Multilevel Modeling: Practical Examples to Illustrate a Special Case of SEM

Lee Branum-Martin
Georgia State University

Author Note: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grants R305D090024 (Paras D. Mehta, PI) and R305A10272 (Lee Branum-Martin, PI) to the University of Houston. The initial data collection was jointly funded by NICHD (HD39521) and IES (R305U010001) to UH (David J. Francis, PI). The opinions expressed are those of the author and do not represent views of these funding agencies. The author would like to thank Paras D. Mehta for his helpful guidance in writing this chapter. The author would like to acknowledge helpful feedback from Yaacov Petscher and Pat Taylor. Inquiries should be sent to Lee Branum-Martin, Georgia State University, Department of Psychology, PO Box 5010, Atlanta GA 30302-5010. BranumMartin@gsu.edu

Multilevel Modeling: Practical Examples to Illustrate a Special Case of SEM

1. Introduction

Multilevel models are becoming commonplace in education to account for the clustering of students within classrooms. Multilevel models have been developed in numerous fields to account for the clustering of observations within other units, such as repeated measures within students, or classrooms within schools (see overview by Kreft & de Leeuw, 1998). In this chapter, it is assumed that readers have a general familiarity with multilevel models. Practical examples of how to fit multilevel models are numerous (West, Welch, & Gałecki 2007), with one of the best and most cited being Singer (1998). An excellent conceptual introduction to multilevel methods with connections to different historical traditions can be found in Kreft and de Leeuw (1998). More complete technical accounts of multilevel models can be found in Raudenbush and Bryk (2002), Goldstein (2003), and Snijders and Bosker (1999).

There is a growing recognition that multilevel models are models of covariance, or dependence among individuals nested within social clusters (Curran, 2003; Mehta & Neale, 2005). A structural equation model (SEM) is a highly flexible method for handling covariance among variables (Bollen, 1989) Lomax, this volume). This chapter will exploit this overlap in modeling covariance in both multilevel models and SEM through the use of applied examples. Most approaches to multilevel models and SEM are tightly tied to a single version of equations, diagrams, and software code, with the connections among these three representations often becoming obscured to the user. In addition, some software contains inherent limitations to the number of levels or the ways in which those levels may be connected.

In order to overcome these limitations of other approaches, we present a new framework for fitting multilevel models which is based in SEM, has a natural connection to our substantive notions of nested levels, and helps to make explicit how equations, diagrams, and software code...
represent the same substantive phenomena we wish to test in our empirical models. The current chapter does this by introducing xxM (Mehta, 2012), a package for the free statistical software, R (R Development Core Team, 2011).

The purpose of the current chapter is to provide practical, step by step examples of fitting simple multilevel models, with equations, diagrams, and code. Singer (1998) and West, Welch, and Galecki (2007) provided practical examples of multilevel models with code for SAS PROC MIXED. Mehta and Neale (2005) provided code and technical details of a complete translation between PROC MIXED and the more general approach of fitting a structural equation model (SEM). The current chapter extends these previous sources in order to emphasize conceptual and practical commonalities across multilevel models and SEM.

The current chapter will focus on two issues: (a) connections among the parameters represented in equations, diagrams, and code, and (b) an introduction to the new but highly extendable programming code of xxM. It is hoped that the current chapter may facilitate connections between multilevel approaches (e.g., mixed effects vs. multilevel SEM) and software (e.g., SAS PROC MIXED, Mplus, HLM, lme4 in R).

Three software packages will be used in this chapter: SAS PROC MIXED, Mplus 6.1, and xxM. SAS PROC MIXED is a general purpose mixed effects program (Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006), coupled with the highly general data management and statistics software, SAS (SAS Institute Inc., 2010). Mplus 6.1 fits numerous kinds of latent variable and two-level models (B. O. Muthén, 2001, 2002; L. K. Muthén & Muthén, 2010). xxM is a package for the free software, R (R Development Core Team, 2011). xxM is designed for fitting SEMs with essentially no limit to the number of levels of nesting, including complicated structures of cross-classification and multiple membership. Demonstrations and updates for xxM
can be found at http://xxm.times.uh.edu. A user’s guide can be downloaded from there with several worked examples of multilevel models and the features of the program (see also Mehta, this volume).

1.1. An example dataset: Passage comprehension scores

The purpose of this chapter is to demonstrate connections between multilevel model parameters in equations, diagrams, and code. To be sure, multilevel analysis opens conceptual and statistical issues important for careful consideration. Some of these issues will be considered here in the context of an applied data analysis. Generally speaking, multilevel models are used in education for examining how scores vary across shared environments, such as how variation in student reading or math scores might differ across classrooms or schools (or how repeated measures may differ across persons, who also may share classrooms and schools). Such variation at the cluster level is important to understand, since it could not only obscure student level effects such as background characteristics or the influence of other skills (an issue of power or appropriateness of standard errors), but cluster-level variation may also indicate environmental level effects, such as cohort/classmate influence, instruction, or other potentially ecological effects (Raudenbush & Sampson, 1999). Some of these issues will be commented upon in the following applied examples. Excellent overviews of the broader conceptual issues can be found in introductory sources (Hox, 1995; Kreft & de Leeuw, 1998; Snijders & Bosker, 1999), with detailed technical accounts for more advanced readers (Goldstein, 2003; Raudenbush & Bryk, 2002).

In order to keep the discussion applied, let us consider an example dataset of 802 students in 93 classrooms (ignoring schools for the current example) measured in the spring of their first grade year on reading passage comprehension. The measure was taken from the Woodcock
Language Proficiency Battery-Revised, a test developed on a nationally representative sample (Woodcock, 1991). The test is scored in W-units, a Rasch-based metric. A substantive question for this dataset is “to what extent do classrooms vary in their English passage comprehension scores?”

**Figure 1: Boxplots of student passage reading comprehension scores within classrooms**

Figure 1 shows boxplots for each of 93 classrooms of student passage comprehension scores. The classrooms have been sorted in order of ascending means. The horizontal line at W-score 443 represents the overall sample mean across all students. Each box represents one classroom, with a diamond representing the classroom mean, the middle line within each box representing the classroom median, the upper and lower ends of the box representing the upper
and lower quartiles (25th and 75th centiles), and the whiskers extending to 1.5 times the range between quartiles. Open circles represent student scores outside the whiskers.

Figure 1 represents the essential problem for nested data: clusters such as classrooms differ in their average level of performance. If we were to randomly assign students to a reading intervention versus regular instruction, variability due to classrooms would weaken power to detect a true treatment difference (Raudenbush, Bryk, & Congdon, 2008; Raudenbush, Martínez, & Spybrook, 2007; Spybrook et al., 2011). That is, we might worry that noisy variation among classrooms might overwhelm our ability to find a true average treatment effect for students. Even small amounts of cluster differences have been shown to reduce the power to detect mean differences (Barcikowski, 1981; Walsh, 1947). This issue of increasing precision is a major motivation for using multilevel models in evaluating educational programs.

On a less technical note, Figure 1 illustrates the problem of clustering in a compelling way: classrooms differ systematically. That is, students in a high-performing classroom are more likely to score higher than students in a low performing classroom. If we put this lack of independence another way, we can say that student scores in the same classroom are likely to be related, perhaps due to the shared influence of their environment (i.e., partly due to instruction, partly due to selection or assignment to particular teachers, as well as partly due to influence of higher levels of nesting at the school, district, and community).

A fundamental assumption of most statistical models is that conditional on all variables used in the model (predictors), student scores (residuals) are independent. In the case of clustered data, as Figure 1 suggests, residual scores are not likely to be independent, unless classroom is used in the model. Insofar as a multilevel model handles this dependence across students due to classrooms, multilevel models are a model of covariance (Mehta & Neale, 2005). Because SEM
is a model of covariance, connecting the concepts between SEM and multilevel modeling can be helpful to increase our understanding of the connections among our theories, data, and models. Explaining this covariance and connecting SEM to multilevel models is one of the goals of this chapter.

