24 Multilevel Models in Design Research
A Case from Mathematics Education

Finbarr C. Sloane
Arizona State University

Introduction

Because of the inherently hierarchical nature of many learning environments, data collected in these environments are nested in structure. More specifically, students work in dyads or groups. These groups are nested in classrooms, classrooms in schools, and schools in local cultures and school districts. Design researchers working in field settings build theory and design products to support learning in such environments. Consequently, design researchers must deal constantly with data structures of the type described here. Although much design research is qualitative in nature, this chapter takes a quantitative perspective and describes how quantitative researchers have begun to deal with nested data structures and the complexities of building theory and drawing inferences when data have this nested structure. In this chapter, the hierarchical linear model (HLM) is described and it is shown how this model provides a conceptual and statistical mechanism for investigating simultaneously how phenomena at different levels interact with each other. In so doing, the aggregation concerns raised by Saari (this volume) are addressed.

In his insightful chapter in this volume, Saari raises a number of issues that mathematics education researchers (in particular) and design researchers (in general) need to be wary of as they study data that have a nested or hierarchical structure. Of central concern in Saari’s discussion is the issue of data aggregation. One purpose of this chapter is to demonstrate a particular resolution for the aggregation problem in hierarchically-ordered data structures known as HLMs. This chapter will show why these models are important for design researchers as they try to build theories with data from individual students and the environments in which they learn. In taking this quantitative approach, design researchers can construct and test stronger, more theoretically valid, inferences that are free from the aggregation and ecological issues raised by Saari.

In schools, classrooms, and other learning environments (e.g., after-school settings), students are often placed in groups for the purposes of play and learning (whether the groups are dyads or larger learning groups). How to investigate these hierarchically-ordered structures, where students are nested in learning environments, has been a concern for educational, psychological and sociological researchers since the middle of the twentieth century (Robinson, 1950). The study of hierarchy has been of intellectual interest in a number of disciplines for quite some time. For example, researchers in sociology (Mason et al., 1983), economics (e.g., Hanushek, 1974), education (e.g., Bryk & Raudenbush, 1992; Burstein, 1980; Murchan & Sloane, 1994), biology (e.g., Laird & Ware, 1982), and statistics (e.g., Longford, 1989, 1993), have all discussed the issues and solutions to problems generated when data are drawn from hierarchically-ordered
systems. From a design science perspective, two themes have dominated these discussions: (a) issues about the aggregation of data, and (b) how to investigate relationships between variables residing at different hierarchical levels (see Saari, this volume).

With regard to aggregation, discussions have focused on whether it is appropriate to aggregate data and on the types of inferences that can be made from aggregated data (see Burstein, 1978; Firebaugh, 1978; Hannan & Burstein, 1974; Robinson, 1950). These discussions, as they relate to theory building, can be found in the learning science literature (see Barab & Kirshner, 2001), and in statistics (Goldstein, 1995; Hannan & Burstein, 1974; Raudenbush & Bryk, 2002) have examined the theoretical and statistical issues associated with the use of aggregate measures to draw inferences about higher-level units.

Although aggregation issues are certainly an important aspect of design research, they are not the focus here. The aim in this chapter is to address the second major theme: namely, how to investigate relationships between variables located at different levels of a learning hierarchy (see Bryk & Raudenbush, 1992) while avoiding the issues attendant to aggregation (for example, ecological fallacies; Robinson, 1950). These inferential fallacies occur, as Saari (in this volume) hints, when the researcher observes a relationship at one level of analysis, say the group, and infers incorrectly that the relationship will hold in the same way at another level of analysis, say the individual student (Murchan & Sloane, 1994).

**Relationships that Cross Hierarchical Levels**

Given the structure of learning environments, it is clear that variables at one hierarchical level (e.g., the group) can and do influence variables at another hierarchical level (e.g., individuals). Numerous theoretical discussions and empirical investigations have identified relationships between variables that reside at different levels. For example, Cobb and Yackel’s (1996) work on the development of sociomathematical norms looks at these relationships from the other direction—from the bottom up. Sloane (2005) has argued the need for a multilevel theory in educational research that goes beyond single-level perspectives and blends both psychological and sociological theoretical lenses with the mutual goals of improved theory and increased capacity to scale designed interventions in real-world settings. In other words, in the scaling of educational interventions in mathematics we need a multilevel theory of implementation that accounts for and maps better to the learning settings in which students find themselves (Fishman et al., 2004).

**Hierarchical Data: Three Possible Options**

Following on the theoretical work of Yackel and Cobb (1996), Sloane (2005) argued that researchers in education need to investigate variables that span multiple levels of analysis. Thus, to study individual behavior in learning environments, one must measure and integrate individual attributes of students and also salient aspects of the environment in which they are performing (Lave & Wenger, 1991). Similarly, in order to investigate the structure of learning environments as a whole, one needs to measure the attributes of the learning settings as well as the organizational environments that support them. For example, we need to investigate students’ learning both in groups and in classrooms that support such group learning. In either case, the resulting data will include variables that reside at different levels of analysis (i.e., variables describing the lower level units as well as the higher-level contexts). Typically, researchers are
interested in investigating both lower-level and higher-level influences on a lower-level outcome variable. This type of investigation has been referred to as cross-level in nature (Sloane, 2005).

