Measuring School Effects With Hierarchical Linear Modeling: Data Handling and Modeling Issues

Eugene P. Adcock, Ph.D., Prince George's County Public Schools, Maryland
Gary W. Phillips, Ph.D. National Center for Education Statistics, U.S. Department of Education

Because public schools do not randomly assign students and teachers across schools (methodological utopia), multilevel evaluation models which account for student and school contextual and practice variables in their natural settings provide the most rigorous means for empirically showing what is actually happening in school classrooms. Still, no statistical methodology can make up for faulty design or bad data. This article presents some important practical issues regarding data handling for multilevel analysis methodology. Also presented are important modeling design issues that need to be considered when applying hierarchical linear models (HLM) to the measurement of schools and for determining which factors impact the value schools add to students' achievement.

The statistical method chosen for an analysis is usually a function of two things: the question being asked and the nature of the recorded data. In the case of measuring school effects, HLM is a multivariate regression-like analysis technique that was developed specifically for use in school effects research. HLM allows the examination of associations among multi-level, nested data such as students within schools by estimating simultaneous linear equations at the student level within schools and the school level between schools. HLM models explain student and school variation in achievement scores, using both student- and school-level variables as explanatory variables, while accounting for the variance at each level. In the Prince George's County Public School district, the HLM model has been used to rank schools on their contribution to student achievement beyond those associated with student poverty, student mobility and school poverty (i.e., *Value-added Index)*, and HLM was also used to evaluate which factors contribute to the value added by schools (Adcock, 1995; Adcock, 1997).

Despite the tremendous potential for HLM to show how schools are doing and what can be done to make them better, the types of evaluation-quality data necessary to support the different levels of analysis – student, teacher, classroom, school, district — are not supported by the data handling practices of most public school districts. The fact that HLM is a non-experimental design involving the analysis of relationships among variables at multiple levels in the educational system makes the integrity of the data support system critical. Analysis of multilevel data must begin with an understanding of relationships among the lowest level variables, how unbiased higher level variables are constructed from lower level variables, and the relationships among the lower level and higher level variables (Cooley, Lloyd, and Mao, 1981).

After the multilevel evaluation design has been determined (e.g., HLM), the availability of specified student, classroom, school and district level evaluation-quality data is a real-life issue to the practical application to school effectiveness studies. The first section of this paper will address the issue of school district data support for multilevel evaluation designs and the second section will address modeling issues important to the successful application of the HLM model.

Section One

Data handling and data analysis are not distinctly different. Due to the increasing popularity of causal analysis and structural equation models (e.g., LISREL, AMOS, HLM) in school effects studies, the problems inherent in the multilevel nature of educational data are becoming more widely recognized (Bentler and Chou, 1988). School district data management systems and school district evaluation offices need to get in sync with the research, evaluation and accountability needs fulfilled by multilevel analysis models.

The formulation of explicit multilevel models with hypotheses about effects occurring at each level and across levels places important structural features and demands on data. Expressing relationships among variables within a given level, and specifying how variables at one level influence relations occurring at another require a data processing system purposefully designed to support such innovative analysis methods. Because multilevel model analysis requirements of school district data are statistical in nature, it is the responsibility of school district evaluation offices to develop a relational database that can provide hierarchically structured and linked data — students, classrooms, schools, and district — for analysis by these powerful and important multilevel evaluation methods. From the perspective of a school district staff member responsible for fulfilling the data requirements of two large scale HLM school effects studies, the following data handling issues are identified among those which are important to the application of HLM analysis, reporting the nature of the analysis to colleagues, and supporting continued multilevel analysis studies:

1. taking control of variable definitions and parameters in determining the unit of analysis;

2. variable selection and measurement standards for evaluation-quality data vs. colleagues' "wish list" for inclusion of "crucial variables" in the analysis model; and

3. harvesting raw data from school district legacy system sources.

Unit of Analysis

Who is a student? Who is a teacher? What constitutes participation behavior, class size and student instructional cost? What is a program, a treatment, a school? Because one can not analyze below the data level that you observe, record, store and manage, it is vital that the unit of analysis parameters for measured predictor variables are established by statistical staff with a definitive vision and understanding of analysis. Once a plausible causal model has been defined, the structural equations implied by that model determine the appropriateness of a particular data analysis scheme. If the causal models are multilevel (e.g., HLM), then analysis will occur at the different levels for a complete understanding of the teaching and learning phenomena under investigation. In particular, the potential contribution of multilevel analysis is a function of recorded data on each individual's singular characteristics, experiences, behaviors, and achievements. Furthermore, since HLM analysis procedures take both student and school information into account simultaneously, it is important that data representing the same variables between these levels are consistent, linked and stable.

