An introduction to hierarchical linear modeling

Heather Woltman, Andrea Feldstain, J. Christine MacKay, Meredith Rocchi
University of Ottawa

This tutorial aims to introduce Hierarchical Linear Modeling (HLM). A simple explanation of HLM is provided that describes when to use this statistical technique and identifies key factors to consider before conducting this analysis. The first section of the tutorial defines HLM, clarifies its purpose, and states its advantages. The second section explains the mathematical theory, equations, and conditions underlying HLM. HLM hypothesis testing is performed in the third section. Finally, the fourth section provides a practical example of running HLM, with which readers can follow along. Throughout this tutorial, emphasis is placed on providing a straightforward overview of the basic principles of HLM.

[*]Hierarchical levels of grouped data are a commonly occurring phenomenon (Osborne, 2000). For example, in the education sector, data are often organized at student, classroom, school, and school district levels. Perhaps less intuitively, in meta-analytic research, participant, procedure, and results data are nested within each experiment in the analysis. In repeated measures research, data collected at different times and under different conditions are nested within each study participant (Raudenbush & Bryk, 2002; Osborne, 2000). Analysis of hierarchical data is best performed using statistical techniques that account for the hierarchy, such as Hierarchical Linear Modeling.

Hierarchical Linear Modeling (HLM) is a complex form of ordinary least squares (OLS) regression that is used to analyze variance in the outcome variables when the predictor variables are at varying hierarchical levels; for example, students in a classroom share variance according to their common teacher and common classroom. Prior to the development of HLM, hierarchical data was commonly assessed using fixed parameter simple linear regression techniques; however, these techniques were insufficient for such analyses due to their neglect of the shared variance. An algorithm to facilitate covariance component estimation for unbalanced data was introduced in the early 1980s. This development allowed for widespread application of HLM to multilevel data analysis (for development of the algorithm see Dempster, Laird, & Rubin, 1977; for its application to HLM see Dempster, Rubin, & Tsutakawa, 1981). Following this advancement in statistical theory, HLM's popularity flourished (Raudenbush & Bryk, 2002; Lindley & Smith, 1972; Smith, 1973).

HLM accounts for the shared variance in hierarchically structured data: The technique accurately estimates lower-level slopes (e.g., student level) and their implementation in estimating higher-level outcomes (e.g., classroom level;

Table 1. Factors at each hierarchical level that affect students' Grade Point Average (GPA)

Hierarchical Level	Example of Hierarchical Level	Example Variables
Level-3	School Level	School's geographic location
		Annual budget
Level-2	Classroom Level	Class size
		Homework assignment load
		Teaching experience
		Teaching style
Level-1	Student Level	Gender
		Intelligence Quotient (IQ)
		Socioeconomic status
		Self-esteem rating
		Behavioural conduct rating
		Breakfast consumption
		GPA[a]

[a] The outcome variable is always a level-1 variable.

Hofmann, 1997). HLM is prevalent across many domains, and is frequently used in the education, health, social work, and business sectors. Because development of this statistical method occurred simultaneously across many fields, it has come to be known by several names, including multilevel-, mixed level-, mixed linear-, mixed effects-, random effects-, random coefficient (regression)-, and (complex) covariance components-modeling (Raudenbush & Bryk, 2002). These labels all describe the same advanced regression technique that is HLM. HLM simultaneously investigates relationships within and between hierarchical levels of grouped data, thereby making it more efficient at accounting for variance among variables at different levels than other existing analyses.

Example

Throughout this tutorial we will make use of an example to illustrate our explanation of HLM. Imagine a researcher asks the following question: *What school-, classroom-, and student-related factors influence students' Grade Point Average?* This research question involves a hierarchy with three levels. At the highest level of the hierarchy (level-3) are school-related variables, such as a school's geographic location and annual budget. Situated at the middle level of the hierarchy (level-2) are classroom variables, such as a teacher's homework assignment load, years of teaching experience, and teaching style. Level-2 variables are nested within level-3 groups and are impacted by level-3 variables.

For example, schools (level-3) that are in remote geographic locations (level-3 variable) will have smaller class sizes (level-2) than classes in metropolitan areas, thereby affecting the quality of personal attention paid to each student and noise levels in the classroom (level-2 variables).

Variables at the lowest level of the hierarchy (level-1) are nested within level-2 groups and share in common the impact of level-2 variables. In our example, student-level variables such as gender, intelligence quotient (IQ), socioeconomic status, self-esteem rating, behavioural conduct rating, and breakfast consumption are situated at level-1. To summarize, in our example students (level-1) are situated within classrooms (level-2) that are located within schools (level-3; see Table 1). The outcome variable, grade point average (GPA), is also measured at level-1; in HLM, the outcome variable of interest is always situated at the lowest level of the hierarchy (Castro, 2002).

For simplicity, our example supposes that the researcher wants to narrow the research question to two predictor variables: *Do student breakfast consumption and teaching style influence student GPA?* Although GPA is a single and continuous outcome variable, HLM can accommodate multiple continuous or discrete outcome variables in the same analysis (Raudenbush & Bryk, 2002).

Methods for Dealing with Nested Data

An effective way of explaining HLM is to compare and contrast it to the methods used to analyze nested data prior to HLM's development. These methods, disaggregation and aggregation, were referred to in our introduction as simple linear regression techniques that did not properly account for the shared variance that is inherent when dealing with hierarchical information. While historically the use of disaggregation and aggregation made analysis of hierarchical data possible, these approaches resulted in the incorrect partitioning of variance to variables, dependencies in the data, and an increased risk of making a Type I error (Beaubien, Hamman, Holt, & Boehm-Davis, 2001; Gill, 2003; Osborne, 2000).

Disaggregation

Disaggregation of data deals with hierarchical data issues by ignoring the presence of group differences. It considers all relationships between variables to be context free and situated at level-1 of the hierarchy (i.e., at the individual level). Disaggregation thereby ignores the presence of possible between-group variation (Beaubien et al., 2001; Gill, 2003; Osborne, 2000). In the example we provided earlier of a researcher investigating whether level-

Table 2. Sample dataset using the disaggregation method, with level-2 and level-3 variables excluded from the data (dataset is adapted from an example by Snijders & Bosker, 1999)

Student ID (Level-1)	Classroom ID (Level-2)	School ID (Level-3)	GPA Score (Level-1)	Breakfast Consumption Score (Level-1)
1	1	1	5	1
2	1	1	7	3
3	2	1	4	2
4	2	1	6	4
5	3	1	3	3
6	3	1	5	5
7	4	1	2	4
8	4	1	4	6
9	5	1	1	5
10	5	1	3	7

1 variable breakfast consumption affects student GPA, disaggregation would entail studying level-2 and level-3 variables at level-1. All students in the same class would be assigned the same mean classroom-related scores (e.g., homework assignment load, teaching experience, and teaching style ratings), and all students in the same school would be assigned the same mean school-related scores (e.g., school geographic location and annual budget ratings; see Table 2).

By bringing upper level variables down to level-1, shared variance is no longer accounted for and the assumption of independence of errors is violated. If teaching style influences student breakfast consumption, for example, the effects of the level-1 (student) and level-2 (classroom) variables on the outcome of interest (GPA) cannot be disentangled. In other words, the impact of being taught in the same classroom on students is no longer accounted for when partitioning variance using the disaggregation approach. Dependencies in the data remain uncorrected, the assumption of independence of observations required for simple regression is violated, statistical tests are based only on the level-1 sample size, and the risk of partitioning variance incorrectly and making inaccurate statistical estimates increases (Beaubien et al., 2001; Gill, 2003; Osborne, 2000). As a general rule, HLM is recommended over disaggregation for dealing with nested data because it addresses each of these statistical limitations.

In Figure 1, depicting the relationship between breakfast consumption and student GPA using disaggregation, the predictor variable (breakfast consumption) is negatively related to the outcome variable (GPA). Despite (X, Y) units being situated variably above and below the regression line, this method of analysis indicates that, on average, unit increases in a student's breakfast consumption result in a lowering of that student's GPA.