In cases where variables exist at more than one level of analysis (e.g., a lower-level outcome and both lower-level and higher-level predictors), there are three main options for data analysis. First, one can disaggregate the group-level data such that each lower-level unit is assigned a score representing the higher-level unit within which it is nested. The data analysis for this option, therefore, would be based on the total number of lower-level units included in the study and represents a traditional psychological approach with an emphasis on individual differences. For example, all individuals might receive a score representing their classroom’s sociomathematical norms, with the investigation centered on the relationship between such norms and individual students’ beliefs about mathematics. The problem with this solution is that multiple individuals are in the same learning group and, as a result, are exposed to similar stimuli in the group. Thus, one cannot satisfy the independence of observations assumption that underlies traditional statistical approaches (Raudenbush & Bryk, 2002). Even when two students share the same classroom we should not infer that each has equal access to the same instructional resources as would be implied in this psychological model setting. In addition to violating this assumption, the disaggregation approach results in another problem. Statistical tests involving the variable at the higher-level unit are based on the total number of lower-level units (e.g., the effect of the group’s cohesion is assessed based on the number of individuals, and not the number of groups). This practice underestimates the standard errors and raises questions about the associated statistical inferences (Bryk & Raudenbush, 1992).

The second major approach is to aggregate the lower-level units and investigate the relationships at the aggregate level of analysis. This we consider a traditional sociological perspective. For example, one could investigate the relationship between group characteristics and individual outcomes by aggregating the individual outcomes to the group level. This parallels the process followed by McClain and her colleagues in their qualitative studies of developmental research in mathematics education (McClain et al., 1996). From a statistical perspective, the disadvantage of this approach is that potentially meaningful individual-level variance in the outcome measure is ignored. Theoretically, one is limited to conceptualizations of the group structure only; that is, no cross-level inferences can be drawn. Sensibly, McClain and her colleagues infer correctly to group characteristics only. However, interactions (rich or otherwise) across levels are not evaluated in this research tradition. In summary, the traditional choice has been between a disaggregated model that violates statistical assumptions and assesses the impact of higher-level units based on the number of lower-level units, or an aggregated model that discards potentially meaningful, lower-level variance. Neither of these two options, the simple psychological approach or the solely sociological approach, is fully satisfactory.

HLMs represent the third major approach to dealing with hierarchically-nested data structures. Statistically, these models are designed specifically to overcome the weakness of the disaggregated and aggregated approaches discussed above. First, these models recognize explicitly that individuals in a particular group may be more similar to one another than individuals in other groups and, therefore, may not provide independent observations. These multilevel statistical approaches explicitly model both individual and group residuals, thus recognizing the partial interdependence of individuals in the same group (this is in contrast to ordinary least-squares [OLS] regression approaches where individual and group residuals are not estimated separately). Second, these
Hierarchical Linear Models

A Framing

As noted above, one of the primary advantages of HLMs is that they allow one to investigate simultaneously relationships within a particular hierarchical level, as well as relationships between or across hierarchical levels. In order to model both the within-level and between-level relationships, the researcher needs to estimate two models simultaneously: one that models relationships within each of the lower-level units and another to model how these relationships within units vary between units. This type of two-level, modeling approach defines HLMs (Bryk & Raudenbush, 1992).

Conceptually, HLMs are relatively straightforward. For clarity, we refer to the two levels here as individuals and groups; however, the methods apply to any situation in which there are lower-level units nested within higher-level units. These models adopt the following two-level approach to cross-level investigations where the level-one model is estimated separately for each group. Typically, this takes the form of a regression-based model such as:

\[
\text{Level 1: } Y_{ij} = B_{0j} + B_{1j} X_{ij} + e_{ij}
\]

where:

- \(Y_{ij}\) = the outcome measure for individual \(i\) in group \(j\)
- \(X_{ij}\) = the value of the predictor for individual \(i\) in group \(j\)
- \(B_{0j}\) and \(B_{1j}\) = intercepts and slopes estimated separately for each group (as noted by the subscript \(j\))
- \(e_{ij}\) = the residual

An example consisting of several different groups will illustrate the nature of these equations. When separate regression equations are estimated for each group, four different patterns can emerge. Figures 24.1(a), (b), (c), and (d) present these four possible options. In Figure 24.1(a), each of the groups in the sample has identical regression lines. Therefore, each group has identical intercepts and slopes. In Figure 24.1(b), the groups still have identical slope terms; now, the intercept terms vary significantly across the groups. Thus, even though the relationship between \(X_{ij}\) and \(Y_{ij}\) is equivalent across the groups, the initial “location” (i.e., the intercept) of this relationship varies across the groups. In Figure 24.1(c), the groups have similar intercept terms, but the relationship between \(X_{ij}\) and \(Y_{ij}\) varies significantly across the groups. In Figure 24.1(d), both the
initial location and the relationship between $X_{ij}$ and $Y_{ij}$ vary significantly across the groups (i.e., both the intercepts and the slopes vary across the groups).

Three of these figures display systematic patterns or differences across the groups (i.e., Figures 24.1(b), (c) and (d)). These differences raise the question of whether there are group-level variables associated with the variation across the groups. For example, group-level variables may be associated with varying intercepts in Figures 24.1(b) and (d) and varying slopes in Figures 24.1(c) and (d). This is precisely the question that the Level 2 analysis in HLMs answers. The Level 2 analysis uses the intercepts and slopes from the Level 1 analysis as dependent variables. For example, a typical, Level 2 model may take the following form:

\[
\text{Level 2a: } B_{0j} = \gamma_{00} + \gamma_{01}Z_j + U_{0j} \\
\text{Level 2b: } B_{1j} = \gamma_{10} + \gamma_{11}Z_j + U_{1j}
\]

where:

- $Z_j$ = a group-level variable
- $\gamma_{00}$ and $\gamma_{10}$, the second-level intercept terms
- $\gamma_{01}$ and $\gamma_{11}$ = the slopes relating $Z_j$ (the group level variables) to the intercept and slope terms from the Level 1 equation
- $U_{0j}$ and $U_{1j}$ = the Level 2 residuals

Depending on the pattern of variance in the Level 1 intercepts and slopes, different Level 2 models would be required. For example, in situations such as Figure 24.1(b), where there is no variance in the slope parameter, the inclusion of $Z_j$ in Equation 3...
would not be meaningful given that $B_{1j}$ is identical for all the groups. Similarly, in situations like Figure 24.1(c), where there is no intercept variance, the inclusion of $Z_j$ in Equation 2 would not be very meaningful because there is no variance in $B_{0j}$ across the groups.

