Paper 433-2013

A Multilevel Model Primer Using SAS® PROC MIXED

Bethany A. Bell, Mihaela Ene, Whitney Smiley, Jason A. Schoeneberger
University of South Carolina

ABSTRACT

This paper provides an introduction to specifying multilevel models using PROC MIXED. After a brief introduction to the field of multilevel modeling, users are provided with concrete examples of how PROC MIXED can be used to estimate (a) two-level organizational models, (b) two-level growth models, and (c) three-level organizational models. Both random intercept and random intercept and slope models are illustrated. Examples are shown using different real world data sources, including the publically available Early Childhood Longitudinal Study–Kindergarten cohort data. For each example, different research questions are examined through both narrative explanations and examples of the PROC MIXED code and corresponding output.

Keywords: MULTILEVEL MODELING, PROC MIXED, MODEL FIT, STATISTICAL ASSUMPTIONS

INTRODUCTION

Hierarchically organized data are commonplace in educational, clinical, and other research settings. For example, students are nested within classrooms or teachers, and teachers are nested within schools. Alternatively, service recipients are nested within social workers providing services, who may in turn be nested within local civil service entities. Conducting research at any of these levels while ignoring the more detailed levels (students) or contextual levels (schools) can lead to erroneous conclusions. As such, multilevel models have been developed to properly account for the hierarchical (correlated) nesting of data (Heck & Thomas, 2000; Hox, 2010; Klein & Kozlowski, 2000; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999).

Research has shown that ignoring a level of nesting in data can impact estimated variances and the available power to detect treatment or covariate effects (Donner & Klar, 2000; Julian, 2001; Moerbeek, 2004; Murray, 1998; Shadish, Cook & Campbell, 2002), can seriously inflate Type I error rates (Wampold & Serlin, 2000), and can lead to substantive errors in interpreting the results of statistical significance tests (Goldstein, 2003; Nich & Caroll, 1997). There are many types of multilevel models, which differ in terms of the number of levels [e.g., students nested within schools (2-level), students nested within schools nested within districts (3-level)], type of design (e.g., cross-sectional, longitudinal with repeated measures, cross-classified), scale of the outcome variable (e.g., continuous, categorical), and number of outcomes (e.g., univariate, multivariate). These models have been used to address a variety of research questions involving different model parameters. For example, if a researcher was interested in how socioeconomic status (SES) is related to reading achievement in students nested within schools, the researcher may be interested in answering questions regarding fixed effects, such as understanding the relationship between SES and reading achievement across schools (i.e. slope across all schools). Questions about random level-1 coefficients can also be answered using multilevel models, such as the relationship between SES and reading achievement at a particular school (i.e. slope for a particular unit). Lastly, researchers can answer questions about variance-covariance components, such as the amount of variation in the relationship between student SES and reading achievement across schools (i.e. slope variance).

Multilevel models (MLMs) can be conceptualized as regression models occurring at different levels. For example, a researcher might be interested in modeling the reading achievement of individual students (at level-1) and exploring the influence of characteristics associated with the school those students attend (at level-2). Alternatively, a third-level could be added by using multiple measurements of reading achievement across time as the level-1 unit, nested within individual students (level-2) nested within schools (level-3). For simplicity, we proceed using the 2-level model where students are nested within schools. The equations necessary for estimating the multilevel model are presented below.

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij} \quad \text{(Eq. 1)}$$

Equation 1 represents a simple level-1 model with one student-level predictor where Y_{ij} is the reading achievement for student i in school j, β_{0j} is the average reading achievement for school j, X_{ij} is a student-level predictor for

student i in school j and β_{1j} is the slope or regression coefficient associated with X_{ij}, thus this value depicts the relationship between the student-level variable and reading achievement. The last term in the equation, e_{ij} is the student-level error term, which is assumed to be normally distributed with covariance R.

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + \mu_{0j} \qquad \text{(Eq. 2a)}$$
$$\beta_{1j} = \gamma_{10}$$

Equation 2a is for the simple level-2 model with one school-level predictor where γ_{00} is the intercept, which represents the grand mean of reading achievement across students and across schools, W_j is a school-level predictor for school j and γ_{01} is the regression coefficient associated with W_j, μ_{0j} is an error term representing a unique effect associated with school j, and γ_{10} estimates the average effect of the student-level predictor. The absence of an error term in the equation for β_{1j} indicates that the effect of the student-level predictor is fixed, or held constant across schools. The level-2 errors are assumed to be normally distributed with covariance G. This type of two-level model is most commonly referred to as a "random intercept" model.