1.2. Analysis: How to handle classrooms?

Analyzing these data while ignoring classrooms would clearly leave out a sizeable systematic factor. Alternatively, analyzing the means of each of 93 classrooms would vastly reduce the amount of information. In a case where there are other variables, such as predictors, the student level relations might be different than those at the classroom level. Collapsing across levels might result in distorted estimates of the relations among variables—an ecological fallacy (Burstein, 1980; Cronbach, 1976; Robinson, 1950; Thorndike, 1939). For the present case with a single outcome and no predictors (yet), it would be substantively interesting to simultaneously investigate how much variability there is in passage comprehension due to students and due to classrooms.

There is a long history of discussions and alternative models to account for multilevel data (Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002). One approach would be to add a set of predictors for each individual classroom, such as 92 dummy variables in a regression or ANOVA model. Such propositions raise substantive issues regarding sampling as well as issues regarding statistical estimation. If the classrooms in the data do not constitute a random sample from a universe of possible classrooms—i.e., we wish to estimate their differences and not generalize beyond this current set of classrooms, then fixed effects represented by dummy indicators or an ANOVA may be appropriate. That is, the estimates are fixed because the sample is fixed. However, if the sample is drawn at random from a universe of classrooms (or at least
plausibly so), and we can assume that the outcome variable is normally distributed (or distributed appropriately according to the model we are using), then it can be efficient to estimate the overall mean and variance, rather than to fit separate models for each classroom: we can estimate two parameters instead of one parameter for each classroom.

2. A univariate model: students nested within classrooms

The most common and widely known model for the nesting of students within classrooms is for a single outcome. Such models have been referred to as multilevel regression models and mixed effects models, as they combine fixed with random regression coefficients (Kreft & de Leeuw, 1998). For a single outcome, \( Y \), with no predictors for \( i \) students in \( j \) classrooms, we can fit a univariate random effects model:

\[
Y_{ij} = \beta_{0j} + e_{ij} \quad [1]
\]

where \( Y_{ij} \) is the model-predicted outcome score for student \( i \) in classroom \( j \), \( \beta_{0j} \) is the regression intercept for classroom \( j \), and \( e_{ij} \) is the deviation (residual—sometimes noted as \( r_{ij} \)) for that student. The residuals are assumed to be distributed normally over students with mean zero and variance \( \sigma^2 \). At level 2 for classrooms, the intercept is decomposed into two parts:

\[
\beta_{0j} = \gamma_{00} + u_{0j} \quad [2]
\]

where \( \gamma_{00} \) is the grand intercept (fixed), and the \( u_{0j} \) is the random deviation from the grand intercept for that classroom. These fixed and random components are the reason such models are often referred to as “mixed effects models.” The random classroom deviations, \( u_{0j} \), are assumed to be distributed normally with variance \( \tau_{00} \). Most readers will recognize that equations 1-2 can be combined:

\[
Y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \quad [3]
\]
Equation 3 compactly states that each individual score is the sum of a grand intercept, a classroom deviation, and a student deviation. The distributional assumptions are worth examining closer. At the student level, residual scores are assumed to be distributed normally\(^1\) with mean zero and variance \(\sigma^2\), noted \(\mathcal{N}[0, \sigma^2]\). The mean is zero because these are deviations from the overall model-predicted mean, \(\gamma_{00}\). At the classroom level, classrooms deviate from this grand mean with a mean of zero and variance \(\tau_{00}\), noted \(\mathcal{N}[0, \tau_{00}]\). It can be confusing to beginners that one variance is squared and the other is not (but some authors choose to use \(\tau^2\); Snijders & Bosker, 1999). We will later change this notation in order to reinforce that these parameters are both variances and to emphasize the connections with SEM.

Equation 3 can be represented graphically, with the helpful addition that variances are shown in Figure 2. Standard rules of path tracing apply with a strict correspondence to each parameter in Equations 1-3. That is, we can follow the arrows additively to calculate the model predicted score for any given child in any given classroom. Additionally, the hierarchical structure of the data and the equation for students nested within classrooms is visually reinforced by vertical stacking.

Figure 2 is a full, explicit SEM representation using typical HLM notation. This representation is useful to bring together Equations 1-3 along with their variance components, as is typical in SEM. Figure 2 also reinforces the idea that random effects are latent variables, usually represented in the Equations 1-3 with implicit (omitted) unit regression weight. That is, these latent variables, \(u_{0j}\) and \(e_{ij}\), have fixed factor loadings (Mehta & Neale, 2005). We will explore the notion of factor loadings in more detail later.

---

\(^1\) Other distributions can be used, and there are extensive diagnostics which can be used to test these assumptions (Hilden-Minton, 1995).
From an SEM perspective, we also notice that $\beta_{0j}$ is a phantom variable: a latent variable with no variance. In this sense, it serves only as a placeholder in the notation (no separate values are estimated in the model). We have included it here only for explicit correspondence to the typical equations.

More importantly, however, we may wish to fit different kinds of models for students and teachers. In the same way that mixed effects models can be represented by equations at each level, we introduce a few changes in order to open interesting possibilities later. First, we separate the levels into two sub-models (one for students and one for classrooms) and explicitly link them. Second, following SEM notation at the classroom level, we adopt $\eta$ as our symbol for latent classroom deviations, with mean $\alpha$ and variance $\psi$. Third, at the student level, we use SEM notation for the residual variance: $\theta$. 
Figure 3 shows the results of these changes, making it correspond closely to standard SEM. The most important parameters, the variances of the random effects ($\psi$ and $\theta$) are shown with the fixed effect ($\alpha$).

**Figure 3: Diagram of a univariate two level model for students within classrooms**

In this model, we are interested in estimating three parameters: the latent mean across classrooms ($\alpha$), the variance of classroom deviations ($\psi$), and the variance of the student residuals ($\theta$). The cross-level link is fixed at unity and not estimated (and as we will see, is usually handled automatically by software). For the sake of completeness, now we can compact the mean structure and rewrite Equation 3 using SEM notation from Figure 3:

$$Y_{ij} = 1 \eta_j + 1 e_{ij} \quad \text{[4]}$$

where the values of $\eta_j$ are distributed $\mathcal{N}[\alpha, \psi]$ and $e_{ij}$ are distributed $\mathcal{N}[0, \theta]$. The first term, $1 \eta_j$, shows the linking of each student score, $Y_{ij}$, by a unit weight design matrix: each student gets counted in their respective classroom. Thus, each student can be thought of as a
variable indicating the latent classroom mean, with a fixed factor loading of unity (Mehta & Neale, 2005). Each classroom has a latent mean, $\eta_j$, with grand mean $^{2} \alpha$, with variance $\psi$.

Additionally, the individual residual, $e_{ij}$, is shown with an explicit unit link. These parameters for this model ($\alpha$, $\psi$, $\theta$) respectively represent those from conventional mixed effects models ($\gamma_{00}$, $\tau_{00}$, $\sigma^2$), such as those in HLM, PROC MIXED, or other multilevel representations (Mehta & Neale, 2005). We have chosen to represent the unit links explicitly in order to invoke the notion of a factor loading from SEM: random effects in multilevel models are latent variables in SEM.