The three equations presented above are not new approaches to investigating relationships occurring across hierarchical levels. A quarter of a century ago, Burstein (1980) discussed this same type of approach under the general label of “slopes-as-outcomes.” Conceptually, this is a very insightful description because the regression parameters (i.e., the intercepts and slopes) estimated for each group at Level 1 are used as outcome measures (i.e., dependent variables) in the Level 2 model.

Although the conceptual approach has been understood for some time, statistical concerns about the adequacy of the Level 1 intercept and slope estimates as well as the estimation of the variance components hindered the full development of these models (Burstein et al., 1989). Throughout the 1980s, however, several, separate, statistical advances improved greatly the estimation strategy for intercepts- and slopes-as-outcomes models (Burstein et al., 1989). Bryk and Raudenbush (1992) enumerated the specific statistical advances and their relationship to hierarchical linear models. These advances have resulted in the development of several different software packages designed specifically for the analysis of multilevel or hierarchical data (e.g., HLM; Bryk et al., 1994; MLn; Rasbash & Woodhouse, 1995; VARCL; Longford, 1990).

### Estimating the Effects

In estimating the Level 1 and Level 2 models discussed above, a distinction is made between fixed effects, random coefficients, and variance components. Fixed effects are parameter estimates that do not vary across groups, for example the $\gamma$s from Equations 2 and 3. Alternatively, random coefficients are parameter estimates that are allowed to vary across groups such as the Level 1 regression coefficients (e.g., $\beta_{0j}$ and $\beta_{1j}$). In addition to these Level 1 and Level 2 regression coefficients, the HLM software also includes estimates of the variance components, which include: (a) the variance in the Level 1 residual (i.e., $e_{ij}$ referred to as $\sigma^2$), (b) the variance in the Level 2 residuals (i.e., $U_{0j}$ and $U_{ij}$), and (c) the covariance of the Level 2 residuals (i.e., $\text{cov}(U_{0j} \text{ and } U_{ij})$). The variance–covariance matrix of the Level 2 residuals is referred to in the hierarchical linear modeling literature as the $\tau$ matrix. The element $\tau_{00}$ represents the variance in $U_{0j}$, element $\tau_{11}$ represents the variance in $U_{ij}$, and element $\tau_{12}$ represents the covariance between $U_{0j}$ and $U_{ij}$. Obviously, the number of elements in the $\tau$ matrix will depend on the number of Level 2 equations estimated, the number of Level 1 predictors that vary and are modeled.

### The Fixed Effects

The predictor weights ($\gamma$) in Equations 2 and 3 represent fixed effects in HLMs. Although these Level 2 regression weights could be estimated using an OLS regression approach, this is not appropriate given that the precision of the Level 1 parameters will vary across the groups (as the groups are not required to be of equal size). Given this variation in precision, an OLS approach is not appropriate because of the violation of the homoscedasticity assumption. HLMs use a generalized least-squares (GLS) estimate for the Level 2 parameters, this generates a weighted Level 2 regression. The groups with more precise Level 1 estimates receive more weight in the Level 2 regression equation.
The Variance–Covariance Components

The variance–covariance components in HLMs represent the variance of the Level 1 residuals (i.e., the variance in the $e_{ij}$) and the variance–covariance of the Level 2 residuals (i.e., the variance–covariance of $U_{0j}$ and $U_{1j}$). These variance components are estimated using maximum likelihood estimation and the EM algorithm (Raudenbush & Bryk, 2002).

Level 1 Random Coefficients

Frequently, especially in the context of education research, a researcher is interested in obtaining the best estimate of a particular, Level 1 random coefficient (Raudenbush, 1988). In the design context, this might be when a researcher is interested in estimating the effectiveness of a particular learning environment where effectiveness is conveyed by the design and significance of the Level 1 slope coefficient. For example, a learning environment may be defined as being most effective when it reduces the effect of socio-economic status within groups on mathematics performance—a classic condition for the study of equity in mathematics performance. One of the simplest ways to estimate the Level 1 coefficient for a particular group or school is to compute an OLS regression equation for that particular unit (e.g., Equation 1). Assuming large sample sizes in each group, this analysis should provide relatively precise estimates. In practice, however, group size can be small (and certainly unequal); fortunately, HLMs deal with this possibility explicitly. When groups are smaller, these estimates will not be stable (Burstein, 1980). Inspection of Level 2 equations (i.e., Equations 2 and 3) reveals that there are two estimates of the Level 1 intercepts and slopes. The first estimate comes from an OLS regression equation estimated for a particular unit (i.e., Equation 1), whereas the second estimate comes from the Level 2 regression model (i.e., the predicted values of $\beta_0$ and $\beta_1$ from Equations 2 and 3). In other words, for any particular unit, two predicted, intercept and slope values can be estimated: the first from the Level 1 regression equation and the second from the Level 2 regression model. Therefore, the question becomes which of these estimates provides a more accurate assessment of the population intercept and slope parameters for that particular unit. Alternatively, we can ask how the two sets of estimates might be weighted optimally.