Multilevel evaluation models which account for student and school contextual and practice variables in their natural settings (e.g., HLM) provide a viable means of empirically showing what is actually happening in school classrooms. Students who are highly mobile and schools with highly mobile populations, for example, represent contextual variables which can be represented at both the student level (Student Mobility) and the school level (School Mobility). Likewise, teachers who have service years in a particular school (School Vested) and total service years in the district (System Vested) provide teaching experience information which naturally vary across schools. Rigorous variable specifications must rely upon an understanding of the school system source data structure and multilevel analysis requirements. These specifications enable the appropriate unit of analysis construction for individual student and individual teacher variables which can, in turn, be aggregated to higher classroom, school and district levels yielding consistent and stable estimates at each level.

Table 1 lists operational examples of how the Prince George's school district evaluation office fulfilled the requirements for evaluation-quality variables included in a recent HLM value-added study of 120 elementary schools (Adcock, 1997). This research study had two foci: the effects of personal characteristics and individual educational experiences on student learning, and how these relations are in turn influenced by classroom organization and the specific behavior and characteristics of the teachers within the school. Correspondingly, the data have a two-level hierarchical structure. The Level-1 units are the persons, who are nested within the Level-2 units of schools.

Table 1

School Year 1994-95 (SY95) HLM Value-Added Assessment Study
Partial List of Individual (Level 1) and Elementary School (Level 2) Variables

Variable	*Definition*	*Parameters*
Student (Level 1)	For the value-added study, student is a SY95 *Maryland School Performance Assessment Program* (MSPAP) eligible examinee with at least one scale score in the content areas of reading, mathematics or science.	"Student" is a *Research, Evaluation and Assimilation Database* (READ) warehouse system data element defined as a child who has an assigned PGCPS enrollment date and location, student number, race code and gender code.
Teacher (Level 1)	For the value-added study, elementary school teacher is a "core teacher" who is responsible for delivering the PGCPS curriculum in the six MSPAP test content areas (i.e., mathematics, science, social studies, reading, writing, and language arts).	**Core teacher** is a READ data element representing a school-based certificated "classroom" teacher employed on the last day of the school year and who has the assigned responsibility to provide students instruction and assign course grades in one or more of the core academic subject areas of language (reading, English, etc.), mathematics, science or social studies.
Class Size (Level 2)	The total number of students enrolled on the last day of the school year divided by the number of core teachers employed on the last day of the school year for each elementary school.	"Core Teacher" is a READ-defined data element: See Level 1 definition for "teacher" listed above. "Class Size" is a constructed class student-teacher ratio similar to that used by R. F. Ferguson (1991).
Teacher College Training (Level 2)	The average academic training index of the core teachers in a school. Seven point scale: 1=Bachelors, 2=Bachelors+30 course credit hours (cch), 3=Masters/Equivalent, 4=Masters+15(cch), 5=Masters+30(cch), 6=Masters+60(cch), 7=Doctorate.	Computed from the sum of teacher college training index divided by the number of core teachers employed on the last day of the school year for each elementary school.
Teacher Cost Per Student (Level 2)	The average salary of the core teachers employed at end-of-year (EOY) multiplied by the number of classroom teachers assigned (= the budgeted[1] number or the actual number of core teachers observed, whichever greater) divided by the total number of students enrolled in school at EOY for each elementary school.	Permanent teachers who are replaced by long-term substitute teachers at EOY required the following correction for computing the school's teacher salary (numerator): The average salary of the observed permanent core teaching staff is multiplied by the number of budgeted core teachers in each school.