The effect of the student-level predictor can also be modeled to vary across level-2 units. The equation that includes a random effect for the student-level predictor (X_{ij}), is shown below in Equation 2b.

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + \mu_{0j} \qquad \text{(Eq. 2b)}$$
$$\beta_{1j} = \gamma_{10} + \mu_{1j}$$

Note the only difference between Equation 2a and 2b is the inclusion of μ_{1j} in the equation for β_{1j}. By including μ_{1j} in the level-2 equation, we are allowing the relationship between the student-level predictor (X_{ij}) and the outcome (Y_{ij}) to vary across level-2 units. This type of two-level model is most commonly referred to as a "random intercept and random slope" model.

Next, by substituting the values of β_{0j} and β_{1j} from the level-2 equation, shown in Equation 2a, into the level-1 equation (Equation 1), the combined level-1 and level-2 random intercept model is created (Equation 3a). From this combined model, the regression element behind multilevel models becomes more apparent. Specifically, as shown in the two-level model of Equation 3a, there is the continuous reading outcome (Y_{ij}), an intercept (γ_{00}), level-1 and level-2 regression coefficients (γ_{10} and γ_{01}, respectively), and level-1 and level-2 error terms (e_{ij} and μ_{0j}, respectively).

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10} X_{ij} + \mu_{0j} + e_{ij} \qquad \text{(Eq. 3a)}$$

Following the same logic, when substituting the values β_{0j} and β_{1j} from the level-2 equation, shown in Equation 2b, into the level-1 equation (Equation 1), the combined level-1 and level-2 random intercept and slope model is created (Equation 3b).

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10} X_{ij} + \mu_{0j} + \mu_{1j} X_{ij} + e_{ij} \qquad \text{(Eq. 3b)}$$

In addition to understanding the equations used to represent multilevel models, it is also important to understand the language that is used when estimating these models. For a comprehensive list of key terminology used with

multilevel model analyses, we recommend Diez Roux (2002). For this paper, we will define key terms as they are introduced, beginning with random effects and fixed effects. Random effects, which are estimated as variance components, are model parameters that are estimated to vary between higher level units whereas fixed effects are estimates that are modeled to not vary between higher level units. At minimum, a 2-level linear model estimated in PROC MIXED will include one random effect --this is the key distinction between 2-level linear models and single-level ordinary least squares (OLS) models.

For example, in Equation 3a above, if the random error term μ_{0j}, which is the part of the model that represents the intercept variance between level-2 units, was removed, we would be left with a traditional OLS equation. It is this error term that allows us to call the model depicted in Equation 3a a "random intercept" model. With Equation 3b, there is still the random error term, but now we also have μ_{1j} which allows the random slope for our level-1 predictor variable X_{ij}. A random slope means that we are allowing the relationship between the student-level predictor (X_{ij}) and the outcome (Y_{ij}) to vary across level-2 units (i.e., we are not forcing the relationship to be the same in each of the level-2 units).   These random effects are key components of multilevel models that differentiate them from single-level models.