Instead of forcing a choice between these two estimates, HLMs (and the HLM software program; Bryk & Raudenbush, 1992) compute an optimally-weighted combination of the two estimates using an empirical Bayes estimation strategy (Raudenbush & Bryk, 2002). In other words, the HLM software program computes an empirical Bayes estimate of the Level 1 intercepts and slopes for each unit, which optimally weights the OLS, Level 1 estimates (Equation 1) and the Level 2 predicted values for these same estimates (Equations 2 and 3). These empirical Bayes estimates are contained in the residual file generated by the HLM software. Raudenbush (1988) provides proofs demonstrating that this composite estimate produces a smaller mean square error term than either the Level 1 estimate or the Level 2 predicted value. Thus, when one is interested in obtaining the best estimate of the Level 1 coefficient for a particular unit, the empirical Bayes estimate will meet this criterion. Of course, this assumes that both the Level 1 and Level 2 models are specified correctly.

The empirical Bayes estimates are a weighted composite of the two estimates discussed above where the weight is based on the reliability of the OLS estimate. The HLM software program provides an estimate of the “reliability” of the OLS Level 1 regression coefficients. First, the software partitions the variance in the OLS regression
parameters for each group into its estimated true parameter variance and error variance (e.g., variance in $B_{0j} = \text{true variance in } B_{0j} + \text{error variance in } B_{0j}$). This parallels the classical test theory model. Then, after obtaining these estimates, one can use the software to compute a “reliability coefficient” for each group’s OLS parameters as the ratio of the true parameter variance to the total parameter variance (i.e., reliability = true variance/total variance). The HLM software reports the reliability of each Level 1 random coefficient averaged across the groups. This reported reliability can be interpreted as the amount of systematic variance in the parameter across the groups (i.e., the variance that is available to be modeled by between-group variables).

**Possible Statistical Tests**

The HLM software contains several statistical tests for hypothesis testing. Specifically, there are $t$-tests for all of the fixed effects (i.e., the second-level regression parameters), which test whether these parameter estimates depart significantly from zero. Chi-square tests are provided for the Level 2 residual variance (e.g., variance in the $U$s; for instance, $[\tau_{00} \text{ and } \tau_{11}]$, indicating whether the residual variance departs significantly from zero). Other more complicated tests are available but we do not discuss them here. For the majority of HLMs, these basic tests should suffice.

In the above introduction we reviewed the background, logic, rationale, and estimation approach of HLMs. In the next section we explore how these models can be applied to answer questions relevant to design researchers in general and design researchers in mathematics education in particular. To illustrate the hierarchical linear modeling approach further, a hypothetical set of research questions is presented first. This deliberation is followed by a discussion of the sequence of the models that would be used to investigate these questions.

**Investigating the Effects of Group Sociomathematical Norms on Individual Performance in Mathematics**

Hierarchical linear models are valuable in the design process at two end points: at the end of each design iteration (see Cobb et al., 2003; Design-Based Research Collective, 2003) to examine a theory (and its possible effect), and then at the end of the design process. In each case, the HLM analysis provides the qualitative design researcher with a quantitative mechanism for checking if the designed product is having the anticipated learning effects.

Suppose that a mathematics education design researcher is interested in predicting mathematics performance at the individual level. Suppose also that the researcher has identified beliefs about mathematics (an individual-level variable) and sociomathematical norms (a group-level variable) as potential predictors of this mathematics performance. Then, the summative analyses proposed an outlined here are likely candidates that a design researcher could employ at the end of the design process. This makes sense when the goal of the researcher is to verify or check the size and direction of these multilevel relationships. Further, comparable analyses could be conducted at the end of the design iterate.

Historically, there has been a dichotomy of views about mathematical learning that distinguish the individual cognitive perspective of constructivism (von Glasersfeld, 1995) from the sociocultural perspective based on symbolic interactionism (Blumer, 1969). The emergent perspective represents the possibility of coordinating the two views, with the underlying assumption that mathematical learning can be characterized
“as both a process of active individual construction and as a process of mathematical enculturation” (Cobb, 1994: 35). Cobb and his colleagues (e.g., Bauersfeld, 1995; Cobb & Bauersfeld, 1995; Yackel & Cobb, 1996) have made significant contributions through their articulation of an interpretive framework that coordinates both the psychological and sociocultural (cognitive) perspectives on students’ learning (see Table 24.1). Table 24.2 provides an adaptation from Cobb and Yackel (1996) to frame this multilevel example.

The goals of framing the theory and analysis from a multilevel perspective allow us to ascertain the strength of these relationships and to investigate whether the shared perspective affects students’ performance in mathematics. This latter goal, the impact of group norms on student performance, is assumed in the Cobb model but never investigated because his theoretical focus is on the emergent production of group-level sociomathematical norms.

Table 24.1 specifies three, rather straightforward, hypotheses about the relationship between mathematics performance, beliefs about mathematics, and group sociomathematical norms. In order for these hypotheses to be supported, several conditions must be met. These conditions are listed in the bottom half of Table 24.1.