1 Pupil Accounting and School Boundary "*Class Size Report: 1994-95.*"

Enrollment Mobility: School (Level 2)	The average total number of days that SY95 *Maryland School Performance Assessment Program* (MSPAP) examinees were NOT enrolled in the school in which they began taking the SY95 MSPAP test for the past 3 years (SY93-SY95) based upon their most recent occurrence of continuous enrollment in that school.	Only the last continuous enrollment period is considered. No school transfers after the start of MSPAP administration date are considered. Continuous school enrollment (i.e., 0 Mobility) for 3 years is 540 days (i.e., 180 * 3 years) for the MSPAP school.
Enrollment Mobility: System (Level 2)	The average total number of days that SY95 MSPAP examinees were NOT enrolled in the PGCPS system for the past 3 years (SY93-SY95) dating back from the start of MSPAP administration date.	Note: continuous system enrollment (i.e., 0 Mobility) for 3 years is 540 days (i.e., 180 * 3 years) for any combination of schools in the system.
Teacher Service Years at MSPAP School (Level 2)	The average total number of years that core teachers employed at SY95 MSPAP schools "belonged" to that school based upon their most recent occurrence of continuous employment in that school.	Only the last continuous "belonging" period is considered.
Teacher Service Years in PG System (Level 2)	The average total number of years that core teachers have been employed as certified teachers in the PGCPS system based upon their most recent occurrence of continuous employment in the system.	Only the last continuous "belonging" period is considered.
% of MSPAP Examinees African-American (Minority) (Level 2)	The proportion of the total SY95 MSPAP examinee population who are African-American for each elementary school.	School aggregate means of Minority = 1 and Other = 0 are actually proportion values of study students who are African-Americans.
% Poverty Among MSPAP Examinees (Level 2)	The proportion of the SY95 MSPAP examinee population who are receiving a free or reduced lunch.	School aggregate means of Poverty = 1 and Non-Poverty = 0 are actually proportion values of study students who are eligible for Free/Reduced meal program.
% of MSPAP Examinees TAG (Level 2)	The proportion of the total SY95 MSPAP examinee population who are identified as "talented and gifted" by the TAG Office.	
Teacher Days Absent in SY95 (Level 2)	The proportion of days the core teachers employed at end-of-year (EOY) were absent during SY95 for each elementary school.	Computed from sum of teacher days absent divided by sum of days "belonging" to school for all end-of -year (EOY) core teachers.
Teacher Salary (Level 2)	The average core teacher salary in a school.	Computed from the sum of the teacher salary, divided by the number of core teachers at EOY in a school. SY95 "A" Scale Tables used for salaries.
Achievement Test Scale Score in Reading, Mathematics and Science (Level 2)	The school's average unweighted third and fifth grade student performance for SY95 MSPAP reading, mathematics and science content areas.	A few elementary or "combination" schools did not have both third and fifth grade levels. Cases deleted from content area school aggregation if missing test scale score.