Figure 1 provides an illustration of random intercepts and random slopes. Ignoring color, the data points as a whole are represented by the thick, black regression line, which would be typical of OLS. The blue data points represent performance of students in school A, red in B and green in C. Equation 3b allows us to specify a unique group effect through random intercepts and random slopes. This yields an estimate of the overall intercept and the overall slope across schools, as well as the deviation from the overall intercept and slope for each school. Thus, each school has their own regression line – in this example, both the intercepts and slopes are varying by school. School A has, on average, lower overall performance on the outcome, but the slope associated with the predictor is stronger in this school than any other. School C, on the other hand, has the highest overall performance on the outcome, but the slope is the weakest.
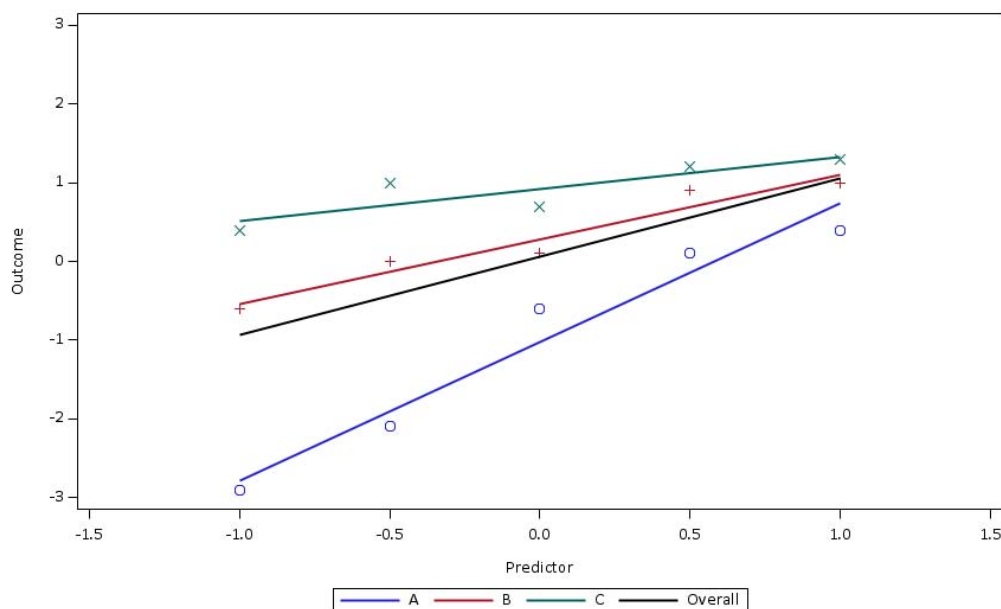
Figure 1. Conceptual illustration of random intercepts and slopes

Whereas MLMs are different than OLS regression models, the fixed effects in MLMs (i.e., in Equation 3b, γ_{00} , γ_{01} , and W_j) are analogous to unstandardized regression coefficients in a single-level OLS model. More specifically, the three fixed effects from Equation 3b represent the (a) model intercept (γ_{00}), which represents the grand mean of

reading achievement across students and across schools; (b) fixed effect for X_{ij} (γ_{10}), which represents the relationship between the level-1 variable reading achievement while controlling for all other variables in the model; and (c) fixed effect for W_j (γ_{01}) which represents the relationship between the level-2 variable and reading achievement while controlling for all other variables in the model.

MODEL BUILDING PROCESS

Although different sources provide different guidelines on the model building process when estimating multilevel models, they all have the same goal -- for researchers to estimate the most parsimonious models that best fit their data. Other common elements of the different model building strategies include (a) always starting with an unconditional model (i.e., a model that has no predictors) as this model is used to calculate the intraclass correlation coefficient (ICC) which estimates how much variation in the outcome exists between level-2 units and (b) gradually estimating more complex models while checking for improvement in model fit after each model is estimated. One major difference between approaches relates to the fit statistic/indices that are used to assess model fit; the appropriate fit statistic/index depends on model estimation [full maximum likelihood (ML) vs. restricted maximum likelihood (REML)] as well as whether the models are nested (i.e., models that have been fit using the same data and where one model is a subset of the other) or not.

For simplicity's sake, in this paper, we will estimate all of our models using ML, which means that we can examine improvement in model fit using Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) when the nested models differ in either random or fixed effects. For both of these measures, smaller values represent better fitting models. Another benefit of using AIC and BIC to examine model fit is that these two indices can be used when examining model fit for both nested and non-nested models, although the majority of researchers who use a model building process do so on nested models. Another commonly used measure of model fit is the likelihood ratio test when examining differences in the - 2 log likelihood (-2LL) values of nested models. However, given the many nuances associated with the correct use of this fit statistic, we have chosen to utilize the more versatile measures of AIC and BIC. For readers who want more details on the different approaches and fit indices for model building and examining model fit, we encourage them to consult any of the many multilevel modeling textbooks – most of them provide detailed explanations of these processes and approaches.

The model building process that we will use in the examples provided below is a relatively simple and straightforward approach to obtain the best fitting and most parsimonious model for a given set of data and specific research questions (see Tables 1 and 2). In addition to listing what effects are included in each of the models, we have also included information about what the output from the various models provides.   It is important to note that the models listed in Tables 1 and 2 are general in nature. Sometimes intermediate models are also examined depending on the model fit indices (i.e., if Model 3 did not seem to be a better fitting model than Model 2, and only 1 of 3 random slopes was significant, a reduced version of Model 3 could be estimated only including the significant random slope from the original Model 3).

