catholic Schools." *Sociology of Education* 58:74–97.

Keesling, J.W., and D.E. Wiley. 1974. "Regression Models for Hierarchical Data." Paper presented at the annual meeting of the Psychometrics Society, Stanford University.

Kilgore, S.B. 1983. "Statistical Evidence, Selectivity Effects, and Program Placement: Response to Alexander and Pallas." *Sociology of Education* 56:182–86.

Laird, N.M., and J.H. Ware. 1982. "Random-effects Models for Longitudinal Data." *Biometrics* 38:963–74.

Lindley, D.V., and A.F.M. Smith. 1972. "Bayes Estimates for the Linear Model." *Journal of the Royal Statistical Society, Series B* 34:1–41.

Longford, N.T. 1985. "A Fast Scoring Algorithm for Maximum Likelihood Estimation in Unbalanced Mixed Models with Nested Effects." Unpublished manuscript, Institute for Applied Statistics, Lancaster University, Lancaster, England.

Mason, W.M., G.Y. Wong, and B. Entwisle. 1983. "Contextual Analysis Through the Multilevel Linear Model." Pp. 72–103 in *Sociological Methodology 1983–1984,* edited by S. Leinhardt. San Francisco: Jossey-Bass.

McPartland, J.M., and E.L. McDill. 1982. "Control and Differentiation in the Structure of American Education." *Sociology of Education* 55:77–88.

Morris, C. 1983. "Parametric Empirical Bayes Inference, Theory and Applications." *Journal of the American Statistical Association* 78:47–65.

Murnane, R. 1981. "Evidence, Analysis, and Unanswered Questions." *Harvard Educational Review* 51:483–89.

Noell, J. 1982. "Public and Catholic Schools: A Reanalysis of 'Public and Private Schools'." *Sociology of Education* 55:123–32.

Page, E., and T. Keith. 1981. "Effects of U.S. Private Schools: A Technical Analysis of Two Recent Claims." *Educational Researcher* 10:7–17.

Rachman-Moore, D., and R.G. Wolfe. 1984. "Robust Analysis of a Non-linear Model for Educational Survey Data." *Journal of Educational Statistics* 9:277–94.

Rao, C.R. 1972. "Estimation of Variance and Covariance Components in Linear Models." *Journal of the American Statistical Association* 67:112–15.

Raudenbush, S.W. 1984. "Applications of a Hierarchical Linear Model in Educational Research." Ph.D. diss., School of Education, Harvard University.

Raudenbush, S.W., and A.S. Bryk. 1985. "Empirical Bayes Meta-analysis." *Journal of Educational Statistics* 10:75–98.

Rosenberg, B. 1973. "Linear Regression with Randomly Dispersed Parameters." *Biometrika* 60:61–75.

Smith, A.F.M. 1973. "A General Bayesian Linear Model." *Journal of the Royal Statistical Society, Series B* 35:61–75.

Strenio, J.L.F., H.I. Weisberg, and A.S. Bryk. 1983. "Empirical Bayes Estimation of Individual Growth Curve Parameters and their Relationship to Covariates." *Biometrics* 39:71–86.

Swamy, P.A.V. 1973. "Criteria, Constraints, and Multicollinearity in Random Coefficient Regression Models." *Annals of Economic and Social Measurement* 4:429–50.

Wiley, D.E. 1970. "Design and Analysis of Evaluation Studies." Pp. 259–69 in *The Evaluation of Instruction,* edited by M.C. Witrock and D.E. Wiley. New York: Holt, Rinehart, and Winston.

Willms, J.D. 1984. "School Effectiveness within the Public and Private Sectors: An Evaluation." *Evaluation Review* 8:113–35.

———. 1985. "Catholic-School Effects on Academic Achievement: New Evidence from the High

School and Beyond Follow-up Study." *Sociology of Education* 58:98–114.

Wisenbaker, J., and W. Schmidt. 1979. "The Structural Analysis of Hierarchical Data." Paper presented at the annual meetings of the American Educational Research Association, San Francisco.