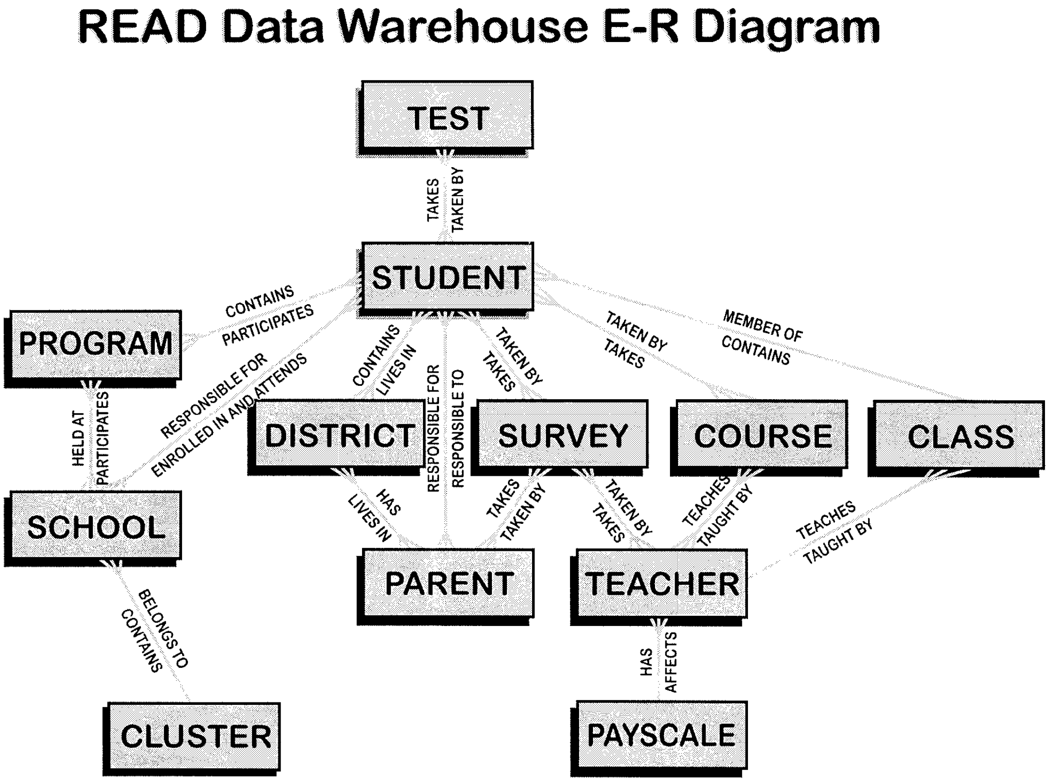

Figure 1

Data Entities And Their Relationships In The READ Warehouse

As can be seen from the list of variable in Table 1, selection of variables for this HLM research study was not limited to "available and easy" but included factors cited in school effects literature and by school policy members as important contributors to teaching and learning. Table 1 lists Level 1 variable definitions for student and teacher, and several Level 2 school aggregate variables used in a recent HLM study. The variable definitions and parameter specifications are also shown. Since Level 1 variables for individual characteristics, behaviors and achievements (e.g., Student SES, Student Mobility, and Teacher Training) are used to build Level 2 aggregate variable values, the Level 1 variables beyond "student" and "teacher" used in the study were omitted from the list because the reader can easily deduce the concomitant Level 1 definitions and specifications from those listed for Level 2.

Evaluation Variables vs. "Wish List"

You cannot analyze what you do not measure. It is around the conference table where evaluation study results are being presented that evaluation staff often learn from colleagues of the plethora of programs and initiatives which "explain everything!" but are missing from the causal evaluation model. For example, where are the: students' beginning achievement levels, gain scores, teacher inservice, Saturday Academies, parent participation, computer labs, dimension of learning instructional practices, content certified teachers, extra resource teachers,...,etc. in the multilevel model analysis? After all, schools are implementing one great thing after another great thing, and there is no measurement of these great things in the analysis model! Actually, there is no evaluation-quality measurement of these great practices at all, otherwise they would be in the analysis model. Statisticians have not been known to shy away from any available evaluation-quality data that may correlate with student achievement.

As presented in the previous section, quality standards for evaluation analysis data must meet rigorous specifications. The evaluation-quality measurement standards include unit of analysis issues for case selection, assessment, scaling and recording. Often these evaluation-quality measurement standards are very difficult to achieve for many of the innovative practices, activities, experiences, and resources implemented by program staff and put forth as correlates of observed student achievement. In fact, when the measurement specifications are delineated for the inclusion of these practices (e.g., teacher inservice), program staff often find them too confining, burdensome, and in some cases menacing. Still, it is the responsibility of the school district evaluation office to provide guidance to staff interested in carrying out evaluation-quality measurement of their program's contribution to the value-added effects of schools.

Harvesting Data From Legacy Sources

" We do so much testing and surveys, plus filling out tons of data forms; how come we don't have any data for evaluating this program, that initiative or these schools?" With respect to research, the choice of data to analyze, debugging and preparation methods, management and storage procedures, and data layout is an act of theoretical preference (Davidson, 1996). A scientifically rigorous approach for research, evaluation and accountability has inherent data handling standards which sometimes render locally developed and administered data gathering information inadequate for evaluation purposes. Still, a database support system which can transform much of a school district's operational system data (e.g., course schedules, grades, tests, attendance, teacher service years, etc.) into a database system which meets the structural and statistical evaluation data standards for multilevel school and program evaluation studies is an indispensable tool for school district evaluation. In response to this vital need for pro-actively prepared evaluation-quality extant data on students, teachers, program/school participation measures, and resources the Research, Evaluation and Accountability staff of the PGCPS system has developed the *Research and Evaluation Assimilation Database (READ)* warehouse support system (Adcock, Haseltine, & Winkler, 1997). This comprehensive relational school district data warehouse model, READ, provides detailed achievement data together with contextual and process information at the various levels of students, classroom, teacher and schools. READ is well-suited for supporting scientifically rigorous multilevel HLM evaluation studies of student and school correlates with student achievement.