Aggregation

Aggregation of data deals with the issues of hierarchical data analysis differently than disaggregation: Instead of ignoring higher level group differences, aggregation ignores lower level individual differences. Level-1 variables are raised to higher hierarchical levels (e.g., level-2 or level-3) and information about individual variability is lost. In aggregated statistical models, within-group variation is ignored and individuals are treated as homogenous entities (Beaubien et al., 2001; Gill, 2003; Osborne, 2000). To the researcher investigating the impact of breakfast consumption on student GPA, this approach changes the research question (Osborne, 2000). Mean classroom GPA becomes the new outcome variable of interest, rather than
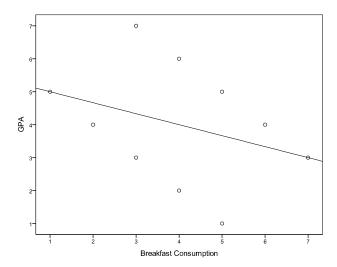

Figure 1. The relationship between breakfast consumption and student GPA using the disaggregation method. Figure is adapted from an example by Snijders & Bosker (1999) and Stevens (2007).

Table 3. Sample dataset using the aggregation method, with level-1 variables excluded from the data (dataset is adapted from an example by Snijders & Bosker, 1999)

Teacher ID (Level-2)	Classroom GPA (Level-2)	Classroom Breakfast Consumption (Level-2)
1	6	2
2	5	3
3	4	4
4	3	5
5	2	6

student GPA. Also, variation in students' breakfast habits is no longer measurable; instead, the researcher must use mean classroom breakfast consumption as the predictor variable (see Table 3 and Figure 2). Up to 80-90% of variability due to individual differences may be lost using aggregation, resulting in dramatic misrepresentations of the relationships between variables (Raudenbush & Bryk, 1992). HLM is generally recommended over aggregation for dealing with nested data because it effectively disentangles individual and group effects on the outcome variable.

In Figure 2, depicting the relationship between classroom breakfast consumption and classroom GPA using aggregation, the predictor variable (breakfast consumption) is again negatively related to the outcome variable (GPA). In this method of analysis, all (X, Y) units are situated on the regression line, indicating that unit increases in a classroom's mean breakfast consumption perfectly predict a lowering of that classroom's mean GPA. Although a negative relationship between breakfast consumption and GPA is found using both disaggregation and aggregation techniques, breakfast consumption is found to impact GPA more unfavourably using aggregation.
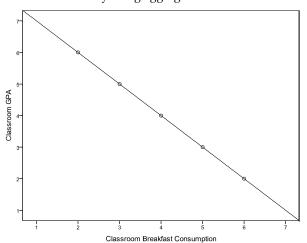
HLM

Figure 3 depicts the relationship between breakfast consumption and student GPA using HLM. Each level-1 (X,Y) unit (i.e., each student's GPA and breakfast consumption) is identified by its level-2 cluster (i.e., that student's classroom). Each level-2 cluster's slope (i.e., each classroom's slope) is also identified and analyzed separately. Using HLM, both the within- and between-group regressions are taken into account to depict the relationship between breakfast consumption and GPA. The resulting analysis indicates that breakfast consumption is positively related to GPA at level-1 (i.e., at the student level) but that the intercepts for these slope effects are influenced by level-2 factors [i.e., students' breakfast consumption and GPA (X, Y) units are also affected by classroom level factors]. Although disaggregation and aggregation methods indicated a negative relationship between breakfast consumption and GPA, HLM indicates that unit increases in breakfast consumption actually positively impact GPA. As demonstrated, HLM takes into consideration the impact of

Figure 2. The relationship between classroom breakfast consumption and classroom GPA using the aggregation method. Figure is adapted from an example by Snijders & Bosker (1999) and Stevens (2007).
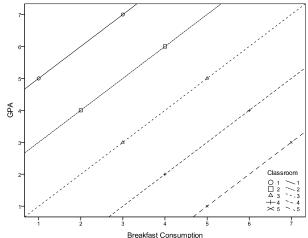

Figure 3. The relationship between breakfast consumption and student GPA using HLM. Figure is adapted from an example by Snijders & Bosker (1999) and Stevens (2007).

factors at their respective levels on an outcome of interest. It is the favored technique for analyzing hierarchical data because it shares the advantages of disaggregation and aggregation without introducing the same disadvantages.

As highlighted in this example, HLM can be ideally suited for the analysis of nested data because it identifies the relationship between predictor and outcome variables, by taking both level-1 and level-2 regression relationships into account. Readers who are interested in exploring the differences yielded by aggregation and disaggregation methods of analysis compared to HLM are invited to experiment with the datasets provided. Level-1 and level-2 datasets are provided to allow readers to follow along with the HLM tutorial in section 4 and to practice running an HLM. An aggregated version of these datasets is also provided for readers who would like to compare the results yielded from an HLM to those yielded from a regression.

In addition to HLM's ability to assess cross-level data relationships and accurately disentangle the effects of between- and within-group variance, it is also a preferred method for nested data because it requires fewer assumptions to be met than other statistical methods (Raudenbush & Bryk, 2002). HLM can accommodate non-independence of observations, a lack of sphericity, missing data, small and/or discrepant group sample sizes, and heterogeneity of variance across repeated measures. Effect size estimates and standard errors remain undistorted and the potentially meaningful variance overlooked using disaggregation or aggregation is retained (Beaubien, Hamman, Holt & Boehm-Davis, 2001; Gill, 2003; Osborne, 2000).

A disadvantage of HLM is that it requires large sample sizes for adequate power. This is especially true when detecting effects at level-1. However, higher-level effects are more sensitive to increases in groups than to increases in observations per group. As well, HLM can only handle missing data at level-1 and removes groups with missing data if they are at level-2 or above. For both of these reasons, it is advantageous to increase the number of groups as opposed to the number of observations per group. A study with thirty groups with thirty observations each (n = 900) can have the same power as one hundred and fifty groups with five observations each (n = 750; Hoffman, 1997).

Equations Underlying Hierarchical Linear Models

We will limit our remaining discussion to two-level hierarchical data structures concerning continuous outcome (dependent) variables as this provides the most thorough, yet simple, demonstration of the statistical features of HLM. We will be using the notation employed by Raudenbush and Bryk (2002; see Raudenbush & Bryk, 2002 for three-level models; see Wong & Mason, 1985 for dichotomous outcome variables). As stated previously, hierarchical linear models allow for the simultaneous investigation of the relationship within a given hierarchical level, as well as the relationship across levels. Two models are developed in order to achieve this: one that reflects the relationship within lower level units, and a second that models how the relationship within lower level units varies between units (thereby correcting for the violations of aggregating or disaggregating data; Hofmann, 1997). This modeling technique can be applied to any situation where there are lower-level units (e.g., the student-level variables) nested within higher-level units (e.g., classroom level variables).

To aid understanding, it helps to conceptualize the lower-level units as individuals and the higher-level units as groups. In two-level hierarchical models, separate level-1 models (e.g., students) are developed for each level-2 unit (e.g., classrooms). These models are also called within-unit models as they describe the effects in the context of a single group (Gill, 2003). They take the form of simple regressions developed for each individual i:

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

where:

Y_{ij} = dependent variable measured for ith level-1 unit nested within the jth level-2 unit,

X_{ij} = value on the level-1 predictor,

β_{0j} = intercept for the jth level-2 unit,

β_{1j} = regression coefficient associated with X_{ij} for the jth level-2 unit, and

r_{ij} = random error associated with the ith level-1 unit nested within the jth level-2 unit.

In the context of our example, these variables can be redefined as follows:

Y_{ij} = GPA measured for student i in classroom j

X_{ij} = breakfast consumption for student i in classroom j

β_{0j} = GPA for student i in classroom j who does not eat breakfast

β_{1j} = regression coefficient associated with breakfast consumption for the jth classroom

r_{ij} = random error associated with student i in classroom j.