In order to understand the connection between Equation 4 and Figure 3, it can be helpful to visualize the data structure implied by the model. We will then illustrate this model in an expanded, explicit way to demonstrate how a multilevel model is a model of covariance. First, we can imagine a simple dataset of three classrooms of two students each:

<table>
<thead>
<tr>
<th>Student</th>
<th>Classroom</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>$Y_{11}$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>$Y_{21}$</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>$Y_{32}$</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>$Y_{42}$</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>$Y_{53}$</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>$Y_{63}$</td>
</tr>
</tbody>
</table>

Each outcome score in italics shows the individual student scores, $Y_{ij}$, where the subscripts indicate the student ($i$) and respective classroom ($j$). Then, in matrix form, Equation 4 would match to the dataset of six students in three classrooms in the following way:

\[ \begin{pmatrix} 
Y_{11} \\
Y_{21} \\
Y_{32} \\
Y_{42} \\
Y_{53} \\
Y_{63} 
\end{pmatrix} \]

\[ = \begin{pmatrix} 
\alpha + \eta_j \\
\alpha + \eta_j \\
\alpha + \eta_j \\
\alpha + \eta_j \\
\alpha + \eta_j \\
\alpha + \eta_j 
\end{pmatrix} + \begin{pmatrix} 
\epsilon_{11} \\
\epsilon_{21} \\
\epsilon_{32} \\
\epsilon_{42} \\
\epsilon_{53} \\
\epsilon_{63} 
\end{pmatrix} \]

\[ = \begin{pmatrix} 
\alpha + \eta_j + \epsilon_{11} \\
\alpha + \eta_j + \epsilon_{21} \\
\alpha + \eta_j + \epsilon_{32} \\
\alpha + \eta_j + \epsilon_{42} \\
\alpha + \eta_j + \epsilon_{53} \\
\alpha + \eta_j + \epsilon_{63} 
\end{pmatrix} \]

Alternatively, we could represent Equations 1 and 2 with classroom deviations in SEM notation (following Figure 2). Depending on the context, one representation may be better than the other for interpretation or communication purposes (they are mathematically equal). Here, we choose the compact version of $\eta_j$ centered around $\alpha$. 

\(^{2}\) Alternatively, we could represent Equations 1 and 2 with classroom deviations in SEM notation (following Figure 2). Depending on the context, one representation may be better than the other for interpretation or communication purposes (they are mathematically equal). Here, we choose the compact version of $\eta_j$ centered around $\alpha$. 
Equation 5 is the matrix version of Equation 4 and shows the outcome scores on the left side \((Y_{ij})\). On the right side, the full matrix of unit links for each classroom’s deviation \((\eta_j)\) is actually a \(i \times j\) design matrix, allocating each block of two students to their respective classroom (Mehta & Neale, 2005). This matrix is usually constructed automatically by software, simply by identifying the classroom cluster variable (e.g., teacher ID). The second matrix is a column vector of latent variables to be estimated for each classroom. In this case, there is one latent \(\eta\) for each of the three classrooms. Finally, each student has a residual score, deviating from the model prediction for the classroom. In the next sections, we will see how this model can be specified in software script.

**Figure 4: SEM diagram of two students each in three classrooms (expanded, multivariate layout)**
Figure 4 completes the representation of Equation 5. Each classroom has a latent factor, indicated by its students. At the top of Figure 4, a single mean is estimated to be common across latent classroom factors (separate latent means are constrained equal)—classroom latent factors deviate from this mean. Similarly, the variances of the classroom factors are constrained equal, since we are interested in estimating the population variability from this sample of classrooms.

At the bottom of Figure 4, each student in the model from Equation 5 is represented explicitly. Residual variances are constrained to a single value ($\theta$), because the model assumes that conditional on classroom factors, student scores behave as if drawn randomly from a homogeneous population.

The cross-level linking matrix in the middle of Figure 4 is represented as classroom-specific paths from three latent variables, one per classroom. This representation conforms to the linking matrix from Equation 5, with blocks of ones in three columns. The factor loadings are fixed to unity for each student within the respective classroom. The links between students not in particular classrooms are constrained to zero and are therefore not shown in the diagram.

Figure 4 emphasizes that the multilevel latent variable for each classroom is a model of covariance between persons. That is, because people’s scores within a classroom are expected to covary due to their shared environment, the latent variable for each classroom is fit to them. Conditional on the classroom latent variable, students are independent of each other. This is akin to a model of parallel tests and in this way, people in a univariate multilevel model act as variables from a single level, multivariate model (Mehta & Neale, 2005).

The representation in Figure 4 can be extended to any number of persons per classroom and any number of classrooms. However, this expanded version can be quite cumbersome to specify. It is shown here only to emphasize that SEM as a model of covariances, multilevel
regression is a special case of SEM (Curran, 2003; Mehta & Neale, 2005), and also to clarify the role of the cross-level linking matrix as a matrix of factor loadings. Subsequently in this chapter, we will return to the more compact notation (Figure 3).

It is worth noting that this same shift from univariate to multivariate layout is conceptually the same shift used for multiple time points within persons, allowing individual growth models to be fit as latent variables in SEM (McArdle & Epstein, 1987; Mehta & West, 2000; Meredith & Tisak, 1990; B. O. Muthén, 1997).

2.1. PROC MIXED estimation of the univariate model

Let us return to the sample of 802 first grade students in 93 classrooms. In SAS PROC MIXED, the univariate random intercepts model would be specified:

```sas
proc mixed data=PCWA method=ML;
  class teacher;
  model PC = / solution;
  random intercept / subject=teacher;
run;
```

Similar to the examples provided in Singer (1998), this code specifies that `teacher` is a classification variable for the clustering of students in classrooms. The `model` statement gives the regression equation in which passage comprehension, PC, is specified without predictors. The `solution` portion of the `model` statement requests the estimate of the fixed effect: the latent mean across classrooms (α). The residual variance (θ) is estimated implicitly by default. The `intercept` is specified to be random across classrooms (`teacher`). The `random` statement produces the variance component (ψ) of the random effect (intercept, η) implicitly by default. The `method=ML` is specified in order to be comparable to typical SEM estimation (other estimators can be chosen).

The first main result of this PROC MIXED code are the covariance parameter estimates:
These show that the variance \( (\psi) \) across classroom intercepts \((\text{teacher}) = 89.8343\) and the residual variance \( (\theta) = 410.00\). If we add these variance components, the total variance of the model is 499.8 \((89.8 + 410.0)\). An important concept in multilevel models is the intraclass correlation, or proportion of variability due to clustering. The ICC is simply the ratio of cluster level variance over total variance \((89.9 / 499.8 = 0.18)\). In this case, the model suggests that 18% of the variance in the student level outcome is due to classroom clustering.

The second important part of the output shows the fixed effect of the intercept (the latent mean, \( \alpha \)) as 444.00 W-score units. This implies that in a perfectly average classroom, we would expect the classroom mean to be 444 W-score units. If we take the square root of the variance components above, then classrooms have an SD of 9.5 W-score units around this mean and students differ within their classrooms with an SD of 20.2 W-score units.

### Solution for Fixed Effects

| Effect  | Estimate | Standard Error | DF | t Value | Pr > |t| |
|---------|----------|----------------|----|---------|------|---|
| Intercept | 444.00 | 1.2806 | 92 | 346.70 | <.0001 |

### 2.2. Mplus estimation of the univariate model

In Mplus (L. K. Muthén & Muthén, 2010), the same model from Figure 3 would be specified by the following code:
In the Mplus specification, the DATA source and VARIABLES are named. The CLUSTER variable, teacher, is specified. The analysis type is TWOLEVEL, which allocates students to their classroom. The MODEL command specifies potentially different models for the WITHIN classroom portion (students), and the BETWEEN classroom portion (classrooms), using the percentage sign to delineate these sections. At the %WITHIN% level, specifying the outcome, PC, estimates a variance for that variable (θ). At the %BETWEEN% level, specifying that same variable estimates a between classroom variance (ψ) for that variable. The overall mean (α) is estimated implicitly by default, but we have chosen here to mark it explicitly (square brackets indicate mean structure in Mplus).