The READ data collection scheme focuses on collecting data for the following **five core database entities**: student, teacher, school, program and instructional finance. The READ warehouse sequential data processing procedures require data "scrubbing" for all incoming data. Scrubbing is a data warehouse term that includes the integration of legacy data from multiple sources and reformatting as necessary to ensure completeness, consistency, and accuracy. In addition, scrubbing data to evaluation requirement specifications often involves enhancement or derivation processing, partitioning and summarization of newly acquired legacy data. Transforming legacy data into evaluation-quality data is given such importance that the READ data warehousing pipeline has dedicated substantial resources to data verification, documentation, scrubbing and enhancement activities.

The design of the READ data warehouse follows logical relational database design with subject areas and their relationships. Figure 1 shows the Entity/Relationship Diagram (ERD) for the READ System's data warehouse.

In READ all input data is initially kept at the individual student (or teacher) level, and then aggregated at higher levels to meet complex evaluation data needs of multilevel analysis. Two-level HLM analysis, for example, may require the extraction of READ student level data for achievement, socio-economic status, ethnicity, etc., and school level data on teacher academic training, cost per student, mobility of student population, etc. The READ warehouse method of collecting, managing and extracting data permit this type of evaluation of the real-life multi-level nature of school district structure to be conducted. The next section describes some of the fundamental issues associated with modeling HLM analysis.

SECTION 2

This section is intended to provide researchers with the basic understanding of several statistical fundamentals of hierarchical linear models (HLM). We will introduce the HLM model, discuss centering, the estimation of school effects, and the empirical

Bayes estimation procedure. Along the way we will provide some practical advice in several other areas.

Simple versions of the HLM

To facilitate understanding we will illustrate all points with a simple 2-level HLM model with only one independent variable at both the student and school levels. We will also adopt the widely used notation provided by Bryk and Raudenbush (1992). At level I the dependent variable, Y_{ij}, will be math achievement and the independent variable, X_{ij} , will be socio-economic status (SES). At level II the independent variable will be the mean SES for school j, W_j.

Level I

$$Y_{ij} = ß_{0j} + ß_{1j} (X_{ij}) + r_{ij}, \qquad (1)$$

where

Y_{ij} = math achievement for student i in school j,

$ß_{0j}$ = expected math achievement. $ß_{0j}$ is an adjusted mean for school j such that

$$ß_{0j} = \mu_{y\,j} - ß_{1j} (X_{ij}),$$

$ß_{2j}$ = expected change in math achievement a unit change in X, and

r_{ij} = residual for student i in school j.

Level II

$$ß_{0j} = \gamma_{00} + \gamma_{01} (W_j) + \mu_{0j}, \qquad (2)$$

where

γ_{00} = predicted grand mean for math achievement for all schools based on W,

γ_{01} = change in expected school mean achievement ($ß_{0j}$) for a unit change in W,

μ_{oj} = unique effect of school j on the expected school achievement after controlling for W.

$$ß_{1j} = \gamma_{10} + \gamma_{11} (W_j) + \mu_{1j}, \qquad (3)$$

where

γ_{10} = SES slope for all schools.

γ_{11} = change in SES slope ($ß_{1j}$) for a unit change in W,

μ_{1j} = unique effect of school j on the SES slope after controlling for W.

At level I we make the assumptions that $E(r_{ij}) = 0$, and $Var(r_{ij}) = \sigma^2$. At level II we assume $E(\mu_{0j}) = E(\mu_{1j}) = 0$, $Var(\mu_{0j}) = \tau_{00}$, $Var(\mu_{1j}) = \tau_{11}$, $Cov(\mu_{0j},\mu_{1j}) = \tau_{01}$, and $Cov(\mu_{1j}, \mu_{0j}) = \tau_{10}$.