### Table 24.1 Working Hypotheses and their Assumed Conditions: Mathematical Performance, Beliefs About Mathematics, and Sociomathematical Norms (Within Groups)

<table>
<thead>
<tr>
<th>Working hypotheses</th>
<th>Assumed conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belief about mathematics is related positively to performance in mathematics.</td>
<td>Systematic within- and between-group variance in mathematics performance.</td>
</tr>
<tr>
<td>Sociomathematical norms are related positively to performance in mathematics after controlling for beliefs about mathematics (i.e., on average, individuals learning in environments with positive sociomathematical norms are more likely to have positive beliefs about mathematics; in other words, such individuals will show a group-level main effect for sociomathematical norms after controlling for mathematical beliefs).</td>
<td>Significant variance in the Level 1 intercept.</td>
</tr>
<tr>
<td>Sociomathematical norms moderate the beliefs about mathematics–mathematics performance relationship (i.e., the relationship between beliefs about mathematics and mathematics performance is stronger in situations where group members are in closer sociomathematical norms to one another).</td>
<td>Significant variance in the Level 1 slope.</td>
</tr>
<tr>
<td>Variance in the intercept is predicted significantly by sociomathematical norms of group members.</td>
<td>Variance in the intercept is predicted significantly by sociomathematical norms of group members.</td>
</tr>
<tr>
<td>Variance in the slope is predicted significantly by the sociomathematical norms of group members.</td>
<td></td>
</tr>
</tbody>
</table>

### Table 24.2 Framework for Interpreting Individual and Social Activity in Learning

<table>
<thead>
<tr>
<th>Psychological perspective</th>
<th>Social perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual student belief about his or her own role, the roles of others, and the general nature of mathematical activity.</td>
<td>Social norms in the classroom.</td>
</tr>
<tr>
<td>A student’s specific mathematical beliefs and values.</td>
<td>Sociomathematical norms in the classroom.</td>
</tr>
<tr>
<td>A student’s conceptions and activities in mathematics.</td>
<td>Mathematical practices in the classroom.</td>
</tr>
</tbody>
</table>

Source: Adapted from Cobb and Yackel (1996).
Hypotheses 1 and 2 suggest that mathematics performance will be related significantly to both an individual-level variable (i.e., beliefs about mathematics), and a group-level variable (i.e., sociomathematical norms). Thus, one should expect meaningful within- and between-group variance in mathematics performance (Condition 1). Hypothesis 2 proposes that, after controlling for beliefs about mathematics, mathematics performance will be associated significantly with sociomathematical norms. In this example of an HLM, the variance in the Level 1 intercept term represents the between-group variance in mathematics performance after controlling for beliefs about mathematics. Thus, for Hypothesis 2 to be supported, there needs to be significant variance in the intercept term (Condition 2), and this variance needs to be related significantly to the sociomathematical norms of group members (Condition 4 and Hypothesis 2). Hypothesis 3 proposes that the relationship between beliefs about mathematics and mathematics performance will vary as a function of the sociomathematical norms of group members. Therefore, for this hypothesis to be supported, there would need to be significant variance in the Level 1 slope coefficient across the groups (i.e., the relationship between beliefs about mathematics and mathematics performance; Condition 3), and this variance would have to be related significantly to the sociomathematical norms of group members (Condition 5 and Hypothesis 3). In the following section we outline a typical sequence of models that would allow the design researcher the opportunity to assess (and statistically test) the viability of each of these necessary conditions as well as the three hypotheses listed in Table 24.1.

One-way Analysis of Variance

The first condition specifies systematic within- and between-group variance in mathematics performance. The investigation of within- and between-group variance suggests that the variance in mathematics performance needs to be partitioned into each component separately. To accomplish the variance partitioning in HLMs, the following set of equations is estimated:

Level 1: Mathematics performance_{ij} = B_{0j} + e_{ij} (4)
Level 2: B_{0j} = \gamma_{00} + U_{0j} (5)

where:

\(B_{0j}\) = mean mathematics performance for group j
\(\gamma_{00}\) = grand mean mathematics performance (across all groups)
Variance (\(e_{ij}\)) = \([\sigma^2]\) = within-group variance in mathematics performance
Variance (\(U_{0j}\)) = \([\tau_{00j}]\) = between-group variance in mathematics performance

In Equations 4 and 5, the Level 1 equation (4) includes no predictors. Therefore, the regression equation includes only an intercept estimate. In order to compute intercept terms in regression, the analysis includes a unit vector as a predictor in the equation. The parameter associated with this unit vector represents the intercept term in the final regression equation. When a researcher specifies no predictors in a Level 1 or Level 2 equation, the variance in the outcome measure is regressed implicitly onto a unit vector producing a regression-based intercept estimate. In the Level 1 equation above (4), mathematics performance is regressed onto a constant unit vector, which is implied when one chooses no predictors. Because there are no additional predictors in the
model, the $B_{0j}$ parameter will be equal to that group’s mean level of mathematics performance (i.e., if a variable is regressed only onto a constant unit vector, the resulting parameter is equal to the mean).

The Level 2 model (Equation 5) regresses each group’s mean mathematics performance onto a constant; that is, $B_{0j}$ is regressed onto a unit vector, resulting in a $\gamma_{00}$ parameter equal to the grand mean of mathematics performance (i.e., the mean of the group means, $B_{0j}$—dyad by dyad, group by group, or classroom by classroom). Given that each of the respective dependent variables is regressed onto a constant, it follows that any within-group variance in mathematics performance is forced into the Level 1 residual (i.e., $e_{ij}$) and any between-group variance in mathematics performance is forced into the Level 2 residual (i.e., $U_{0j}$).

Although hierarchical linear modeling does not provide a significance test for the within-group variance component, it does provide a significance test for the between-group variance (i.e., $\tau_{00}$). In addition, the ratio of the between-group variance to the total variance is presented as an intraclass correlation (ICC). In the model above (Equations 4 and 5), the total variance in mathematics performance has been decomposed into its within- and between-group components [i.e., Variance (mathematics performance) = Variance ($U_{0j} + e_{ij}$) = ($\tau_{00} + \sigma^2$)]. Therefore, an intraclass correlation can be computed by investigating the following ratio: ICC = $[(\tau_{00})/(\tau_{00} + \sigma^2)]$. This intraclass correlation represents a ratio of the between-group variance in mathematics performance to the total variance in mathematics performance (i.e., the percentage of variance in mathematics performance that resides between groups).