<table>
<thead>
<tr>
<th>Code</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE: Model 1 Passage Comprehension;</td>
<td>name of data file</td>
</tr>
<tr>
<td>DATA: FILE IS PCWA.txt;</td>
<td>data file</td>
</tr>
<tr>
<td>VARIABLE: NAMES = teacher student PC;</td>
<td>variable names for columns in file</td>
</tr>
<tr>
<td>CLUSTER = teacher;</td>
<td>cluster identification is teacher</td>
</tr>
<tr>
<td>USEVARIABLES = PC;</td>
<td>variable to be used in model</td>
</tr>
<tr>
<td>ANALYSIS: TYPE = TWOLEVEL;</td>
<td>two-level model</td>
</tr>
<tr>
<td>ESTIMATOR = ML;</td>
<td>estimator chosen to match MIXED</td>
</tr>
<tr>
<td>MODEL:</td>
<td></td>
</tr>
<tr>
<td>%WITHIN%</td>
<td>within classroom variance</td>
</tr>
<tr>
<td>PC;</td>
<td></td>
</tr>
<tr>
<td>%BETWEEN%</td>
<td>between classroom variance</td>
</tr>
<tr>
<td>PC;</td>
<td></td>
</tr>
<tr>
<td>[PC];</td>
<td>latent mean (freely estimated by default)</td>
</tr>
</tbody>
</table>

The loglikelihood of this model was identical to the MIXED result. The output is divided in the same way as the input, with the WITHIN level containing the residual variance, θ =

<table>
<thead>
<tr>
<th>MODEL RESULTS</th>
<th>Two-Tailed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>S.E.</td>
</tr>
<tr>
<td>Within Level</td>
<td></td>
</tr>
<tr>
<td>Variances</td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td>409.866</td>
</tr>
<tr>
<td>Between Level</td>
<td></td>
</tr>
<tr>
<td>Means</td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td>444.011</td>
</tr>
<tr>
<td>Variances</td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td>90.459</td>
</tr>
</tbody>
</table>
409.866, and the BETWEEN level containing the classroom variance, $\psi = 90.459$. The latent mean $\alpha = 444.011$. Standard errors (S.E.), Wald ratios (Est./S.E.), and p-values are printed by default. The differences of less than a unit in the two variance estimates between PROC MIXED and Mplus are negligible, given that the outcome has a standard deviation of more than 22 W-score units (square root of the total variance: $\sqrt{410+90}$).

### 2.3. xxM estimation of the univariate model

The above univariate model can also be specified in xxM (Mehta, 2012), a package for the free software, R (R Development Core Team, 2011). There are three stages to fitting a model in xxM: setup, matrix specification, and estimation. Each of these three stages will be presented in turn.

#### 2.3.1. xxM setup for the univariate model

In setting up an xxM model, we name the model and then name the submodels and their connections. As noted in Figure 3, the level 1 model is for students and level 2 is for classrooms. For the current model of a random intercept model for English passage comprehension, we can name the model “EngPC” and define the two levels of student and teacher. Each level requires a separate dataframe (this is the R term for a dataset: a table of rows for observations and columns for variables).

```r
EngPC <- xxmModel(levels = c("student", "teacher"))
```

The `xxmModel` function creates an R object with placeholders for student and teacher submodels. The "c( )" operator in R concatenates objects—in this case, a list of submodel names are created. Accordingly, we need to read the dataset into R, creating separate dataframes (i.e., data files) for each level. There are a number of methods for importing and exporting data.
in R (see sources in Appendix A). Next, we need to specify what these submodels are. The student level model is specified in the following way:

```
Code                          comment
EngPC <- xxmSubmodel(
  model = EngPC,             model to which this submodel belongs
  level = "student",        name of this submodel
  parents = "teacher",      levels to which this submodel is connected
  ys = "PC",               outcome variable(s)
  xs = ,                    predictors (none)
  latent = ,                latent variables (none)
  data = student)           name of R dataframe for this submodel
```

This code adds a `Submodel` to the model object, `EngPC`. Each of these seven submodel parameters is required, separated by commas. The model and level are named. The “parent” of this submodel, which represents the higher order clustering unit, is defined as “teacher.” This notes to `xxM` that a linkage will later be provided for students nested within teachers—in this sense, teachers are the model object “parents” of the students. There is a single dependent variable, named as the `ys`. There are no predictors or latent variables at the student level, so `xs` and `latent` variables are left blank. Finally, the dataframe is named `student`, to match the identification variable at that level.

Next, the teacher level `Submodel` is specified.

```
Code                          comment
EngPC <- xxmSubmodel(
  model = EngPC,             model to which this submodel belongs
  level = "teacher",        name of this submodel
  parents = ,                this submodel is not nested within another level
  ys = ,                    outcome variables (none)
  xs = ,                    predictors (none)
  latent = c("int"),        latent variable named for intercept (η)
  data = teacher )          name of R dataframe for this submodel
```
The only difference in this specification is that the `latent` keyword expects a list, so even if there is only one element, the `c()` operator must be used: `latent = c("int")`.

Specification for a Submodel is essentially the same for any given level.

### 2.3.2. xxM matrix specification for the univariate model

In xxM, matrices are specified as in the software program, R (R Development Core Team, 2011). The basic definition of a matrix in R follows the general form:

```r
matrix(vector, nrows, ncolumns)
```

where `vector` is a list of numbers, `nrows` is the number of rows in the matrix, and `ncolumns` is the number of columns. Readers interested in learning more about specifying matrices or the general use of R can see references in Appendix A.

In xxM, we need to specify the pattern of fixed and free elements of each matrix, followed by start values. In this univariate two-level model, we have three elements to be estimated: the latent mean ($\alpha$), classroom level variance ($\psi$), and student level/residual variance ($\theta$). These estimated elements are therefore “free.” In addition, there is a fixed element—the cross-level linking matrix—which is not estimated, but held to its value of unity. Figure 5 shows an overview of the diagram, matrices, and xxM matrix specifications. We show this to give an advanced overview of the model pieces before they are all put together in xxM.
Figure 5: Diagram, matrix, and xxM code for the model matrices of the univariate random intercepts model.

The left column of Figure 5 repeats the SEM diagram for the univariate model. The outer box shows that the model we have created is called “EngPC.” The “teacher” and “student” submodels are also shown (layering indicates implied repetition over units of observation, which in principle could be distinguished in a more complex model). In this diagram, we have left the residual, $e_{ij}$, to be implicit and instead only show its variance, $\theta$. Otherwise, the diagram matches Figure 3.

The middle column of Figure 5 shows the matrices of parameters in the model. These single-value matrices are compact versions of the matrices shown in Equation 5. The right-hand column shows the xxM code specification for each matrix. For each matrix, a pattern of free and fixed elements must be defined, where “1” represents a freely estimated parameter and “0” represents a fixed parameter. Thus, the line

\[
Y_{ij} = \theta_1 \psi_1 \alpha_1 \eta_j \theta
\]

The $SEM$ diagram...
alpha_pattern <- matrix(1, 1,1)

specifies a matrix named “alpha_pattern” consisting of a freely estimated element in a $1 \times 1$ matrix: the matrix $\alpha$ has one row and one column containing the estimated latent mean. The “alpha_value” matrix specifies that the start value will be 440 for the latent mean. In practice, any reasonable value will usually suffice, such as the observed mean of the variable.

Each of the subsequent $1 \times 1$ matrices are specified in the same way. The psi matrix has one freely estimated element ($\psi$). The residual variance matrix also has one freely estimated element ($\theta$). The residual variance represents the variability in the outcome not explained by the model, and any reasonable value less than the observed variance of the outcome can be used$^3$.

The cross-level linking matrix is fixed, specified as “(0, 1, 1)”. That is, the zero specifies a fixed (non-estimated) element in a $1 \times 1$ pattern matrix. The value matrix sets this fixed element to unity$^4$.