There are two important statistics that are based on these variances and covariances. The first is the intra-class correlation coefficient, p, (which indicates the overall degree of clustering within schools)

(1)

$$p = \tau_{00}/(\tau_{00} + \sigma^2), \qquad (4)$$

and the second is the reliability with which μ_{0j} is etimated by the ordinary least-squares estimate (OLS) $Y_{\cdot j} - ß_{1j} (X_{\cdot j})$ within each school

$$\lambda_j = \tau_{00}/(\tau_{00} + \sigma^2/n_j). \qquad (5)$$

The above first three equations can be expressed as a single level I model
by substituting equations 2 and 3 into 1. This yields the reduced form of the HLM model as follows

$$Y_{ij} = [\gamma_{00} + \gamma_{01} (W_j) + \mu_{0j}] + [\gamma_{10} + \gamma_{11} (W_j) + \mu_{1j}] (X_{ij}) + r_{ij}. \qquad (6)$$

As a general rule the coefficients of the level I model are treated as random while the level II (or the highest level in the model) are treated as fixed. Treating a level I coefficient as random indicates that the coefficient varies across schools (or level II units). One good way to better understand the HLM is to contrast it with other simpler models frequently used by education researchers. A number of commonly used simpler models can be obtained from equations 1-3 by fixing the level I parameters. For example, if there are no level I or level II independent variables then equation 1 becomes

$$Y_{ij} = ß_{0j} + r_{ij}, \quad (7)$$

and equation 2 becomes

$$ß_{0j} = \gamma_{00} + \mu_{0j}. \quad (8)$$

Substituting equation 8 into 7 yields

$$Y_{ij} = \gamma_{00} + \mu_{0j} + r_{ij}, \quad (9)$$

which is the one way analysis of variance (ANOVA) model. Another often used model can be derived from equations 1-3 by assuming no level II independent variable, and assuming that the $ß_{0j}$ at level I are fixed. When this is the case then equation 1 becomes

$$Y_{ij} = ß_{0j} + ß_{1j} (X_{ij} - X..) + r_{ij}, \quad (10)$$

equation 2 becomes

$$ß_{0j} = \gamma_{00} + \mu_{0j}, \quad (11)$$

and equation 3 becomes

$$ß_{1j} = \gamma_{10}, \quad (12)$$

which is the pooled within-school regression coefficient. Substituting equation 11 and 12 into 10 yields

$$Y_{ij} = [\gamma_{00}] + [\gamma_{10}](X_{ij} - X..) + \mu_{0j} + r_{ij}. \quad (13)$$

which is the analysis of covariance (ANCOVA) model (except for the fact that μ_{0j} is random instead of fixed).

Centering

Notice that in equation 8, X_{ij} was centered around the grand mean. In fact it is important to spend some time to be sure that the centering (especially at level I) is done in such a way that the interpretations of $ß_{0j}$ and γ_{00} are meaningful. There are essentially three ways to center at level I (uncentering, grand mean centering, and group mean centering) and two ways to center at level II (uncentering and grand mean centering). At level II, group mean centering and grand mean centering are really the same thing. Centering at level I determines the meaning of the Level I intercept and centering at level II determines the meaning of the level II intercept. In all cases the interpretation of the intercept is that it is the value of the dependent variable when the independent variable equals zero. In the following section we will only discuss centering at level I since the same interpretations apply to level II.

Uncentering

When X_{ij} is uncentered it means we wish to use the zero point in the original metric of X_{ij} as the defining point for $ß_{0j}$. In many areas of science the natural zero of X_{ij} has a practical interpretation. For example, if X_{ij} is the Celsius scale and Y_{ij} is the barometric pressure, then $ß_{0j}$ equals the barometric pressure when water freezes. In most situations in the social sciences there is not a natural zero point for X_{ij}. One notable exception to this is when dummy variable coding is used. For example, if $X_{ij} = 1$ for minority students and $X_{ij} = 0$ for non-minority students, then, $ß_{0j}$ equals the mean of Y_{ij} for non-minority students. If another dummy variable, Z_{ij}, is added to the level I equation, such as gender (where $Z_{ij} = 1$ for females and $Z_{ij} = 0$ for males), then $ß_{0j}$ equals the mean of Y_{ij} for non-minority males.

Group Mean Centering

In the social sciences the group mean of X_{ij} is often used as the zero point for X_{ij}. In group mean centering, $ß_{0j}$ equals the student's math achievement when $(X_{ij} - X_{.j})$ equals zero(which is at the group mean of X_{ij}). For example if X_{ij} is the SES of students, and Y_{ij} is the student's math achievement, then, $ß_{0j}$ equals the student's math achievement at the mean of SES. Another characteristic of group mean centering is that $ß_{0j}$ is always equal to the mean of Y_{ij}, or μ_{Yj}. Therefore, group mean centering is often used as the method of choice when the researcher is primarily interested in studying the variation in school means.