In summary, the one-way analysis of variance provides the following information about the mathematics performance measure: (a) the amount of variance residing within groups, (b) the amount of variance residing between groups, and (c) the intraclass correlation specifying the percentage of the total variance residing between groups.

The Random Coefficient Model

After assessing the degree of within- and between-group variance in mathematics performance, one can investigate now whether there is significant variance in the intercepts and slopes across groups (Conditions 2 and 3). In other words, for Hypothesis 2 to be supported there needs to be significant variance across groups in the intercepts. For Hypothesis 3 to be supported, there needs to be significant variance across groups in the slopes. In addition to providing evidence in support of Conditions 2 and 3, this model also will test Hypothesis 1 directly. The random coefficient regression model takes on the following form:

\[
\text{Level 1: Math performance}_{ij} = B_{0j} + B_{1j} \text{ (Beliefs about math}_{ij} \right) + e_{ij} \tag{6} \]

\[
\text{Level 2a: } B_{0j} = \gamma_{00} + U_{0j} \tag{7} \]

\[
\text{Level 2b: } B_{1j} = \gamma_{10} + U_{1j} \tag{8} \]

where:

$\gamma_{00}$ = mean of the intercepts across groups

$\gamma_{10}$ = mean of the slopes across groups (a check of Hypothesis 1)

Variance ($e_{ij}$) = Level 1 residual variance

Variance ($U_{0j}$) = [$\tau_{00}$] = variance in intercepts

Variance ($U_{1j}$) = [$\tau_{11}$] = variance in slopes
Because there are no Level 2 predictors of either $B_{0j}$ or $B_{1j}$, the Level 2 regression equations (7 and 8) are equal to an intercept term and a residual. In this form, the $\gamma_{00}$ and the $\gamma_{10}$ parameters represent the Level 1 coefficients averaged across groups (i.e., they represent the pooled $B_{0j}$ and $B_{1j}$ parameters). Similarly, given that $B_{0j}$ and $B_{1j}$ are regressed onto constants, the variance of the Level 2 residual terms (i.e., $U_{0j}$ and $U_{ij}$) represents the between-group variance in the Level 1 parameters.

Hierarchical linear modeling provides a $t$-test related to the $\gamma_{00}$ and $\gamma_{10}$ parameters, where a significant $t$-value indicates that the parameter departs significantly from zero. In the case of the $\gamma_{10}$ parameter, this $t$-test provides a direct test of Hypothesis 1. In other words, this tests whether beliefs about mathematics are related significantly to mathematics performance. This test assesses whether the pooled Level 1 slope between beliefs about mathematics and mathematics performance differs significantly from zero. Thus, this test investigates whether, on average, the relationship between beliefs about mathematics and mathematics performance is significant.

Hierarchical linear modeling also provides a chi-square test for the two residual variances (i.e., $\tau_{00}$ and $\tau_{11}$). These chi-square tests indicate whether the variance components differ significantly from zero and afford a direct test of Conditions 2 and 3. In other words, these tests determine whether the variance in the intercepts and slopes across groups is significantly different from zero. Thus, the random regression model furnishes two primary pieces of information: (a) it tests the significance of the pooled Level 1 slopes, which are used to test Level 1 hypotheses, and (b) it evaluates whether there is significant variance surrounding the pooled Level 2 intercepts and slopes. In other words, the random regression model provides a significance test for the mean of the Level 1 regression coefficients, as well as a significance test for the amount of variance in each of the Level 1 regression coefficients.

In addition to estimating the fixed ($\gamma$) and random ($\tau$) effects, hierarchical linear modeling also estimates the Level 1 residual variance (i.e., the variance in $e_{ij}$ or $\sigma^2$). Recall that in the one-way analysis of variance model (Equation 4), $\sigma^2$ was equal to the within-group variance in mathematics performance. Because the random regression model adds a Level 1 predictor (Equation 5), $\sigma^2$ is now equal to the Level 1 residual variance. Therefore, comparing these two values of $\sigma^2$ provides an estimate of the Level 1 variance in mathematics performance accounted for by beliefs about mathematics. More specifically, the $R^2$ (i.e., the variance accounted for) in mathematics performance can be obtained by computing the following ratio:

\[
R^2 \text{ for Level 1 model} = \frac{(\sigma^2 \text{ one-way ANOVA}) - (\sigma^2 \text{ random regression})}{(\sigma^2 \text{ one-way ANOVA})}.
\]

This ratio represents the percentage of the Level 1 variance in mathematics performance accounted for by beliefs about mathematics.

The Intercepts-As-Outcomes Model

Assuming that Condition 2 was satisfied in the random regression model (i.e., there was significant variance in the intercept term), the intercepts-as-outcomes model assesses whether this variance is related significantly to the sociomathematical norms of group members. This model tests Condition 4, which is also a test of Hypothesis 2.

The HLM would take the following form:

\[
\text{Level 1: Math performance}_{ij} = B_{0j} + B_{1j} \text{(Beliefs about math}_{ij}) + e_{ij} \tag{9}
\]
Level 2a: $B_{0j} = \gamma_{00} + \gamma_{01} (\text{sociomathematical norms}_j) + U_{0j}$ \hspace{1cm} (10)
Level 2b: $B_{1j} = \gamma_{10} + U_{1j}$ \hspace{1cm} (11)

where:
- $\gamma_{00}$ = the intercept for Level 2
- $\gamma_{01}$ = the slope for Level 2 (a test of Hypothesis 2)
- $\gamma_{10}$ = mean (pooled) slopes
- Variance ($e_{ij}$) = $[\sigma^2]$ = residual variance for Level 1
- Variance ($U_{0j}$) = $[\tau_{00}]$ = residual intercept variance Level 2a
- Variance ($U_{1j}$) = $[\tau_{11}]$ = variance in slopes Level 2b