The pattern and values of the above four matrices, $\alpha, \psi, \lambda, \theta$, are specified in R code just as shown in Figure 5. Next, each of these four model matrices in Figure 5 must be connected to its role within the overall model. Matrices $\alpha, \psi, \theta$, belong within a particular level and are connected to the overall model with the function, xxmWithinMatrix. The cross-level connection matrix, $\lambda$, has its own function and will be shown later. Starting with the classroom level latent mean, $\alpha$, we would specify:

---

$^3$ For univariate (single outcome) models like the present one, starting values will not cause problems in the estimation. However, as models get larger and more complex, poor start values can increase the probability of estimation trouble, since the combination of parameters might produce illegal solutions, such as variances less than zero or correlations greater than 1.0. For complex models, we recommend users build them up in small model subsets so that such boundary conditions are understood and can be avoided.

$^4$ The interesting feature of this specification is that alternative linkages could be specified, such as fixing the link to student-specific values in the data set. For example, instead of each link being 1.0, link values could represent classroom attendance, instruction received, or individual-specific testing dates in a growth model (Mehta & Neale, 2005; Mehta & West, 2000). The second example in this chapter will explore this with a random slope model.
The above code specifies that an alpha matrix will be connected to the teacher level. The pattern and value matrices from Figure 5 are named here to be used in the model. There are two additional parameters at the end of the specification, “label” and “name” not used in this model (see random slopes model in the next section).

Next, we have the classroom level variance, \( \psi \), connected to the teacher level:

Finally, we have the cross-level linking matrix, \( \lambda \), which has its own function, \text{xxmBetweenMatrix}:
The “parent” statement notes the parent model within which the “child” model units are nested. That is, the model posits that influence flows from the “parent” to the “child” (alternative models would require different specification and are not covered here). As noted before, the lambda matrix has no free elements in this case. The lambda matrix has unit weights linking students to their classroom teachers.

### 2.3.3. xxM estimation of the univariate model

Now that all matrices have been attached to the model, we are ready to estimate the model with the xxmRun command:

```r
EngPC <- xxmRun(model = EngPC)
```

The results of the EngPC xxM model are the following.

<table>
<thead>
<tr>
<th>xxM Output</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>teacher_alpha</td>
<td>classroom mean matrix</td>
</tr>
<tr>
<td></td>
<td>colnames</td>
</tr>
<tr>
<td>rownames Mean or Intercept</td>
<td></td>
</tr>
<tr>
<td>int</td>
<td>444.0026</td>
</tr>
<tr>
<td>student theta</td>
<td>student covariance</td>
</tr>
<tr>
<td></td>
<td>[,1]</td>
</tr>
<tr>
<td>[1,] 410.0291</td>
<td>residual variance θ</td>
</tr>
<tr>
<td>teacher psi</td>
<td>classroom covariance</td>
</tr>
<tr>
<td></td>
<td>[,1]</td>
</tr>
<tr>
<td>[1,] 89.86626</td>
<td>classroom variance</td>
</tr>
</tbody>
</table>

The parameter estimates are the same as those shown previously. The specification of each matrix is given, with results shown in matrix format.
3. A model with a random slope

Predictors can be added to the model at any given level, with code that is quite straightforward for multilevel regression (Singer, 1998) or SEM (Lomax, this volume). The standard regression of a predictor is a simple idea to represent in an equation or diagram: the predictor relates to the outcome in a single, fixed way for all students. However, if we consider the possibility that the relation is different for each classroom—that each classroom has its own regression slope—then there are two latent variables for each classroom: one for the classroom intercept of the outcome ($\eta_{1j}$), and the other for the classroom relation of predictor to outcome ($\eta_{2j}$). Equation 5 expands with two columns for each classroom in the cross-level linking matrix, and the $\eta$ matrix is stacked for each of the three classrooms as a column vector:

\[
\begin{bmatrix}
Y_{11} \\
Y_{21} \\
Y_{32} \\
Y_{42} \\
Y_{53} \\
Y_{63}
\end{bmatrix} = \begin{bmatrix}
1 & X_{11} & 0 & 0 & 0 & 0 \\
1 & X_{21} & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & X_{32} & 0 & 0 \\
0 & 0 & 1 & X_{42} & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & X_{53} \\
0 & 0 & 0 & 0 & 1 & X_{63}
\end{bmatrix} \begin{bmatrix}
\eta_{11} \\
\eta_{21} \\
\eta_{12} \\
\eta_{22} \\
\eta_{13} \\
\eta_{23}
\end{bmatrix} + \begin{bmatrix}
e_{11} \\
e_{21} \\
e_{32} \\
e_{42} \\
e_{53} \\
e_{63}
\end{bmatrix} \tag{6}
\]

In this model, $\eta_{1j}$ represents the latent deviation of the outcome variable for classroom $j$ (i.e., the intercept, as in the prior model), while $\eta_{2j}$ represents the latent relation for classroom $j$ of the predictor, $X_{ij}$, with the outcome, $Y_{ij}$ (i.e., it is a classroom-specific regression weight or slope). We can see that the cross-level linking matrix has a $2 \times 2$ block for each classroom: one row for each student and one column for each of the two latent variables per classroom. This linking matrix maps to the single column of $X_{ij}$ in the dataset (Mehta & Neale, 2005).

Representing variability in such a regression can be accomplished through the notion of a definition variable (Mehta & Neale, 2005; Mehta & West, 2000). That is, if we have a student level predictor, $X_{ij}$, centered at its grand mean, we can represent it in the multilevel SEM by a diamond:
In Figure 6, we have updated the model matrices with superscripts to explicitly show the level to which they belong, following xxM convention. We have chosen to name this model for passage comprehension with a random slope for word attack, “PCranWA.” At the student level (level 1), matrix $\Theta^{1,1}$ has the student level residual variance, as before.

Moving one matrix upward in the diagram, the cross-level linking matrix is now explicitly represented with a superscript of 1,2, indicating that it links parent level 2 (“teacher”) to child level 1 (“student”). The first column of matrix $\lambda^{1,2}$ is fixed to unity as before (Equation 5, Figure 3). The second column of matrix $\lambda^{1,2}$ is fixed to each individual’s specific value of the predictor, $X_{ij}$. Thus, there is a complete and accurate correspondence between the equation representation of the model and the SEM diagram, even when there is a random slope.
Additionally, this representation reinforces the notion of factor loadings for latent variables, even when those loadings vary across person and if they occur across levels.

3.1. PROC MIXED estimation of the random slope model

We extend the previous example to consider how text decoding relates to passage comprehension. Text decoding, or the mapping of letters to the sounds and words they represent, is a crucial basic skill for efficient reading (National Institute of Child Health & Human Development, 2000; Scarborough, 2001). The word attack subtest of the Woodcock Language Proficiency Battery—Revised (Woodcock, 1991) is a widely used test of phonetic decoding in which students try to pronounce nonsense words. While it would be fairly uninformative to find that decoding has a positive relation to passage comprehension (we already know this from theory and decades of prior research), a more interesting question would be, “to what extent does the relation between decoding and comprehension differ across classrooms?” We therefore fit a model of passage comprehension predicted by word attack (centered at its grand mean), with a random effect for word attack at the classroom level:

<table>
<thead>
<tr>
<th>Code</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>proc mixed data=PCWA method=ML;</td>
<td>teacher ID as classification variable</td>
</tr>
<tr>
<td>class teacher;</td>
<td>PC predicted by Word Attack</td>
</tr>
<tr>
<td>model PC = WA / solution;</td>
<td>intercept &amp; WA random at teacher level</td>
</tr>
<tr>
<td>random intercept WA / subject= teacher</td>
<td>effects are correlated UNstructured</td>
</tr>
<tr>
<td>type=un</td>
<td>display random effects matrix and correlation</td>
</tr>
<tr>
<td>g gcorr;</td>
<td>run;</td>
</tr>
</tbody>
</table>

5 As in standard regression, centering can be important when using predictors. Because the intercept is defined where all predictors are zero, it can extend far beyond the observed data (in the current dataset, a W-score of zero is not possible). Therefore, subtracting the sample mean from a predictor, centers the predictor values at their mean and coefficients can be interpreted more conveniently. Broadly speaking, variables can be centered at their grand mean (as is done in the current chapter) or can be centered within clusters (i.e., classrooms). Centering in multilevel models is a critical issue (Kreft, de Leeuw, & Aiken, 1995; Mehta & West, 2000).
The model line now has the grand-mean centered word attack variable, $\bar{WA}$, which will produce the latent mean for word attack (the intercept is included by default). The random line also has $\bar{WA}$, indicating that its effect varies across classrooms ($subject = teacher$). Because we wish to estimate variances for the intercept and slope, as well as freely estimated covariance between them, the covariance is specified as “un” for unstructured (for other covariance structures, see Littell, et al., 2006). In addition, the random effects can optionally be requested to be given in matrix form ($g$), as well as a correlation matrix ($gcorr$).