Grand Mean Centering

In the social sciences it is also common practice to center around the grand mean. An example of this was used in the above ANCOVA equation 13. In grand mean centering, $ß_{0j}$ equals the student's math achievement when (X_{ij} - X..) equals zero (which is at the grand mean of X_{ij}). $ß_{0j}$ has a different interpretation in grand mean centering than it does in centering within groups. In grand mean centering $ß_{0j}$ is an adjusted mean such that $ß_{0j} = \mu_{Yj} + ß_{1j} (X_{ij} - X..)$. Grand mean centering is often used when the researcher is interested in estimating school effects.

Estimating School Effects

One of the main uses of HLM is to provide an index of school effectiveness. Once the school effects have been estimated then the researcher can rank schools on their effectiveness or use the effectiveness index as a dependent variable to investigate school factors that are related to effectiveness. A good example of school effects can be derived from the simple HLM model provided in equations 1-3. We rewrite these equations under the assumption that we use grand mean centering and for level I the intercept is random but the slope is fixed (i.e., constant across schools). Under these assumptions, equations 1-3 become

$$Y_{ij} = ß_{0j} + ß_{1j} (X_{ij}) + r_{ij}, \quad (14)$$

$$ß_{0j} = \gamma_{00} + \gamma_{01} (W_j) + \mu_{0j}, \quad (15)$$

$$ß_{1j} = \gamma_{10}. \quad (16)$$

Equations 14-16 are similar to the ANCOVA model except we have added W_j at level II. Substituting equations 15 and 16 into 14 yields the following reduced form

$$Y_{ij} = [\gamma_{00} + \gamma_{01} (W_j) + \mu_{0j}] + [\gamma_{10}](X_{ij} - X..) + r_{ij}.$$

Rearranging terms provides

$$\mu_{0j} = Y_{ij} - [\gamma_{00} + \gamma_{01} (W_j) + \gamma_{10} (X_{ij} - X..) + r_{ij}].$$

Averaging over student i within school j gives the estimate of school effects

$$\mu_{0j} = Y._j - [\gamma_{00} + \gamma_{01} (W_j) + \gamma_{10} (X._j - X..)]. \quad (17)$$

Empirical Bayes Estimation

In HLM the level I coefficients are usually estimated with an empirical Bayes procedure (Lindley and Smith, 1972). This procedure is different from the OLS used in most multiple regression procedures in that the level I estimates are weighted by the collateral estimates in level II. An example of this is found by inspecting more closely equations 1 and 2. We see that $ß_{0j}$ in equations 1 and 2 has two different OLS estimates, $ß^*_{0j}$

$$ß^*_{0j} = Y._j - ß^*_{1j} (X._j) \text{, and}$$

$$ß^*_{0j} = \gamma^*_{00} + \gamma^*_{01} (W_j).$$

The empirical Bayes estimate combines these two OLS estimates by weighting them according to the reliability, λ_j, of $[Y._j - ß^*_{1j} (X._j)]$ as an estimate of $ß_{0j}$. The empirical Bayes estimate, $ß^{**}_{0j}$, is found by

$$ß^{**}_{0j} = \lambda_j[Y._j - ß^*_{1j} (X._j)] + (1-\lambda_j)[\gamma^*_{00} + \gamma^*_{01} (W_j)]. \quad (18)$$

This approach was first introduced within the context of psychometrics by Kelley (1927). The weight λ_j is found by equation 5, and understanding this weight is key to appreciating the usefulness of the empirical Bayes estimation in HLM. As the reliability of the OLS estimate at level I approches unity, the best estimate of the within-school is from the data collected from within the school. However, as the reliability approaches zero (as when the number of students within the school is very low), then the best estimate of the within-school regression parameter is based on the regression parameters of similar schools within the system. The logic of this approach is identical to imputation in a survey sampling context. When data elements are missing for a school, a common practice is to substitute (or impute) data elements from similar schools to replace the missing value. Even treating the data as missing is the same as assuming that the missing data element is equal to the mean of the population.