This model is similar to the random regression model discussed above, with the addition of the variable “sociomathematical norms” as a Level 2 predictor of $B_{0j}$. Therefore, the $t$-test associated with the $\gamma_{01}$ parameter provides a direct test of Hypothesis 2; that is, the relationship between sociomathematical norms and mathematics performance after controlling for individual beliefs about mathematics. Given that the Level 2 equation for $B_{0j}$ now includes a predictor (i.e., sociomathematical norms), the variance in the $U_{0j}$ parameter (i.e., $\tau_{00}$) represents the residual variance in $B_{0j}$ across groups. If the chi-square test for this parameter is significant, it indicates that there remains systematic Level 2 variance that could be modeled by other theoretically valid Level 2 predictors. If the chi-square test of this residual variance is not significant, the researcher may use an option in hierarchical linear modeling to fix this variance component to zero (i.e., implying that all of the systematic, between-group variance in $B_{0j}$ has been accounted for by sociomathematical norms). All other parameters take on the same meaning as they did under the estimation of the random regression model (i.e., the chi-square for $\tau_{11}$ provides an assessment of Condition 3).

To obtain information about the percentage of variance accounted for by inclusion of the predictor variable “sociomathematical norms” in the Level 2 model, the same type of procedure described above is invoked. In the random regression model, $\tau_{00}$ was equal to the between-group variance in the intercept term (i.e., $B_{0j}$). In this intercepts-as-outcomes model, a Level 2 predictor (sociomathematical norms) has been added to the equation, rendering $\tau_{00}$ equal to the residual or between-group variance in the intercept term. Thus, by comparing these two variance estimates, one can obtain the $R^2$ for sociomathematical norms. The $R^2$ is computed as follows:

$$R^2 \text{ for Level 2 intercept model} = \frac{(\tau_{00}\text{-random regression}) - (\tau_{00}\text{-intercepts-as-outcomes})}{(\tau_{00}\text{-random regression})}.$$  

Once again, this ratio compares the amount of variance across the intercept terms accounted for by sociomathematical norms.

**Slopes-As-Outcomes Model**

Assuming that Condition 3 was supported in the intercepts-as-outcomes model, one can now investigate whether the variance in the slope across groups is related significantly to the sociomathematical norms of group members. Therefore, the slopes-as-outcomes model provides a direct test of Condition 5, which is also a test of Hypothesis 3.

The HLM would take the following form:
Level 1: \[ \text{Math performance}_{ij} = B_{0j} + B_{1j} \text{ (Beliefs about math)}_{ij} + e_{ij} \] (12)

Level 2a: \[ B_{0j} = \gamma_{00} + \gamma_{01} \text{ (Sociomathematical norms)}_{j} + U_{0j} \] (13)

Level 2b: \[ B_{1j} = \gamma_{10} + \gamma_{11} \text{ (Sociomathematical norms)}_{j} + U_{1j} \] (14)

where:

\[ \gamma_{00} = \text{the intercept at Level 2a} \]
\[ \gamma_{01} = \text{the slope at Level 2 (a test of Hypothesis 2)} \]
\[ \gamma_{10} = \text{the intercept at Level 2b} \]
\[ \gamma_{11} = \text{the slope at Level 2b (a test of Hypothesis 3)} \]
\[ \text{Variance (} e_{ij} \text{)} = [\sigma^2] = \text{Level 1 residual variance} \]
\[ \text{Variance (} U_{0j} \text{)} = [\tau_{00}] = \text{residual intercept variance} \]
\[ \text{Variance (} U_{1j} \text{)} = [\tau_{11}] = \text{residual slope variance} \]

The differences between this model and the intercepts-as-outcomes model above are that the variable sociomathematical norms is included as a predictor of the \( B_{1j} \) parameter, and, as a result, the \( U_{1j} \) variance is now the residual variance in the \( B_{1j} \) parameter across groups, as opposed to, or instead of, the total variance across groups. Once again, if the chi-square test associated with this parameter variance is significant, it indicates that there remains systematic variance in the \( B_{1j} \) parameter that could be modeled by additional Level 2 predictors. In addition, the \( t \)-test associated with the \( \gamma_{11} \) parameter provides a direct test of Hypothesis 3. This hypothesis represents a cross-level interaction because a group-level variable is hypothesized to moderate the relationship between two, individual-level variables (the intercept itself and the estimated effect of the beliefs variable). As Saari notes (this volume), making inferences across levels requires appropriate analytic tools and the HLM is one such tool.

We now compute the \( R^2 \) for sociomathematical norms as a Level 2 moderator of the relationship between individual-level beliefs about mathematics and mathematics performance using the value of \( \tau_{ii} \) from the intercepts-as-outcomes model (i.e., the total, between-group variance in \( B_{ij} \) and the value of \( \tau_{11} \) from the slopes-as-outcomes model). We obtain an estimate of the \( R^2 \) as follows:

\[ R^2 \text{ for Level 2 slope model} = (\tau_{11} \text{-intercepts-as-outcomes} - \tau_{11} \text{-slopes-as-outcomes}) / (\tau_{11} \text{-intercepts-as-outcomes}). \]

This ratio compares the percentage of variance accounted for by sociomathematical norms to the total variance in the belief performance behavior slope across groups.

The preceding sequence of models provides a general introduction to HLMs and the HLM software. The extension of these models to include more Level 1 and Level 2 predictors is relatively straightforward. The purpose of this overview is to provide a general introduction to the ways in which researchers might ask and answer multilevel questions in the hierarchical modeling framework. Additional details about more complex estimation strategies and the statistical intricacies of HLMs can be found in Bryk and Raudenbush (1992), Goldstein (1995), and Longford (1993).
Additional Comments

Before concluding, several additional issues are worth mentioning: the application of HLMs to longitudinal data, centering issues, and expanding the models to include more levels.