The covariance parameters appear in somewhat cryptic format:

<table>
<thead>
<tr>
<th>Covariance Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cov Parm</td>
</tr>
<tr>
<td>UN(1,1)</td>
</tr>
<tr>
<td>UN(2,1)</td>
</tr>
<tr>
<td>UN(2,2)</td>
</tr>
<tr>
<td>Residual</td>
</tr>
</tbody>
</table>

These results are only interpretable when we recall that we specified the random effects of intercept first and slope second. The $2 \times 2$ covariance matrix (see Figure 6) appears in the output in list format. Thus, the estimate of the intercept variance ($\psi_{11}$) is covariance element UN(1,1) = 37.0984. The variance of the slope ($\psi_{22}$) is element UN(2,2) = 0.04237 and the covariance between them ($\psi_{21}$) = UN(2,1) = -0.3402. The student level residual variance $\theta = 234.56$. Because we have formulated the covariances as a matrix, this numbering of rows and columns is familiar.

Fortunately, we also requested the classroom random effect in matrix form (“g” option), which has more informative labels and the same estimates as above, but in their $2 \times 2$ matrix ($\psi$) format:
The standardized form of this matrix is given by the option “gcorr” which shows that the classroom intercepts correlate -0.27 with classroom slopes (MIXED is displaying the correlation matrix for teacher ID number 1048, but this matrix is common across all teachers). This suggests a tendency for classrooms with higher model-implied averages of passage comprehension to have a slightly lower relation between word attack and passage comprehension.

The fixed effect—latent means—of the intercept for passage comprehension ($\alpha_1$) was estimated to be 443.39. The mean slope for the relation between word attack and passage comprehension was $\alpha_2 = 0.8501$, indicating that for every unit above or below the mean of WA (pre-centered at zero), there is a unit change in passage comprehension.

Together, these results imply that in a completely average classroom, the model-implied mean passage comprehension for students with completely average word attack skills would be
443 W-score units, with a regression coefficient of 0.85 W-score units. The square roots of the variance components yield standard deviations, which can give a sense of the variation around these averages.

### 3.2. Mplus estimation of the random slope model.

This random slope model can also be fit in Mplus with code fairly similar to that before.

```plaintext
Title: Model 2 Random Slope for WA;
Data: File is PCWA.txt;
Variable: Names = teacher student PC WA;
Cluster = teacher;
Usevariables = PC WA;
Analysis: Type = Twolevel random;
Estimator = ML;
Model:
  %Within%
  Slope | PC on WA;
  %Between%
  PC Slope;
  PC with Slope;
  [PC Slope];
```

The only new feature here is that the `Analysis` is specified as `RANDOM` and the vertical bar operator “|” notes that the regression of PC on WA has a random effect at the classroom level to be called “Slope” (we can choose to name it anything). The results for the random slope model are essentially identical to those from PROC MIXED:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>S.E.</th>
<th>Est./S.E.</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within Level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual Variances</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td>234.536</td>
<td>12.851</td>
<td>18.250</td>
<td>0.000</td>
</tr>
<tr>
<td>Between Level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC WITH SLOPE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>-0.341</td>
<td>0.342</td>
<td>-0.997</td>
<td>0.319</td>
</tr>
<tr>
<td>Means</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td>443.382</td>
<td>0.905</td>
<td>490.124</td>
<td>0.000</td>
</tr>
<tr>
<td>SLOPE</td>
<td>0.850</td>
<td>0.044</td>
<td>19.371</td>
<td>0.000</td>
</tr>
<tr>
<td>Variances</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td>37.112</td>
<td>10.959</td>
<td>3.386</td>
<td>0.001</td>
</tr>
<tr>
<td>SLOPE</td>
<td>0.042</td>
<td>0.023</td>
<td>1.853</td>
<td>0.064</td>
</tr>
</tbody>
</table>
3.3. **xxM estimation of the random slope model**

As noted before, the three steps to fitting a model in xxM are setup, matrix specification, and estimation. Each of these steps will be presented in turn.

### 3.3.1. **xxM setup for the random slope model**

The setup to this model of passage comprehension with a random slope for word attack by naming the model “PCranWA” (any name will do):

```r
PCranWA <- xxmModel(levels = c("student", "teacher"))
```

Next, we would specify the content of each of the submodels for student and teacher.

```r
PCranWA <- xxmSubmodel(model = PCranWA, level = "student", parents = "teacher", ys = c("PC"), xs = c("WA"), latent = , data = student)
```

In the student Submodel, as before, there is the single outcome, PC, with no latent variables, but we have added the predictor, WA.

```r
PCranWA <- xxmSubmodel(model = PCranWA, level = "teacher", parents = "teacher", ys = , xs = , latent = c("int", "slp"), data = teacher)
```

In the classroom Submodel, we have added two latent variables, int and slp, for the intercept and slope, respectively. The crucial difference is in how these latent variables get defined by the cross-level linking matrix, defined in the next section.

### 3.3.2. **xxM matrix specification for the random slope model**

The four matrices shown in Figure 6 ($\alpha$, $\psi$, $\lambda$, $\theta$) are specified in the following way:
The crucial new concept here is fixing the values of the cross-level loading matrix, $\lambda^{1,2}$, to student-specific values of WA. The `lambda_pattern` matrix sets the loadings for intercept and slope to be fixed (zero indicates they are not estimated). The `lambda_value` matrix sets the loadings for intercept to be fixed at unity and the value for slope at zero, but this value will be substituted for the label of the variable next. The `lambda_label` matrix sets names for the loadings: we use an arbitrary name, “lll” to represent $\lambda_{11}$, but the actual name of the predictor, preceded by its SubModel name “student.WA” for the slope. Thus, the label matrix gives a name, “lll” to the fixed value of one for the intercept, and names the variable “student.WA” to take the fixed value of the linking values for the slopes (see Figure 6). Readers interested in a more detailed account of definition variables can find it in Mehta and Neale (2005).