## APPENDIX: THE LOGIC OF EM ESTIMATION

Under the assumption that $\tau$ and $\sigma^2$ are known, equations (12) and (13) provide empirical Bayes estimates for $\beta_j$ and $\mu_\beta$ for the model specified in (6) and (7). Now, let us suppose that $\beta_j$ and $\mu_\beta$ are known. It can readily be shown that

$$\hat{\sigma}^2 = \frac{\sum_j \sum_i (y_{ij} - \beta_j X_{ij})^2}{\sum_j n_j} \quad \text{(A1)}$$

and

$$\hat{\tau} = \frac{\sum_j (\beta_j - \mu_\beta)^2}{J} \quad \text{(A2)}$$

are maximum likelihood estimates for $\sigma^2$ and $\tau$ respectively. Thus, it is easy to derive maximum likelihood estimates for either the structural parameters, $\beta_j$ and $\mu_\beta$, assuming the variances, $\tau$ and $\sigma^2$, are known, or the variances, assuming the structural parameters are known. The troublesome part is deriving estimates for all these parameters simultaneously. However, the dependence of the structural parameters upon the variances and vice versa provides the basis for the application of the EM algorithm to this problem.

Equations (A1) and (A2) are recognizable as residual mean squares. The EM algorithm requires that we determine the expected value of these mean squares when $\beta_j$ and $\mu_\beta$ are unknown, in terms of the empirical Bayes estimates $\beta_j^*$ and $\mu_\beta^*$. The expectations are taken over the posterior distribution of $\beta_j$ and $\mu_\beta$ given $y$. It can be shown that

$$E(\sigma^2) = \frac{\sum_j \sum_i (y_{ij} - \beta_j^* X_{ij})^2}{\sum_j n_j} + \frac{\sum_j \sum_i X_{ij}^2\{w_j v_j + (1 - w_j)^2 \sum_j [\text{var}(\beta_j^*)]^{-1}\}}{\sum_j n_j} \quad \text{(A3)}$$

$$E(\tau) = \frac{\sum_j (\beta_j^* - \mu_\beta^*)^2}{J} + \frac{\sum_j \left[w_j v_j + w_j^2\left\{\sum_j [\text{var}(\beta_j^*)]^{-1}\right\}\right]}{J}. \quad \text{(A4)}$$

The first components in equations (A3) and (A4) are identical to equations (A1) and (A2) except that we have now substituted the empirical Bayes estimates for the $\beta_j$ and $\mu_\beta$ parameters. Intuitively, we expect that when $\beta_j$ and $\mu_\beta$ are unknown, the sum of squares in equations (A1) and (A2) will be too small, since they fail to take into account the uncertainty associated with the estimation of $\beta_j$ and $\mu_\beta$ from the data. The second components in equations (A3) and (A4) represent adjustments to the mean squares to account for this additional variability.

The actual computations in applying the EM algorithm proceed as follows:

1. Starting values for $\tau$ and $\sigma^2$ are determined. The estimated residual variance pooled from the OLS regressions within each school provides a natural initial value, $\sigma_0^2$, for $\sigma^2$. Recall that the observed variance in $\beta_j$ = var($\hat{\beta}_j$) = $v_j + \tau$. Both $v_j$ and var($\hat{\beta}_j$) are estimable from the data. This suggests that var($\hat{\beta}_j$) $- \bar{v}_j$ can be used as the initial value for $\tau$, $\tau_0$, where $\bar{v}_j = (1/J)(\Sigma_j v_j)$.
2. These starting values are substituted into equations (9), (12), (13), and (14) to yield initial estimates for $\beta_j^*$ and $\mu_\beta^*$.
3. The structural parameter estimates from step 2 are substituted into equations (A3) and (A4) to derive new estimates for $\tau$ and $\sigma^2$.
4. These new variance estimates are substituted back into equations (9), (12), (13), and (14) to yield new values for $\beta_j^*$and $\mu_\beta^*$.
5. The process in steps 3 and 4 iterates back and forth between the estimation of variances and the estimation of structural parameters until convergence.