As with most statistical models, an important assumption of HLM is that any level-1 errors (r_{ij}) follow a normal distribution with a mean of 0 and a variance of σ^2 (see Equation 2; Sullivan, Dukes & Losina, 1999). This applies to any level-1 model using continuous outcome variables.

$$E(r_{ij}) = 0; \quad \text{var}(r_{ij}) = \sigma^2 \qquad (2)$$

In the level-2 models, the level-1 regression coefficients (β_{0j} and β_{1j}) are used as outcome variables and are related to each of the level-2 predictors. Level-2 models are also referred to as between-unit models as they describe the variability across multiple groups (Gill, 2003). We will consider the case of a single level-2 predictor that will be modeled using Equations 3 and 4:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}G_j + U_{0j} \qquad (3)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}G_j + U_{1j} \qquad (4)$$

where:

β_{0j} = intercept for the jth level-2 unit;

β_{1j} = slope for the jth level-2 unit;

G_j = value on the level-2 predictor;

γ_{00} = overall mean intercept adjusted for G;

γ_{10} = overall mean intercept adjusted for G;

γ_{01} = regression coefficient associated with G relative to level-1 intercept;

γ_{11} = regression coefficient associated with G relative to level-1 slope;

U_{0j} = random effects of the jth level-2 unit adjusted for G on the intercept;

U_{1j} = random effects of the jth level-2 unit adjusted for G on the slope.

In the context of our example, these variables can be redefined as follows:

β_{0j} = intercept for the jth classroom;

β_{1j} = slope for the jth classroom;

G_j = teaching style in classroom j;

γ_{00} = overall mean intercept adjusted for breakfast consumption;

γ_{10} = overall mean intercept adjusted for breakfast consumption;

γ_{01} = regression coefficient associated with breakfast consumption relative to level-2 intercept;

γ_{11} = regression coefficient associated with breakfast consumption relative to level-2 slope;

U_{0j} = random effects of the jth level-2 unit adjusted for breakfast consumption on the intercept;

U_{1j} = random effects of the jth level-2 unit adjusted for breakfast consumption on the slope.

It is noteworthy that the level-2 model introduces two new terms (U_{0j} and U_{1j}) that are unique to HLM and differentiate it from a normal regression equation. Furthermore, the model developed would depend on the pattern of variance in the level-1 intercepts and slopes (Hofmann, 1997). For example, if there was no variation in the slopes across the level-1 models, G_j would no longer be meaningful given that β_{1j} is equivalent across groups and

would thus be removed from Equation 3 (Hofmann, 1997). Special cases of the two-level model Equations 1, 3 and 4 can be found in Raudenbush & Bryk (1992).

The assumption in the level-2 model (when errors are homogeneous at both levels) is that β_{0j} and β_{1j} have a normal multivariate distribution with variances defined by τ_{00} and τ_{11} and means equal to γ_{00} and γ_{10}. Furthermore, the covariance between β_{0j} and β_{1j} (defined as τ_{01}) is equal to the covariance between U_{0j} and U_{1j}. As in the level-1 assumptions, the mean of U_{0j} and U_{1j} is assumed to be zero and level-1 and level-2 errors are not correlated. Finally, the covariance between U_{0j} and r_{ij} and the covariance of U_{1j} and r_{ij} are both zero (Sullivan et al., 1999). The assumptions of level-2 models can be summarized as follows (Raudenbush & Bryk, 2002; Sullivan et al., 1999):

$$E(U_{0j}) = 0; \quad E(U_{1j}) = 0$$
$$E(\beta_{0j}) = \gamma_{00}; \quad E(\beta_{1j}) = \gamma_{01}$$
$$\text{var}(\beta_{0j}) = \text{var}(U_{0j}) = \tau_{00}; \quad \text{var}(\beta_{1j}) = \text{var}(U_{1j}) = \tau_{11} \ (5)$$
$$\text{cov}(\beta_{0j}, \beta_{1j}) = \text{cov}(U_{0j}, U_{1j}) = \tau_{01}$$
$$\text{cov}(U_{0j}, r_{ij}) = \text{cov}(U_{1j}, r_{ij}) = 0$$

In order to allow for the classification of variables and coefficients in terms of the level of hierarchy they affect (Gill, 2003), a combined model (i.e., two-level model; see Equation 6) is created by substituting Equations 3 and 4 into Equation 1:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}G_j \qquad (6)$$
$$+ \gamma_{11}G_jX_{ij} + U_{1j}X_{ij} + U_{0j} + r_{ij}$$

The combined model incorporates the level-1 and level-2 predictors (X_{ij} or breakfast consumption and G_j or teaching style), a cross-level term (G_jX_{ij} or teaching style × breakfast consumption) as well as the composite error ($U_{1j}X_{ij} + U_{0j} + r_{ij}$). Equation 6 is often termed a mixed model because it includes both fixed and random effects (Gill, 2003). Please note that fixed and random effects will be discussed in proceeding sections.

A comparison between Equation 6 and the equation for a normal regression (see Equation 7) further highlights the uniqueness of HLM.

$$Y_i = \beta_0 + \beta_1X_{i1} + ... + \beta_nX_{in} + \varepsilon_i \qquad (7)$$

As stated previously, the HLM model introduces two new terms (U_{0j} and U_{1j}) that allow for the model to estimate error that normal regression cannot. In Equation 6, the errors are no longer independent across the level-1 units. The terms U_{0j} and U_{1j} demonstrate that there is dependency among the level-1 units nested within each level-2 unit. Furthermore, U_{0j} and U_{1j} may have different values within level-2 units, leading to heterogeneous variances of the error terms (Sullivan et al., 1999). This dependency of errors has

Table 4. Hypothesis and necessary conditions: Does student breakfast consumption and teaching style influence student GPA?

	Hypotheses
1	Breakfast consumption is related to GPA.
2	Teaching style is related to GPA, after controlling for breakfast consumption.
3	Teaching style moderates the breakfast consumption-GPA relationship.
	Conditions
1	There is systematic within- & between-group variance in GPA.
2	There is significant variance at the level-1 intercept.
3	There is significant variance in the level-1 slope.
4	The variance in the level-1 intercept is predicted by teaching style.
5	The variance in the level-1 slope is predicted by teaching style.

important implications for parameter estimation, which will be discussed in the next section.

Estimation of Effects

Two-level hierarchical models involve the estimation of three types of parameters. The first type of parameter is fixed effects, and these do not vary across groups (Hofmann, 1997). The fixed effects are represented by γ_{00}, γ_{01}, γ_{11} and γ_{10} in Equations 3 and 4. While the level-2 fixed effects could be estimated via the Ordinary Least Squares (OLS) approach, it is not an appropriate estimation strategy as it requires the assumption of homoscedasticity to be met. This assumption is violated in hierarchical models as the accuracy of level-1 parameters are likely to vary across groups (e.g., classrooms; Hofmann, 1997). The technique used to estimate fixed effects is called a Generalized Least Squared (GLS) estimate. A GLS yields a weighted level-2 regression which ensures that groups (e.g., classrooms) with more accurate estimates of the outcome variable (i.e., the intercepts and slopes) are allocated more weight in the level-2 regression equation (Hofmann, 1997). Readers seeking further information on the estimation of fixed effects are directed to Raudenbush & Bryk (2002).

The second type of parameter is the random level-1 coefficients (β_{0j} and β_{1j}) which are permitted to vary across groups (e.g., classrooms; Hofmann, 1997). Hierarchical models provide two estimates for random coefficients of a given group (e.g., classroom): (1) computing an OLS regression for the level-1 equation representing that group (e.g., classroom); and (2) the predicted values of β_{0j} and β_{1j} in the level-2 model [see Equations 3 and 4]. Of importance is which estimation strategy provides the most precise values of the population slope and intercept for the given group (e.g., classroom; Hofmann, 1997). HLM software programs use an empirical Bayes estimation strategy, which takes into consideration both estimation strategies by computing an optimally weighted combination of the two (Raudenbush & Bryk, 2002; Raudenbush, Bryk, Cheong,

Congdon & du Toit, 2006). This strategy provides the best estimate of the level-1 coefficients for a particular group (e.g., classroom) because it results in a smaller mean square error term (Raudenbush, 1988). Readers interested in further information concerning empirical Bayes estimation are directed to Carlin and Louis (1996).