Definition variables, while not new in SEM, require some explanation. While random effects for an observed variable (like word attack) are natural in mixed-effects regression, they can be easily accommodated in SEM (Mehta, 2001; Mehta & Neale, 2005; Neale, 2000). In the same way that we have shown that the cross-level linking matrix captures random variation at the classroom level, the effect of a predictor on the outcome may vary across classrooms. While

<table>
<thead>
<tr>
<th>Code</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha_pattern &lt;- matrix(c(1,1), 2,1)</td>
<td>freely estimated 2x1 start values</td>
</tr>
<tr>
<td>alpha_value &lt;- matrix(c(400,0), 2,1)</td>
<td>start values</td>
</tr>
<tr>
<td>psi_pattern &lt;- matrix(c(1,1,1,1), 2,2)</td>
<td>freely estimated 2x2 start values (from prior output)</td>
</tr>
<tr>
<td>psi_value &lt;- matrix(c(37,-.34,-.34,.04), 2,2)</td>
<td></td>
</tr>
<tr>
<td>lambda_pattern &lt;- matrix(c(0,0), 1,2)</td>
<td>fixed 1 x 2</td>
</tr>
<tr>
<td>lambda_value &lt;- matrix(c(1,0), 1,2)</td>
<td>value of one, other not specified</td>
</tr>
<tr>
<td>lambda_label &lt;- matrix(c(&quot;lll&quot;,&quot;student.WA&quot;), 1,2)</td>
<td>definition variable</td>
</tr>
<tr>
<td>theta_pattern &lt;- matrix(1, 1,1)</td>
<td>estimated variance 1x1 start value</td>
</tr>
<tr>
<td>theta_value &lt;- matrix(234, 1,1)</td>
<td></td>
</tr>
</tbody>
</table>
the cross-level linking matrix for the intercept allocates students within classrooms via a one-to-one correspondence for their physical location, the relation of predictor to outcome may vary across classrooms. Specifically, there may be a specific covariance for each value of the predictor (Mehta & Neale, 2005; Neale, 2000). In this respect, the role of the predictor can be viewed as a factor loading for the random latent variable for this relation. This is why the notion of a definition variable is important: the diagram is mathematically complete and consistent with how random effects function in the model. Now, for conceptual simplicity, random slopes can be incorporated into SEM, taking full advantage of the communicative strengths of SEM diagrams.

Next, we add all of these matrices to their respective model matrices.

<table>
<thead>
<tr>
<th>Code</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCranWA &lt;- xxmWithinMatrix(model = PCranWA, level = &quot;teacher&quot;, matrixType = &quot;alpha&quot;, pattern = alpha_pattern, value = alpha_value, )</td>
<td>classroom level latent means (alpha)</td>
</tr>
<tr>
<td>PCranWA &lt;- xxmWithinMatrix(model = PCranWA, level = &quot;teacher&quot;, matrixType = &quot;psi&quot;, pattern = psi_pattern, value = psi_value, )</td>
<td>classroom level covariance matrix (psi)</td>
</tr>
<tr>
<td>PCranWA &lt;- xxmWithinMatrix(model = PCranWA, level = &quot;student&quot;, matrixType = &quot;theta&quot;, pattern = theta_pattern, value = theta_value, )</td>
<td>student level residual variance (theta)</td>
</tr>
<tr>
<td>PCranWA &lt;- xxmBetweenMatrix(model = PCranWA, parent = &quot;teacher&quot;, child = &quot;student&quot;, type = &quot;lambda&quot;, pattern = lambda_pattern, value = lambda_value, label = lambda_label, )</td>
<td>cross-level linking matrix with definition variables specified in the &quot;label&quot; option (&quot;name&quot; option still not used)</td>
</tr>
</tbody>
</table>

In this specification, the WithinMatrix lines still do not use the "label" and "name" features, so those are left empty at the end of each list of parameters. However, the cross-level linking matrix fixes slope loadings to values of WA, so that variable is named as a label variable. While not used in the current examples, the "name" parameter can be invoked in order to force some parameters to be equal, as in multiple group modeling in which we would desire invariance tests of particular parameters (e.g., loadings or intercepts; see Meredith, 1993).
3.3.3. xxM estimation of the random slope model

The random slope model is run, as before with the xxmRun command:

```r
PCranWA <- xxmRun(model = PCranWA)
```

The xxM output results for the random slope model are:

<table>
<thead>
<tr>
<th>xxM Results</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>teacher_student_lambda</td>
<td>Cross-level linking matrix</td>
</tr>
<tr>
<td>labels</td>
<td>names of the matrix elements</td>
</tr>
<tr>
<td>colnames</td>
<td>slope renamed “student_WA”</td>
</tr>
<tr>
<td>rownames int slp</td>
<td>PC “l11” “student_WA”</td>
</tr>
<tr>
<td>teacher_alpha</td>
<td>Latent classroom means</td>
</tr>
<tr>
<td>colnames</td>
<td></td>
</tr>
<tr>
<td>rownames int slp</td>
<td>Mean intercept ( \alpha_1 )</td>
</tr>
<tr>
<td></td>
<td>Mean slope ( \alpha_2 )</td>
</tr>
<tr>
<td>student_theta</td>
<td>Student level (residual) covariance</td>
</tr>
<tr>
<td>[ ,1 ]</td>
<td></td>
</tr>
<tr>
<td>[1,] 234.5195</td>
<td>Residual variance, ( \theta )</td>
</tr>
<tr>
<td>teacher_psi</td>
<td>Latent classroom covariance</td>
</tr>
<tr>
<td>[ ,1 ]</td>
<td>( \psi ) matrix</td>
</tr>
<tr>
<td>[1,] 37.102141 -0.33879696</td>
<td>( \psi_{11} \ \psi_{12} )</td>
</tr>
<tr>
<td>[2,] -0.338797 0.04235558</td>
<td>( \psi_{21} \ \psi_{22} )</td>
</tr>
</tbody>
</table>

3.4. Extending the random slopes model to three levels

With a respectable number of schools (perhaps 100 or more—see sample size discussion below), we might consider fitting a three level model of students within classrooms within schools. Such a model seeks to examine variance and covariance at this higher level, and conditional variance and covariance at the classroom level. While the current sample only contains 23 schools, we present the specifications (but not estimated results) in order to demonstrate the extension to higher levels.
In PROC MIXED, extending the previous random slope model to three levels only requires using the school ID variable in the following way:

<table>
<thead>
<tr>
<th>Code</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>proc mixed data=PCWA method=ML;</td>
<td>level IDs as classification variables</td>
</tr>
<tr>
<td>class teacher school;</td>
<td>PC predicted by Word Attack</td>
</tr>
<tr>
<td>model PC = WA / solution;</td>
<td>intercept &amp; WA random at school level</td>
</tr>
<tr>
<td>random intercept WA / subject= school type=un g gcorr;</td>
<td>effects are correlated UNstructured</td>
</tr>
<tr>
<td>random intercept WA / subject= teacher(school) type=un g gcorr;</td>
<td>random teacher effects nested within schools</td>
</tr>
<tr>
<td>run;</td>
<td>effects are correlated UNstructured</td>
</tr>
</tbody>
</table>

There are only two small changes from the previous two-level random slopes model. First, there is a random statement for the school level (identified as a class variable). Second, the classroom level random statement notes that teachers are nested within schools: subject= teacher(school). Again, alternative covariance structures can be tried, such as dropping the covariance between the intercept and slope (i.e., a variance components model), or dropping the random slope at the third level, since such effects can be difficult to estimate with small numbers of topmost clusters.

Currently, Mplus 6.1 cannot fit three level models. While a multivariate layout such as that shown in Figure 4 could be specified, such a restructuring of the data would be highly inefficient for models of students within classrooms within schools.

In xxM, the extension to a third level is straightforward. Here, we extend the prior notation of Figure 6 to more explicitly show the levels to which each parameter belongs. Figure 7 shows the SEM diagram and model matrices for the three level random slopes model. Specification of these in code would follow from the previous examples, and is not shown here.
In Figure 7, we can define a model to be called “PC3” with three submodels: student, teacher, and school. At the student level (level 1), as before, there is a single residual variance, $\theta^{1,1}$. The cross-level linking matrix to level 1 from level 2, $\Lambda^{1,2}$, is the same as before (Figure 6).

**Figure 7: Diagram and matrices for the three-level random slope model.**

The teacher submodel (level 2) contains the latent covariance matrix, $\psi^{2,2}$, much the same as before, except that it will now be interpreted as a residual covariance matrix, conditional on the school level random effects. These school level random effects at level 3 have a straightforward substantive interpretation: they are the variance of the intercept across schools ($\psi_1^{3,3}$), the variance of the effect of word attack ($\psi_2^{3,3}$), and the covariance between them ($\psi_1^{3,3} \psi_2^{3,3}$).
These are the random effects produced by the two separate random statements in PROC MIXED before.