The empirical Bayes estimate is an optimal estimate of $ß_{0j}$ in the sense that it has the smallest mean-squared error even though it is biased toward $\gamma^*_{00} + \gamma^*_{01}(W_j)$. The amount of bias is inversely related to λ_j. As a general rule the bias is negligible in schools with large sample sizes.

The empirical Bayes residual, μ^{**}_{0j}, is usually used by HLM researchers as the estimate of the school effect. Like the empirical Bayes estimate, μ^{**}_{0j}, is particulary biased in small schools. The relationship between the empirical Bayes residual and the OLS residual is as follows

$$\mu^{**}_{0j} = \lambda_j \mu^*_{0j}. \quad (19)$$

As λ_j approaches zero, μ^{**}_{0j} also approaches zero. Even though the empirical Bayes residual is biased it is still considered by most educational researchers to be a better estimate than the OLS residual. This is because when the sample sizes are small the OLS residual will be unstable resulting in more chance occurences of extreme valus of μ^*_{0j}. Selecting out such extreme values of μ^*_{0j} for praise or blame will result in more false-positives and false-negatives than the empirical Bayes residual.

Authors' Note:
The discussion in this paper represents the views of the authors and does not represent those of the U.S. Department of Education.

References

Adcock, E. P. (1995). *Value-Added Effective Schools Study for Elementary Schools: 1994 Maryland School Performance Assessment Program Results*, Research, Evaluation & Accountability, Prince George's County Public Schools, MD., Research Report: 36-9-95.

Adcock, E. P. (1996). *Value-Added Assessment Study: Prince George's County Elementary Schools' 1995 MSPAP Results in Reading, Mathematics and Science*, (2nd Draft), Research, Evaluation & Accountability, Prince George's County Public Schools, MD., Research Report: 56-12-96.

Adcock, E. P., Haseltine, R., Winkler, L. H. (1997). *A Systemic Processing Model For Accountability*, Paper presented at the Panasonic Foundation Annual Partnership Conference IX, Colorado Springs, CO., March 1-3, 1997, Research Report: 60-3-97.

Arnold, C. L., (1995) *Using HLM and NAEP Data to Explore School Correlates of 1990 Mathematics and Geometry Achievement in Grades 4, 8, and 12: Methodology and Results*, National Center For Education Statistics (NAEP), Research and Development Report, Jan., 1995.

Bentler, P. M., & Chou, C. P. (1988). *Practical Issues in Structural Modeling*, Chapter 7 of *Common Problems/Proper Solutions: Avoiding Error in Quantitative Research*, Edited by J. Scott Long, Sage Publications, Newbury Park, CA.

Bryk, A.S., & Rondenbush, S.W., *Hierarchical Linear Models: Application and Data. Analysis Methods*, Sage, 1992.

Bryk, A. S., & Weisberg, H. I. Value-added analysis: A dynamic approach to the estimation of treatment effects. *Journal of Educational Statistics*, 1976, 1, 127-155.

Cooley, W. W., Bond, L., Mao, B. J., (1981). *Analyzing Multilevel Data*, Chapter 3 of book entitled: *Educational Evaluation Methodology: The State Of The Art*, Edited by Ronald A. Berk, The Johns Hopkins University Press, Baltimore, MD.

Davidson, F. (1996). *Principles Of Statistical Data Handling*, Sage Publications, Inc., Thousand Oaks, CA.

Kelley, T.L. *The Interpretation of Educational Measurements*, New York: World Books, 1927.

Lindy, D.V. and Smith, A.F.M., *Bayes Estimates From the Linear Model, Journal of the Royal Statistical Society*, Series B, 1972, 34, pp. 1-41.

Phillips, G. W., & Adcock, E. P., (1997). Practical Applications of Hierarchical Linear Models to District Evaluations, *Journal of Multi-Linear Regression Viewpoints (*Spring, 1997).

Phillips, G. W., & Adcock, E. P., *Using Hierarchical Linear Models to Evaluate Schools*, Paper presented at American Educational Research Association, 1996 Annual Meeting, NY, Presentation 45.13, *Business Meeting/HLM Discussion/Symposium*, "Multiple Linear Regression: The General Model" (4/11/96).

Raudenbush, S. W., & Willms, J.D. The estimation of School Effects, *Journal of Educational and Behavioral Statistics,* Winter 1995, Vol. 20, No. 4, pp. 307-335.