The final type of parameter estimation concerns the variance-covariance components which include: (1) the covariance between level-2 error terms [i.e., cov(U_{0j} and U_{1j}) or cov(β_{0j} and β_{1j}) defined as τ_{01}]; (2) the variance in the level-1 error term (i.e., the variance of r_{ij} denoted by σ^2); and (3) the variance in the level-2 error terms (i.e., the variance in U_{0j} and U_{1j} or β_{0j} and β_{1j} defined as τ_{00} and τ_{11}, respectively). When sample sizes are equal and the distribution of level-1 predictors is the same across all groups (i.e., the design is balanced), closed-form formulas can be used to estimate variance-covariance components (Raudenbush & Bryk, 2002). In reality, however, an unbalanced design is more probable. In such cases, variance-covariance estimates are made using iterative numerical procedures (Raudenbush & Bryk, 2002). Raudenbush & Bryk (2002) suggest the following conceptual approaches to estimating variance-covariance in unbalanced designs: (1) full maximum likelihood; (2) restricted maximum likelihood; and (3) Bayes estimation. Readers are directed to chapters 13 and 14 in Raudenbush & Bryk (2002) for more detail.

Hypothesis Testing

The previous sections of this paper provided an introduction to the logic, rationale and parameter estimation approaches behind hierarchical linear models. The following section will illustrate how hierarchical linear models can be used to answer questions relevant to research in any sub-field of psychology. It is prudent to note that for the sake of explanation, equations in the following section (which we will refer to as sub-models) purposely ignore one or a few facets of the combined model (see Equation 6) and are not

ad hoc equations. Through this section we will sequentially show how these sub-models can be used in order to run specific tests that answer hierarchical research questions. Thus the reader is reminded that all analyses presented in this section could be run all at once using the combined model (see Equation 6; Hofmann, 1997; Hofmann, personal communication, April 25, 2010) and an HLM software program. The following example was adapted from the model in Hofmann (1997). For more complex hypothesis testing strategies, please refer to Raudenbush and Bryk (2002).

Suppose that we want to know how GPA can be predicted by breakfast consumption, a student-level predictor, and teaching style, a classroom-level predictor. Recall that the combined model used in HLM is the following:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}G_j \\ + \gamma_{11}G_jX_{ij} + U_{1j}X_{ij} + U_{0j} + r_{ij} \quad (8)$$

Substituting in our variables the combined model would look like this:

$$GPA_{ij} = \gamma_{00} + \gamma_{10}(Breakfast_{ij}) + \gamma_{01}(Teaching_j) \\ + \gamma_{11}(Teaching_j)(Breakfast_{ij}) \\ + U_{1j}(Breakfast_{ij}) \\ + U_{0j} + r_{ij} \quad (9)$$

Our three hypotheses are a) breakfast consumption is related to GPA; b) teaching style is related to GPA, after controlling for breakfast consumption; and c) teaching style moderates the breakfast consumption-GPA relationship. In order to support these hypotheses, HLM models require five conditions to be satisfied. Our hypotheses and necessary conditions to be satisfied are summarized in Table 4.

Condition 1: There is Systematic Within- and Between-Group Variance in GPA

The first condition provides useful preliminary information and assures that there is appropriate variance to investigate the hypotheses. To begin, HLM applies a one-way analysis of variance (ANOVA) to partition the within- and between-group variance in GPA, which represents breakfast consumption and teaching style, respectively. The relevant sub-models (see Equations 10 and 11) formed using select facets from Equation 9 are as follows:

$$\text{Level-1: } GPA_{ij} = \beta_{0j} + r_{ij} \quad (10)$$

$$\text{Level-2: } \beta_{0j} = \gamma_{00} + U_{0j} \quad (11)$$

where:

β_{0j} = mean GPA for classroom j;

γ_{00} = grand mean GPA ;

$\text{Variance}(r_{ij}) = \sigma^2$ = within group variance in GPA;

$\text{Variance}(U_{0j}) = \tau_{00}$ = between group variance in GPA.

The level-1 equation above [see Equation 10] includes only an intercept estimate; there are no predictor variables. In cases such as this, the intercept estimate is determined by regressing the variance in GPA onto a unit vector, which yields the variable's mean (HLM software performs this implicitly when no predictors are specified). Therefore, at level-1, GPA is equal to the classroom's mean plus the classroom's respective error. At level-2 [see Equation 11], each classroom's GPA is regressed onto a unit vector, resulting in a constant (γ_{00}) that is equal to the mean of the classroom means. As a result of this regression, the variance within groups (σ^2) is forced into the level-1 residual (r_{ij}) while the variance between groups (τ_{00}) is forced into the level-2 residual (U_{0j}).

HLM tests for significance of the between-group variance (τ_{00}) but does not test the significance of the within-group variance (σ^2). In the abovementioned model, the total variance in GPA becomes partitioned into its within and between group components; therefore Variance(GPA$_{ij}$) = Variance $(U_{0j} + r_{ij})$ = $\tau_{00} + \sigma^2$. This allows for the calculation of the ratio of the between group variance to the total variance, termed the intra-class correlation (ICC). In other words, the ICC represents the percent of variance in GPA that is between classrooms. Thus by running an initial ANOVA, HLM provides: (1) the amount of variance within groups; (2) the amount of variance between groups; and (3) allows for the calculation of the ICC using Equation 12.

$$ICC = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \quad (12)$$

Once this condition is satisfied, HLM can examine the next two conditions to determine whether there are significant differences in intercepts and slopes across classrooms.

Conditions 2 and 3: There is Significant Variance in the Level-1 Intercept and Slope

Once within- and between-group variance has been partitioned, HLM applies a random coefficient regression to test the second and third conditions. The second condition supports hypothesis 2 because a significant result would indicate significant variance in GPA due to teaching style when breakfast consumption is held constant. The third condition supports hypothesis 3 by indicating that GPAs differ when students are grouped by the teaching style in their classroom. This regression is also a direct test of hypothesis 1, that breakfast consumption is related to GPA. The following sub-models (see Equations 13- 15) are created

using select facets of Equation 9:

Level-1: $GPA_{ij} = \beta_{0j} +$
$$\beta_{1j}(\text{BreakfastConsumption})_{ij} + r_{ij} \tag{13}$$

$$\text{Level-2: } \beta_{0j} = \gamma_{00} + U_{0j} \tag{14}$$

$$\text{Level-2: } \beta_{1j} = \gamma_{10} + U_{1j} \tag{15}$$

where:

γ_{00} = mean of the intercepts across classrooms;

γ_{10} = mean of the slopes across classrooms (Hypothesis 1);

$\text{Variance}(r_{ij}) = \sigma^2$ = Level-1 residual variance;

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

$\text{Variance}(U_{1j}) = \tau_{11}$ = variance in slopes.

The γ_{00} and γ_{10} parameters are the level-1 coefficients of the intercepts and the slopes, respectively, averaged across classrooms. HLM runs a t-test on these parameters to assess whether they differ significantly from zero, which is a direct test of hypothesis 1 in the case of γ_{10}. This t-test reveals whether the pooled slope between GPA and breakfast consumption differs from zero.

A χ^2 test is used to assess whether the variance in the intercept and slopes differs significantly from zero (τ_{00} and τ_{11}, respectively). At this stage, HLM also estimates the residual level-1 variance and compares it to the estimate from the test of Condition 1. Using both estimates, HLM calculates the percent of variance in GPA that is accounted for by breakfast consumption (see Equation 16).

$$R^2_{\text{level-1 model}} = \frac{(\sigma^2_{\text{oneway ANOVA}} - \sigma^2_{\text{random regression}})}{\sigma^2_{\text{oneway ANOVA}}} \tag{16}$$

Of note is that in order for the fourth and fifth conditions to be tested, the second and third conditions must first be met.

Condition 4: The Variance in the Level-1 Intercept is Predicted by Teaching Style

The fourth condition assesses whether the significant variance at the intercepts (found in the second condition) is related to teaching style. It is also known as the intercepts-as-outcomes model. HLM uses another random regression model to assess whether teaching style is significantly related to the intercept while holding breakfast consumption constant. This is accomplished via the following sub-models (see Equations 17-19) created from using select variables in Equation 9:

Level-1: $GPA_{ij} = \beta_{0j}$
$$+ \beta_{1j}(\text{BreakfastConsumption})_{ij} + r_{ij} \tag{17}$$

$$\text{Level-2: } \beta_{0j} = \gamma_{00} + \gamma_{01}(\text{TeachingStyle})_j + U_{0j} \tag{18}$$

$$\text{Level-2: } \beta_{1j} = \gamma_{10} + U_{1j} \tag{19}$$

where:

γ_{00} = Level-2 intercept;

γ_{01} = Level-2 slope (Hypothesis 2);

γ_{10} = mean (pooled) slopes;

$\text{Variance}(r_{ij}) = \sigma^2$ = Level-1 residual variance;

$\text{Variance}(U_{0j}) = \tau_{00}$ = residual intercept variance;

$\text{Variance}(U_{1j}) = \tau_{11}$ = variance in slopes.