Finally, at the top of Figure 7, the means of the intercept and slope factors are given: $\alpha_{1}^{3,3}$ and $\alpha_{2}^{3,3}$. The grand mean of the intercept, $\alpha_{1}^{3,3}$, is the model-predicted level of performance on the outcome (passage comprehension) where the predictor is completely average (grand mean centered in this case). The grand mean of the slope factor, $\alpha_{2}^{3,3}$, is the model-predicted average relation of word attack ($X_{ij}$) to passage comprehension ($Y_{ij}$). Thus, schools might differ from this average relation (reflected in their variance, $\psi_{22}^{3,3}$), and classrooms might also differ in this relation within schools (reflected in the classroom variance, $\psi_{22}^{2,2}$). Specification of the model shown in Figure 7 in xxM would follow the outline of code presented previously.

This three level model raises interesting substantive possibilities. For example, if the covariance between intercept and slope is opposite in direction, this would imply that schools function in very different ways than do classrooms. If high performing classrooms tend to have a positive relation between word attack and passage comprehension, but schools have the opposite relation, then perhaps there are mitigating forces at work at the school level. Alternatively (and possibly more typically), these relations may be in the same direction, suggesting similar support. Considering such relations at separate levels is common in multilevel models, where including the mean of a predictor as a separate, higher level effect can indicate context differences in how a predictor affects an outcome. Such substantive separation between levels is crucial to unpacking ecological differences (Burstein, 1980; Cronbach, 1976; Robinson, 1950).

More importantly, with the explicit SEM layout, we can see how additional observed variables could be added to make measured latent factors at the classroom or school levels (as in multilevel SEM: see Mehta, this volume).
4. **Comparing across the specifications**

For a single outcome, the xxM specification is cumbersome to be sure. However, this step by step illustration helps to reinforce several key concepts which are common across multilevel modeling and SEM, and form a foundation for SEM with arbitrary numbers of levels and potentially complicated nesting.

Strengths of SAS PROC MIXED include the simplicity of working from a single dataset and a number of ready-made covariance structures (e.g., compound symmetric, autoregressive; for extensive examples, see Littell, et al., 2006). In addition, cross-classified random effects (e.g., the nesting of students in schools and neighborhoods) are also easily specified. This flexibility, however, can yield code which can become cryptic for complex models. The generality of PROC MIXED can also come at a price, in that large models become computationally burdensome (however, see PROC HPMIXED for large variance-component models). Several practical examples of using PROC MIXED not covered here can be found in Singer (1998) and in West, Welch, & Galeycki (2007).

Mplus allows specification of observed and latent variable models at two levels, as well as random effects as in mixed effects models. With appropriate structure imposed, three-level models can be specified as two-level multivariate models (Bollen & Curran, 2006; Mehta & Neale, 2005; Mehta & West, 2000). The specification is exceedingly simple with many helpful default features for the novice, but nesting must be strictly hierarchical and limited to two levels (as of the time of this writing in Mplus 6.12). Large models can become computationally burdensome, but alternate estimators are also available. Full translations between Mplus and PROC MIXED are given by Mehta and Neale (2005).
By comparison, xxM requires explicit matrix specification for the role of variables, many of which are handled automatically in other software. In exchange, however, the number of levels is not limited by design and each level can be specified separately as its own SEM, with observed and latent variables. Cross-level linkages, which are essentially automatic in PROC MIXED and Mplus, must be specified explicitly in xxM. This notion of cross-level links as explicit factor loadings makes clear the isomorphism between latent variables in SEM and random effects in multilevel regression. Additionally, there is a one-to-one correspondence between the equations, diagram, and script.

Regardless of software or representation, a primary goal of scientific research is clarity of representation such that readers can independently evaluate the findings and methods. Because of the diversity of representations, this chapter has emphasized the connections between equations, diagrams, and software script. Beginners may also find useful a very concise overview of reporting issues for multilevel analyses (McCoach, 2010). Ferron and colleagues (2008) provide a checklist of general and applied questions for evaluating studies and reporting results of multilevel models.

5. Other Software

It is worth noting that there are other software programs for fitting multilevel models, such as lme4 in R, HLM, SuperMix, and MLwiN. For example, in HLM, if the user can specify the equations, then variables can be easily added to the model with a drag-and-drop interface. Each of these programs has special strengths which differentiate it from other programs and may make it particularly useful for some applications. The intention of this chapter is not to neglect these other programs (see reviews by Goldstein, 2003; Roberts & McLeod, 2008)(Fielding, 2006; West, 2007), but software reviews become outdated nearly as fast as they are printed. For
detailed applications of multilevel models in several different software packages, including SAS, SPSS, R, Stata, and HLM, see West, Welch, and Gałecki (2007). The Centre for Multilevel Modelling maintains reviews of multilevel software (http://www.bristol.ac.uk/cmm/learning/mmsoftware/). Readers interested in fitting models in other software and providing comparison code with xxM are invited to participate in the forum section of the xxM website: http://xxm.times.uh.edu.

6. Sample size considerations and power

With the potential to fit a SEM at any of multiple levels, power becomes limited by the sample sizes at each of those levels. The limitation usually becomes most severe at the highest level of nesting. While estimating means can be done with relatively few units, estimating variances and covariances, or fitting latent variable models requires larger numbers of observations. There are few rules of thumb which can apply to every situation, especially when SEMs become complex (see discussion by Lomax, this volume). One advantage to the xxM framework is that the number of parameters must be explicitly specified at each level. Thus, when the number of parameters relative to the number of observations at a particular level becomes large, users can expect low precision and estimation trouble. Oftentimes, the complexity of models at higher levels must necessarily be limited by the smaller number of clusters.

In a simulation study varying ICC, number of clusters, and number of observations per cluster, Maas and Hox (2005) found no bias in the the estimation of fixed or random effects, even for small samples. However, the precision (i.e., standard errors) of the second level variance components were inappropriately small for samples with fewer than 100 clusters. When studies focus on the precise interpretation of second level variances (and covariances), the
quality of inferences will depend on having enough clusters to accurately estimate the covariance structure at that level—reinforcing the idea that the second level is also a SEM. In this respect, the precision of the model is most severely limited by the topmost level, with recommendations for multilevel SEM suggesting that 100 clusters is a reasonable minimum (Hox & Maas, 2001).

7. Conclusion

Practical introductions to mixed effects models (e.g., Singer, 1998) and full translations between multilevel models and SEM (Mehta & Neale, 2005) are growing and can be found in many sources. The current chapter used two simple examples to highlight the similarities between the models as fit from a random effects framework versus a SEM framework. While the code from PROC MIXED and Mplus bear little resemblance to each other, the code from xxM makes explicit the connections between script, diagram, and model. Consequently, the xxM approach lays bare the underlying equivalence in these approaches. More importantly, xxM highlights the conceptual necessity of thinking about cross-level links and opens the possibility of fitting latent variable models not only at arbitrary levels, but across levels in substantively interesting ways. That is, cross-level random effects are latent variables which may be subject to direct scrutiny.

The concepts emphasized in this chapter are that random effects are latent variables at higher levels, linked by cross-level matrices. While most multilevel software programs handle these linkages automatically, the explicit link matrices required by xxM can be specified for essentially any arbitrary configuration of nesting, including multiple membership (such as students served by multiple tutors), switching classification (as when students change teachers across grades), and even subject-specific values of links (such as attendance or individually-varying times of observations)—the random slopes as shown in this chapter.
References


Thorndike, E. L. (1939). On the fallacy of imputing the correlations found for groups to the individuals or smaller groups composing them. The American Journal of Psychology, 52(1), 122-124.


Appendix A: Resources for learning R

The canonical reference for R data structures and manipulation:
http://cran.r-project.org/doc/manuals/R-intro.html

A jump-start with exercises:

Quick-R has short examples for users familiar with SAS or SPSS:
http://www.statmethods.net/