A Longitudinal Formulation

Although it might not be apparent, virtually all longitudinal investigations conducted by design researchers are hierarchical in nature (Raudenbush & Bryk, 2002). The nested nature of these data would include multiple observations within a unit and a sample of multiple units. Thus, one would have a within-unit, Level 1 model and a between-unit Level 2 model. From a theoretical perspective, one is investigating inter-unit differences in intra-unit change (Nesselroade, 1991). The resulting data structure is one where a time series of data is nested within a larger number of students, thus allowing for an investigation of inter-student differences in change or growth. Sloane, Holding, and Kelly (this volume) highlight the value associated with this possibility in the context of design research and interrupted time-series analysis.

Centering the Intercepts: The Need for Theory

Because HLMs use the Level 1 regression parameters (i.e., intercepts and slopes) as outcome variables to be predicted by the Level 2 equation(s), it is imperative that researchers understand fully the specific interpretation of these parameters. As noted in basic regression texts (e.g., Cohen et al., 2003), the slope parameter represents the expected increase in the outcome variable for a unit increase in the predictor variable, whereas the intercept parameter represents the expected value of the outcome measure when all the predictors are zero. In the ongoing example used in this chapter, the slope merely represents the predicted increase in mathematics performance given a unit increase in belief. The intercept term represents the predicted level of mathematics performance for a person with zero belief about mathematics. However, an obvious question about the meaning of the intercept comes to mind: “How can someone have zero beliefs?”

Like the belief example above, a value of zero is not particularly meaningful for many of the constructs studied in design research (e.g., the following standard equity variables: ethnicity, gender, and socio-economic status). For the Level 2 model we can, and should, ask: what does it mean for a learning environment to have zero norms? To make intercepts more interpretable level by level, a number of researchers have discussed different ways in which to rescale the Level 1 predictors. “Centering” describes the rescaling of the Level 1 predictors, for which three primary options are now available:

1. A raw metric approach, where no centering takes place and the Level 1 predictors retain their original metric—as in the example above.
2. A grand mean centering, where the grand mean is subtracted from each individual’s score on the predictor (e.g., $\text{belief}_i - \text{belief}_{\text{grand mean}}$).
3. A group mean centering, where the group mean is subtracted from each individual’s score on the predictor (e.g., $\text{belief}_i - \text{belief}_{\text{group mean}}$).

With grand mean centering, the intercept represents the expected level of the outcome for a person with an “average” level on the predictor. In the current case, it would be the
expected mathematics performance for a person with average belief. With group mean centering, the intercept represents the expected mathematics performance for a student with his or her group’s average beliefs about mathematics. In both cases, the intercept is theoretically more substantive, and more interpretable than the raw metric alternative. However, centering issues do not begin and end with intercept interpretation.

Recently, several researchers have discussed how the various centering options can change the estimation and meaning of the HLM as a whole (Longford, 1989; Raudenbush & Bryk, 2002). The choice of centering options goes well beyond the interpretation of the intercept term. A researcher must consider the overarching theoretical paradigm primarily and, from that, discern what centering option represents the paradigm best.

Dealing With More Complexity: Adding More Levels

So far the discussion has focused on two-level models, but it is quite obvious that learning settings can represent more than two hierarchical levels. The extension of the two-level model to higher-level models is relatively straightforward but adds theoretical difficulty. For example, in the current example, if individuals were sampled across different learning groups within classrooms, a three-level model could be estimated, where the Level 1 model would describe individuals within groups, the Level 2 model would be groups within classrooms, and the Level 3 model would be a between-classrooms model. The HLM software is available for up to three levels, and a revised version of the Mln (Rasbash & Woodhouse, 1995) software program allows the researcher to work with as many as 15 levels of hierarchy (or nesting). One should keep in mind, however, that very large numbers of levels put more demands on the researcher’s capacity to theorize sensibly, not to mention the overt need for additional data resources.

Conclusions

As the call for developing multilevel theories in education and design research continues (Sloane, 2005), it is important to acknowledge and use methodological advances from other disciplines to begin testing hypothesized relationships across levels. Although HLMs have been discussed for several years in the methodological literature in education and in other disciplines, they have yet to gain much attention in the design sciences. HLMs represent an avenue by which more complex theories of learning and the environments that support such learning can be investigated, tested, and understood (Barab & Kirshner, 2001). Moreover, they are built to address the aggregation issues raised by Saari (this volume). HLMs are far from perfect but they represent a great technical leap forward and provide a mechanism for testing the relationships between variables that cross multiple nested levels. The continuing thrust for the integration of macro and micro concepts into design theories, coupled with these technical advances, should lead to a better understanding of how to design learning environments while honoring their complexity.

Acknowledgments

The author would like to thank Anthony E. Kelly (George Mason University) and Daniel Battey (Mary Lou Fulton College of Education, Arizona State University) for their helpful comments on an earlier draft. The author acknowledges funding from the
Notes

1 In Chapter 23, Sloane, Helding, and Kelly describe how the framework of HLM can be employed to model longitudinal data.

2 This idea is elaborated by Sloane, Helding, and Kelly (this volume). The elaboration is important for design researchers given that iterations over the same group of students will require some form of time-series analytic tools for estimates to adjust for the autocorrelated error structure of the data. From a design perspective, the fact that some, if not all, students use various versions of the design tool to improve their learning has to be considered before any statement can be made about the efficacy of the designed product. This can be ameliorated by using different groups of students across the various design cycles.

References