The intercepts-as-outcomes model is similar to the random coefficient regression used for the second and third conditions except that it includes teaching style as a predictor of the intercepts at level-2. This is a direct test of the second hypothesis, that teaching style is related to GPA after controlling for breakfast consumption. The residual variance (τ_{00}) is assessed for significance using another χ^2 test. If this test indicates a significant value, other level-2 predictors can be added to account for this variance. To assess how much variance in GPA is accounted for by teaching style, the variance attributable to teaching style is compared to the total intercept variance (see Equation 20).

$$R^2_{\text{level-2 intercept model}} =$$
$$\frac{(\tau_{00 \text{ - random regression}} - \tau_{00 \text{ - intercepts-as-outcomes}})}{\tau_{00 \text{ - random regression}}} \tag{20}$$

Condition 5: The Variance in the Level-1 Slope is Predicted by Teaching Style

The fifth condition assesses whether the difference in slopes is related to teaching style. It is known as the slopes-as-outcomes model. The following sub-models (see Equations 21-23) formed with select variables from Equation 9 are used to determine if condition five is satisfied.

Level-1: $GPA_{ij} = \beta_{0j} +$
$$\beta_{1j}(\text{BreakfastConsumption})_{ij} + r_{ij} \tag{21}$$

$$\text{Level-2: } \beta_{0j} = \gamma_{00} + \gamma_{01}(\text{TeachingStyle})_j + U_{0j} \tag{22}$$

$$\text{Level-2: } \beta_{1j} = \gamma_{10} + \gamma_{11}(\text{TeachingStyle})_j + U_{1j} \tag{23}$$

where:

γ_{00} = Level-2 intercept;

γ_{01} = Level-2 slope (Hypothesis 2);

γ_{10} = Level-2 intercept;

γ_{11} = Level-2 slope (Hypothesis 3);

$\text{Variance}(r_{ij}) = \sigma^2$ = Level-1 residual variance;

$\text{Variance}(U_{0j}) = \tau_{00}$ = residual intercept variance;

$\text{Variance}(U_{1j}) = \tau_{11}$ = residual slope variance.

With teaching style as a predictor of the level-1 slope, U_{1j} becomes a measure of the residual variance in the averaged level-1 slopes across groups. If a χ^2 test on U_{1j} is significant, it indicates that there is systematic variance in the level-1 slopes that is as-of-yet unaccounted for, therefore other level-2 predictors can be added to the model. The slopes-as-

outcomes model is a direct test of hypothesis 3, that teaching style moderates the breakfast consumption-GPA relationship. Finally, the percent of variance attributable to teaching style can be computed as a moderator in the breakfast consumption-GPA relationship by comparing its systematic variance with the pooled variance in the slopes (see Equation 24).

$$R^2_{\text{level-2 slope model}} =$$

$$\frac{(\tau_{11 \text{ - intercept-as-outcomes}} - \tau_{11 \text{ - slopes-as-outcomes}})}{\tau_{11 \text{ - intercept-as-outcomes}}} \quad (24)$$

Model Testing – A Tutorial

To illustrate how models are developed and tested using HLM, a sample data set was created to run the analyses. Analysis was performed using HLM software version 6, which is available for download online (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2006). For the purposes of the present demonstration, a two-level analysis will be conducted using the logic of HLM.

Sample Data

The sample data contains measures from 300 basketball players, representing 30 basketball teams (10 players per team). Three measures were taken: *Player Successful Shots on Net* (Shots_On_5), *Player Life Satisfaction* (Life_Satisfaction), and *Coach Years of Experience* (Coach_Experience). Scores for Shots_On_5 ranged from 0 shots to 5 shots; where higher scores symbolized more success. Life_Satisfaction scores ranged from 5 to 25 with higher scores representing life satisfaction and lower scores representing life dissatisfaction. Finally, Coach_Experience scores ranged from 1 to 3, with the number representing their years of experience. The level-1 predictor (independent; individual) variable is Shots_On_5; the level-2 predictor (independent; group) variable is Coach_Experience, and the outcome (dependent) variable is Life_Satisfaction. The main hypotheses were as follows: 1) the number of successful shots on net predicts ratings of life satisfaction, and 2) coach years of experience predict variance in life satisfaction.

For the purposes of the present analysis, it is assumed that all assumptions of HLM are adequately met. Specifically, there is no multicollinearity, the Shots_On_5 residuals are independent and normally distributed, and Shots_On_5 and Coach_Experience are independent of their level-related error and their error terms are independent of each other (for discussion on the assumptions of HLM, see Raudenbush and Bryk, 2002).

Preparation

It is essential to prepare the data files using a statistical software package before importing the data structure into the HLM software. The present example uses PASW (Predictive Analytics SoftWare) version 18 (Statistical Package for the Social Sciences; SPSS). A separate file is created for each level of the data in PASW. Each file should contain the participants' scores on the variables for that level, plus an identification code to link the scores between levels. It is important to note that the identification code variable must be in string format, must contain the same number of digits for all levels, and must be given the exact same variable name at all levels. The data file must also be sorted, from lowest value to highest value, by the identification code variable (see Figure 4).

In this example, the level-1 file contains 300 scores for the measures of Shots_on_5 and Life_Satisfaction, where participants were assigned identification codes (range: 01 to 30) based on their team membership. The level-2 file contains 30 scores for the measure of Coach_Experience and identification codes (range: 01 to 30), which were associated with the appropriate players from the level-1 data. Once a data file has been created in this manner for each level, it is possible to import the data files into the HLM software.

HLM Set-Up

The following procedures were conducted according to those outlined by Raudenbush and Bryk (2002). After launching the HLM program, the analysis can begin by clicking *File → Make New MDM File → Stat Package Input*. In the dialogue box that appears, select the *MDM* (Multivariate Data Matrix). We will select *HLM2* to continue because our example has two levels. A new dialogue box will open, in which we will specify the file details, as well as load the level-1 and level-2 variables.

First, specify the variables for the analysis by linking the file to the level-1 and level-2 SPSS data sets that were created. Once both have been selected, click *Choose Variable* to select the desired variables from the data set (check the box next to *In MDM*) and specify the identification code variables (check the box next to *ID*). Please note that you are not required to select all of the variables from the list to be in the MDM, but you must specify an ID variable. You must also specify whether there are any missing data and how missing data should be handled during the analyses. If you select *Running Analyses* for the missing data, HLM will perform a pairwise deletion; if you select *Making MDM*, HLM will perform a listwise deletion. In the next step,

Figure 4. Example of SPSS data file as required by HLM. The image on the left represents the data for level-1. The image on the right represents the data for level-2.

ensure that under the *Structure of Data* section, *Cross sectional* is selected. Under *MDM File Name,* provide a name for the current file, add the extension ".mdm", and ensure that the input file type is set to *SPSS/Windows.* Finally, in the *MDM Template File* section, choose a name and location for the template files.

To run the analyses, click *Make MDM,* and then click *Check Stats.* Checking the statistics is an invaluable step that should be performed carefully. At this point, the program will indicate any specific missing data. After this process is complete, click *Done* and a new window will open where it is possible to build the various models and run the required analyses. Before continuing, ensure that the optimal output file type is selected by clicking *File → Preferences.* In this window, it is possible to make a number of adjustments to the output; however, the most important is to the *Type of Output.* For the clearest and easiest to interpret output file, it is strongly recommended that *HTML output* is selected as well as *view HTML in default browser.*

Unconstrained (null) Model

As a first step, *a one-way analysis of variance* is performed to confirm that the variability in the outcome variable, by level-2 group, is significantly different than zero. This tests whether there are any differences at the group level on the outcome variable, and confirms whether HLM is necessary. Using the dialogue box, Life_Satisfaction is entered into the model as an "outcome variable" (see Figure 5). The program will also generate the level-2 model required to ensure that the level-1 model is estimated in terms of the level-2 groupings (Coach_Experience). Click *Run Analysis,* then *Run the Model Shown* to run the model and view the output screen. The generated output should be identical to Figure 6.

The results of the first model test yield a number of different tables. For this model, the most important result to examine is the chi-square test (x^2) found within the *Final Estimation of Variance Components* table in Figure 6. If this result is statistically significant, it indicates that there is variance in the outcome variable by the level-2 groupings, and that there is statistical justification for running HLM analyses. The results for the present example indicate that $x^2(29) = 326.02$, $p < .001$; which supports the use of HLM.

As an additional step, the ICC can be calculated to determine which percentage of the variance in Life_Satisfaction is attributable to group membership and which percentage is at the individual level. There is no consensus on a cut-off point, however if the ICC is very low, the HLM analyses may not yield different results from a traditional analysis. The ICC (see Equation 12) can be calculated using the σ^2 (level-1) and τ (level-2) terms at the top of the output, under the *Summary of the model specified*
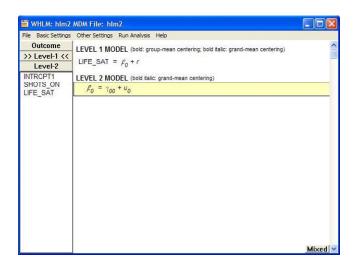
Figure 5. Building the unconstrained (null) model in HLM.

heading in Figure 6 (see Equation 25).

$$\text{ICC} = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$
$$= \frac{14.96}{14.96 + 14.61} \qquad (25)$$
$$= 0.506$$

In the present example, σ^2 = 14.61 and τ = 14.96, which results in an ICC of 0.506. This result suggests that 51% of the variance in Life Satisfaction is at the group level and 49% is at the individual level.

Random Intercepts Model

Next, test the relationship between the level-1 predictor variable and the outcome variable. To test this, return to the dialogue box and add Shots_on_5 as a *variable group centered* in level-1. In most cases, the level-1 predictor variable is entered as a group centered variable in order to study the effects of the level-1 and level-2 predictor variables independently and to yield more accurate estimates of the intercepts. We would select *variable grand centered* at level-1 if we were not interested in analyzing the predictor variables separately (e.g. an ANCOVA analysis, which tests one variable while controlling for the other variable, would require grand centering). Leave the outcome variable (Life_Satisfaction) as it was for the first model and ensure that both error terms (U_{0j} and U_{1j}) are selected in the "Level 2 Model" (see Figure 7). By selecting both error terms, the analyses include estimates of both the between- and within-error. Specifically, U_{0j} starts with the assumption that life satisfaction varies from team to team and U_{1j} starts with the assumption that strength of the relationship between Shots_on_5 and Life_Satisfaction varies from group to

Specifications for this HLM2 run

Problem Title: no title

The data source for this run = m.test
The command file for this run = C:\DOCUME~1\hlm2.html
Output file name = C:\Documents and Settings\hlm2.html
The maximum number of level-1 units = 300
The maximum number of level-2 units = 30
The maximum number of iterations = 100

Method of estimation: restricted maximum likelihood

The outcome variable is LIFE_SAT

Summary of the model specified

Level-1 Model

$$LIFE\_SAT_{ij} = \beta_{0j} + r_{ij}$$

Level-2 Model

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

Mixed Model

$$LIFE\_SAT_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

Final Results - Iteration 2

Iterations stopped due to small change in likelihood function

$\sigma^2 = 14.60563$

τ

INTRCPT1,β_0 14.95920

Random level-1 coefficient	Reliability estimate
INTRCPT1,β_0	0.911

The value of the log-likelihood function at iteration 2 = -8.621511E+002

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	14.806670	0.739815	20.014	29	<0.001

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	14.806670	0.727380	20.356	29	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, u_0	3.86771	14.95920	29	326.02031	<0.001
level-1, r	3.82173	14.60563			

Statistics for current covariance components model

Deviance = 1724.302126
Number of estimated parameters = 2

Figure 6. HLM output tables – Unconstrained (null) model. This represents a default output in HLM.
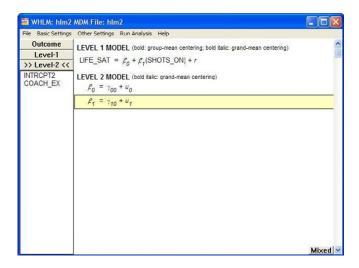
Figure 7. Building the random intercepts model in HLM.

group. Click *Run Analysis* to run this model and view the output screen. The generated output screen should be identical to Figure 8.

A regression coefficient is estimated and its significance confirms the relationship between the level-1 predictor variable and the outcome variable. To view results of this analysis, consult the significance values for the INTRCPT2, γ_{01} in the *Final estimation of fixed effect* output table (refer to Figure 8), which is non-standardized. The non-standardized final estimation of fixed effects tables will be similar to the standardized table (i.e. with robust standard errors) unless an assumption has been violated (e.g. normality), in which case, use the standardized final estimation. The results of the present analysis support the relationship between Shots_on_5 and Life_Satisfaction, $b = 2.89$, $p < .001$. Please note that the direction (positive or negative) of this statistic is interpreted like a regular regression.

To calculate a measure of effect size, calculate the variance (r^2) explained by the level-1 predictor variable in the outcome variable using Equation 26.

$$r^2 = \frac{(\sigma^2_{null} - \sigma^2_{random})}{\sigma^2_{null}}$$
$$= \frac{(14.61 - 4.61)}{14.61} \tag{26}$$
$$= 0.715$$

Note that σ^2_{null} is the sigma value obtained in the previous step (null-model testing) under the *Summary of the model specified* heading in Figure 6 ($\sigma^2_{null} = 14.61$). The σ^2_{random} is sigma value found at the top of the output under the *Summary of the model specified* heading in Figure 8 (in the present example, $\sigma^2_{random} = 4.61$). Using the values and the specified equation, the results indicate that Shots_on_5

Specifications for this HLM2 run

Problem Title: no title

The data source for this run = m.test
The command file for this run = C:\DOCUME~1\hlm2.html
Output file name = C:\Documents and Settings\hlm2.html
The maximum number of level-1 units = 300
The maximum number of level-2 units = 30
The maximum number of iterations = 100

Method of estimation: restricted maximum likelihood

The outcome variable is LIFE_SAT

Summary of the model specified

Level-1 Model

$$LIFE\_SAT_{ij} = \beta_{0j} + \beta_{1j} * (SHOTS\_ON_{ij}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

SHOTS_ON has been centered around the group mean.

Mixed Model

$$LIFE\_SAT_{ij} = \gamma_{00} + \gamma_{10}*SHOTS\_ON_{ij} + u_{0j} + u_{1j}*SHOTS\_ON_{ij} + r_{ij}$$

Final Results - Iteration 11

Iterations stopped due to small change in likelihood function

$\sigma^2 = 4.60875$

τ

INTRCPT1,β_0	15.95889	1.39404
SHOTS_ON,β_1	1.39404	0.66499

τ (as correlations)

INTRCPT1,β_0	1.000	0.428
SHOTS_ON,β_1	0.428	1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1,β_0	0.972
SHOTS_ON,β_1	0.539

The value of the log-likelihood function at iteration 11 = -7.182666E+002

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	14.806670	0.739822	20.014	29	<0.001
For SHOTS_ON slope, β_1					
INTRCPT2, γ_{10}	2.889137	0.202235	14.286	29	<0.001

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	14.806670	0.727380	20.356	29	<0.001
For SHOTS_ON slope, β_1					
INTRCPT2, γ_{10}	2.889137	0.198575	14.549	29	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, u_0	3.99486	15.95889	29	1033.19332	<0.001
SHOTS_ON slope, u_1	0.81547	0.66499	29	64.93933	<0.001
level-1, r	2.14680	4.60875			

Statistics for current covariance components model

Deviance = 1436.533205
Number of estimated parameters = 4

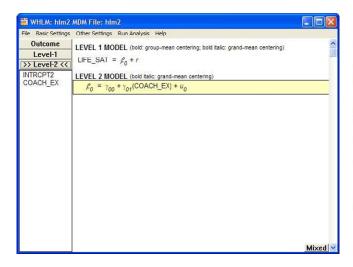*Figure 8.* HLM output tables – Random intercepts model.

Figure 9. Building the means as outcomes model in HLM. explains 71.5% of the variance in Life_Satisfaction.

Means as Outcomes Model

The next step is to test the significance and direction of the relationship between the level-2 predictor variable and the outcome variable. To test this, return to the dialogue box and remove Shots_on_5 as a *group centered* predictor variable in level-1 by selecting *delete variable from model* and leave the outcome variable (Life_Satisfaction). Add Coach_Experience as a *grand centered* predictor variable at level-2 (see Figure 9). The issue of centering at level-2 is not as important as it is at level-1 and is only necessary when we are interested in controlling for the other predictor variables. When examining the level-1 and level-2 predictor variables separately, centering will not change the regression coefficients but will change the intercept value. When the level-2 predictor variable is centered, the level-2 intercept is equal to the grand mean of the outcome variable. When the level-2 predictor variable is not centered, the level-2 intercept is equal to the mean score of the outcome variable when the level-2 predictor variables equal zero. In the current example, a mean score of zero at level-2 is not of much interest given that coach experience scores ranged from 1 through 3, therefore the *grand centered* option was appropriate. When interested in the slopes and not the intercepts, centering is not usually an issue at level-2. Click *Run Analysis* to run the model and view the output screen. The output generated should be identical to Figure 10.

A regression coefficient is estimated and, as before, its significance confirms the relationship between the level-2 predictor variable and the outcome variable (at level-1). To view the results, see COACH_EX y_{01} in the output under the *Final estimation of fixed effects* table in Figure 10. The results

Specifications for this HLM2 run

Problem Title: no title

The data source for this run = m.test
The command file for this run = C:\DOCUME~1\hlm2.html
Output file name = C:\Documents and Settings\hlm2.html
The maximum number of level-1 units = 300
The maximum number of level-2 units = 30
The maximum number of iterations = 100

Method of estimation: restricted maximum likelihood

The outcome variable is LIFE_SAT

Summary of the model specified

Level-1 Model

$$LIFE\_SAT_{ij} = \beta_{0j} + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * (COACH\_EX) + u_{0j}$$

COACH_EX has been centered around the grand mean.

Mixed Model

$$LIFE\_SAT_{ij} = \gamma_{00} + \gamma_{01} * COACH\_EX_j + u_{0j} + r_{ij}$$

Final Results - Iteration 2

Iterations stopped due to small change in likelihood function

$\sigma^2 = 14.60563$

τ
INTRCPT1,β_0 1.67564

Random level-1 coefficient	Reliability estimate
INTRCPT1,β_0	0.534

The value of the log-likelihood function at iteration 2 = -8.384908E+002

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	14.806670	0.323327	45.795	28	<0.001
COACH_EX, γ_{01}	4.784303	0.429936	11.128	28	<0.001

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	14.806670	0.312363	47.402	28	<0.001
COACH_EX, γ_{01}	4.784303	0.373833	12.798	28	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, u_0	1.29447	1.67564	28	60.12326	<0.001
level-1, r	3.82173	14.60563			

Statistics for current covariance components model

Deviance = 1676.981685
Number of estimated parameters = 2

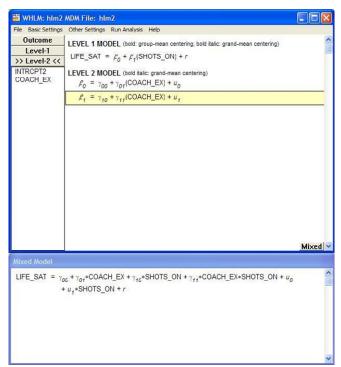*Figure 10.* HLM output tables – Means as outcomes model.

Figure 11. Building the random intercepts and slopes model in HLM. The mixed model must be obtained by clicking on "Mixed".

of this analysis support that Coach_Experience predicts Life_Satisfaction, $b = 4.78$, $p < .001$. For a measure of effect size, the explained variance in the outcome variable, by the level-2 predictor variable can be computed using Equation 27.

$$r^2 = \frac{(\tau^2_{null} - \tau^2_{means})}{\tau^2_{null}}$$
$$= \frac{(14.96 - 1.68)}{14.96} \qquad (27)$$
$$= 0.888$$

where τ^2_{null} is the τ value obtained in the first step (null-model testing) under the *Summary of the model specified* table in Figure 6 ($\tau^2_{null} = 14.96$). Next τ^2_{means} is the τ value obtained under the *Summary of the model specified* table in the present analysis ($\tau^2_{means} = 1.68$; Figure 10). The results confirm that Coach_Experience explains 88.8% of the between measures variance in Life_Satisfaction.

Random Intercepts and Slopes Model

The final step is to test for interactions between the two predictor variables (level-1 and level-2). Please note that if only interested in the main effects of both predictor variables (level-1 and level-2), this final step is not necessary. Alternatively, this final model could be used to test the two previous models instead of running them separately. If you

Problem Title: no title

The data source for this run = m.test
The command file for this run = C:\DOCUME~1\hlm2.html
Output file name = C:\Documents and Settings\hlm2.html
The maximum number of level-1 units = 300
The maximum number of level-2 units = 30
The maximum number of iterations = 100

Method of estimation: restricted maximum likelihood

The outcome variable is LIFE_SAT

Summary of the model specified

Level-1 Model

$$LIFE\_SAT_{ij} = \beta_{0j} + \beta_{1j}*(SHOTS\_ON_{ij}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(COACH\_EX_j) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}*(COACH\_EX_j) + u_{1j}$$

SHOTS_ON has been centered around the group mean.
COACH_EX has been centered around the grand mean.

Mixed Model

$$LIFE\_SAT_{ij} = \gamma_{00} + \gamma_{01}*COACH\_EX_j$$
$$+ \gamma_{10}*SHOTS\_ON_{ij} + \gamma_{11}*COACH\_EX_j*SHOTS\_ON_{ij}$$
$$+ u_{0j} + u_{1j}*SHOTS\_ON_{ij} + r_{ij}$$

Final Results - Iteration 16

Iterations stopped due to small change in likelihood function

$\sigma^2 = 4.60669$

τ
INTRCPT1,β_0 2.67553   0.34947
SHOTS_ON,β_1 0.34947   0.62695

τ (as correlations)
INTRCPT1,β_0 1.000   0.270
SHOTS_ON,β_1 0.270   1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1,β_0	0.853
SHOTS_ON,β_1	0.525

The value of the log-likelihood function at iteration 16 = -6.940699E+002

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	14.806670	0.323326	45.795	28	<0.001
COACH_EX, γ_{01}	4.784303	0.429935	11.128	28	<0.001
For SHOTS_ON slope, β_1					
INTRCPT2, γ_{10}	2.884745	0.199480	14.461	28	<0.001
COACH_EX, γ_{11}	0.381153	0.269730	1.413	28	0.169

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	14.806670	0.312363	47.402	28	<0.001
COACH_EX, γ_{01}	4.784303	0.373833	12.798	28	<0.001
For SHOTS_ON slope, β_1					
INTRCPT2, γ_{10}	2.884745	0.193475	14.910	28	<0.001
COACH_EX. γ_{11}	0.381153	0.260682	1.462	28	0.155

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, u_0	1.63571	2.67553	28	190.62228	<0.001
SHOTS_ON slope, u_1	0.79180	0.62695	28	62.86446	<0.001
level-1, r	2.14632	4.60669			

Statistics for current covariance components model

Deviance = 1388.139702
Number of estimated parameters = 4

Figure 12. HLM output tables – Random intercepts and slopes model.

choose to run this final model instead of testing the main effects separately, be aware that the results will differ slightly because of the maximum likelihood estimation methods used to calculate the models.

To test this final model, return to the dialogue box and add Shots_on_5 as a *group centered* predictor variable in level-1, leave the remaining terms from the 3rd model, and add the level 2 predictor variable, Coach_Experience as a *grand centered* variable to both equations (f_0 and f_1). By adding it to both equations, the interaction term does not accidentally account for all of the variance. The error terms (U_{0j} and U_{1j}) should be selected for both equations (see Figure 11). Finally, click *Run Analysis* to run the model and view the output screen. The generated output should be identical to Figure 12.

For this output, we will focus on the interaction term only. The results of the interaction can be found under the *Final estimation of fixed effects* table of Figure 12 (see COACH_EX y_{11}). HLM results reveal that the interaction was not significant ($b = 0.38$, $p = .169$), providing support that there is no cross-level interaction between the level-1 and level-2 predictors.

Reporting the Results

Now that the analyses are complete, it is possible to summarize the results of the HLM analysis. The statistical analyses conducted in the present example can be summarized as follows:

Hierarchical linear modeling (HLM) was used to statistically analyze a data structure where players (level-1) were nested within teams (level-2). Of specific interest was the relationship between player's life satisfaction (level-1 outcome variable) and both the number of shots on the net (level-1 predictor variable) and their coach's experience (level-2 predictor variable). Model testing proceeded in 4 phases: unconstrained (null) model, random intercepts model, means-as-outcome model, and intercepts- and slopes-as-outcomes model.

The intercept-only model revealed an ICC of .51. Thus, 51% of the variance in life satisfaction scores is between-team and 49% of the variance in life satisfaction scores is between players within a given team. Because variance existed at both levels of the data structure, predictor variables were individually added at each level. The random-regression coefficients model was tested using players' shots on net as the only predictor variable. The regression coefficient relating player shots on net to life satisfaction was positive and statistically significant ($b = 2.89$, $p < .001$). Player's life satisfaction levels were higher when

their shots on net levels were also higher (relative to those whose shots on net were lower). Next, the means-as-outcomes model added coaches' experience as a level-2 predictor variable. The regression coefficient relating coaches' experience to player life satisfaction was positive and statistically significant ($b = 4.784$, $p < .001$). Life satisfaction levels were higher in teams with coaches who had more experience (relative to coaches who had less experience). Finally, the intercepts model and slopes-as-outcomes model were simultaneously tested with all predictor variables tested in the model to test the presence of any interactions between predictor variables. The cross-level interaction between shots on net and coaches' experience was not statistically significant ($b = 0.38$, $p = .169$); which means that the degree of coach experience had no influence on the strength of the relationship between shots on net and life satisfaction.

Conclusion

Since its inception in the 1970s, HLM has risen in popularity as the method of choice for analyzing nested data. Reasons for this include the high prevalence of hierarchically organized data in social sciences research, as well as the model's flexible application. Although HLM is generally recommended over disaggregation and aggregation techniques because of these methods' limitations, it is not without its own challenges.

HLM is a multi-step, time-consuming process. It can accommodate any number of hierarchical levels, but the workload increases exponentially with each added level. Compared to most other statistical methods commonly used in psychological research, HLM is relatively new and various guidelines for HLM are still in the process of development (Beaubien et al., 2001; Raudenbush & Bryk, 2002). Prior to conducting an HLM analysis, background interaction effects between predictor variables should be accounted for, and sufficient amounts of within- and between-level variance at all levels of the hierarchy should be ensured. HLM presumes that data is normally distributed: When the assumption of normality for the predictor and/or outcome variable(s) is violated, this range restriction biases HLM output. Finally, as previously mentioned, outcome variable(s) of interest must be situated at the lowest level of analysis in HLM (Beaubien et al., 2001).

Although HLM is relatively new, it is already being used in novel ways across a vast range of research domains. Examples of research questions analyzed using HLM include the effects of the environment on aspects of youth development (Avan & Kirkwood, 2001; Kotch, et al., 2008; Lyons, Terry, Martinovich, Peterson, & Bouska, 2001),

longitudinal examinations of symptoms in chronic illness (Connelly, et al., 2007; Doorenbos, Given, Given, & Verbitsky, 2006), relationship quality based on sexual orientation (Kurdek, 1998), and interactions between patient and program characteristics in treatment programs (Chou, Hser, & Anglin, 1998).

Throughout this tutorial we have provided an introduction to HLM and methods for dealing with nested data. The mathematical concepts underlying HLM and our theoretical hypothesis testing example represent only a small and simple example of the types of questions researchers can answer via this method. More complex forms of HLM are presented in *Hierarchical Linear Models: Applications and Data Analysis Methods, Second Edition* (Raudenbush & Bryk, 2002). Readers seeking information on statistical packages available for HLM and how to use them are directed to *HLM 6: Hierarchical Linear and Nonlinear Modeling* (Raudenbush et al., 2006).

References

Avan, B. I. & Kirkwood, B. (2010). Role of neighbourhoods in child growth and development: Does 'place' matter? *Social Science & Medicine, 71*, 102-109.

Beaubien, J. M., Hamman, W. R., Holt, R. W., & Boehm-Davis, D. A. (2001). The application of hierarchical linear modeling (HLM) techniques to commercial aviation research. *Proceedings of the 11th annual symposium on aviation psychology,* Columbus, OH: The Ohio State University Press.

Carlin, B. P., & Louis, T. A. (1996). *Bayes and empirical bayes methods for data analysis*. London, CRC Press LLC.

Castro, S. L. (2002). Data analytic methods for the analysis of multilevel questions: A comparison of intraclass correlation coefficients, rwg(j), hierarchical linear modeling, within- and between-analysis, and random group resampling. *The Leadership Quarterly, 13*, 69-93.

Chou, C-P., Hser, Y. I., & Anglin, M. D. (1998). Interaction effects of client and treatment program characteristics on retention: An exploratory analysis using hierarchical linear models. *Substance Use & Misuse, 33*(11), 2281-2301.

Connelly, M., Keefe, F. J., Affleck, G., Lumley, M. A., Anderson, T., & Waters, S. (2007). Effects of day-to-day affect regulation on the pain experience of patients with rheumatoid arthritis. *Pain, 131*(1-2), 162-70.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological), 39*(1), 1-38.

Dempster, A. P., Rubin, D. B., & Tsutakawa, R. K. (1981). Estimation in covariance components models. *Journal of the American Statistical Association, 76*(374), 341-353.

Doorenbos, A. Z., Given, C. W., Given, B., & Verbitsky, N. (2006). Symptom experience in the last year of life among individuals with cancer. *Journal of Pain & Symptom Management, 32*(5), 403-12.

Gill, J. (2003). Hierarchical linear models. In Kimberly Kempf-Leonard (Ed.), *Encyclopedia of social measurement.* New York: Academic Press.

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management*, 23, 723-744.

Kotch, J. B., Lewis, T., Hussey, J. M., English, D., Thompson, R., Litrownik, A. J., … & Dubowitz (2008). Importance of early neglect for childhood aggression. *Pediatrics, 121*(4), 725-731.

Kurdek, L. (1998). Relationship outcomes and their predictors: Longitudinal evidence from heterosexual married, gay cohabitating, and lesbian cohabitating couples. *Journal of Marriage and the Family, 60*, 553-568.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B (Methodological), 34*(1), 1-41.

Lyons, J. S., Terry, P., Martinovich, Z., Peterson, J., & Bouska, B. (2001) Outcome trajectories for adolescents in residential treatment: A statewide evaluation. *Journal of Child and Family Studies, 10*(3), 333-345.

Osborne, J. W. (2000). Advantages of hierarchical linear modeling. *Practical Assessment, Research, and Evaluation, 7*(1), 1-3.

Raudenbush, S. W. (1998). Educational applications of hierarchical linear models: A review. *Journal of Educational Statistics*, 13, 85-116.

Raudenbush, S. W. & Bryk, A. S. (1992). *Hierarchical linear models.* Newbury Park, CA: Sage.

Raudenbush, S. W. Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods, second edition*. Newbury Park, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2006). *HLM 6: Hierarchical linear and nonlinear modeling.* Linconwood, IL: Scientific Software International, Inc.

Smith, A. F. M. (1973). A general Bayesian linear model. *Journal of the Royal Statistical Society, Series B (Methodological), 35*(1), 67-75.

Snijders, T. & Bosker, R. (1999). Multilevel analysis. London: Sage Publications.

Stevens, J. S. (2007). Hierarchical linear models. Retrieved April 1, 2010, from http://www.uoregon.edu/~stevensj /HLM/data/.

Sullivan, L. M., Dukes, K. A., & Losina, E. (1999). Tutorial in

biostatistics: An introduction to hierarchical linear modeling. *Statistics in Medicine, 18*, 855-888.

Wong, G. Y., & Mason, W. M. (1985). The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association*, *80,* 513-524.