Model 1	Model 2	Model 3	Model 4
No predictors, just random effect for the intercept	Model 1 + level-1 fixed effects	Model 2 + random slopes for level-1 predictors	Model 3 + level-2 fixed effects
Output used to calculate ICC – provides information on of how much variation in the outcome exists between level-2 units	Results indicate the relationship between level-1 predictors and the outcome	Fixed effect results provide the same information as Model 2; random slope results reveal if the relationships between level-1 predictors and the outcome vary between level-2 units	Level-2 fixed effect results indicate the relationship between level-2 predictors and the outcome. Rest of the results provide the same information as listed for Model 3

Table 1. Model building process for 2-level linear models

Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
No predictors, just random effect for the intercept	Model 1 + level-1 fixed effects	Model 2 + random slopes for level-1 predictors	Model 3 + level-2 fixed effects	Model 4 + random slopes for level-2 predictors	Model 5 + level-3 fixed effects
Output used to calculate ICC – provides information on of how much variation in the outcome exists between level-2 and level-3 units	Results indicate the relationship between level-1 predictors and the outcome	Fixed effect results provide the same information as Model 2; random slope results reveal if the relationships between level-1 predictors and the outcome vary between level-2 units and between level-3 units	Level-2 fixed effect results indicate the relationship between level-2 predictors and the outcome. Rest of the results provide the same information as listed for Model 3	Results from Model 4 still apply. Random slope results of level-2 variables reveal if the relationships between level-2 predictors and the outcome vary between level-3 units	Level-3 fixed effect results indicate the relationship between level-3 predictors and the outcome. All the other results from level-1 and level-2 provide the same type of information as previously stated.

Table 2. Model building process for 3-level linear models

As you progress through the suggested model building process outlined in Tables 1 and 2, you also examine model fit as measured by AIC and BIC. As the models become more complex and the model fit improves, then AIC and BIC values for the more complex model will be smaller than the values for the comparison model. For example, if in a 2-level model, allowing the level-1 predictors to vary between level-2 units improves model fit, then the AIC and BIC values for Model 3 will be smaller than the values from Model 2. However, the decrease in fit values necessary to declare a model as better-fitting is relatively subjective. Raftery (1995 as cited in O'Connell & McCoach, 2008) does offer some basic rules of thumb for changes in BIC -- differences of 0-2 provide weak evidence favoring the more complex model; differences of 2-6 provide positive evidence for favoring the more complex model, differences of 6-10 provide strong evidence; and differences above 10 provide very strong evidence favoring the more complex model. Again, these are general guidelines. The ultimate determination is left to the researcher.

PROC MIXED STATEMENTS AND OPTIONS

Before explaining the PROC MIXED statements and options, two data management issues will be discussed – centering of predictor variables and data file format requirements for organizational and growth models. As with other forms of regression models, the analysis of multilevel data should incorporate the centering of predictor variables to yield the correct interpretation of intercept values and predictors of interest (Enders & Tofighi, 2007; Hofmann & Gavin, 1998). Predictor variables can be group-mean or grand-mean centered, depending on the primary focus of the research question being examined. For a well-written discussion on the pros and cons of the two approaches to centering predictor variables, we refer readers to Enders & Tofighi (2007). For simplicity and because theoretically our primary interests in the examples we use are at the higher levels of the models (i.e., level-2 and/or level-3), unless specified, we have grand-mean centered all of our continuous predictor variables. Likewise, we did not center our binary predictors as they already had a "meaningful zero" value for interpretation of the intercept and other predictors of interest. Regarding the formatting requirements of data files used to conducted multilevel analyses using PROC MIXED, the data file to be used with organizational models needs to be wide (i.e., one row of data for each observation with each variable in the data file contained in separate columns) whereas the data file to be used with growth models needs to be tall or long (i.e., one row of data for each observation for each time point).

Next, we present details about the PROC MIXED statements and options. Like several other modeling procedures in SAS, PROC MIXED includes a class statement as well as a model statement. In addition, PROC MIXED also includes a random statement – this is where users specify which random effects they want to estimate (i.e., random intercept only or random intercept and slopes). More details about each of the statements and their options are provided below.

Below is example code for PROC MIXED to estimate a simple random intercept and slope model.

```
PROC MIXED data=SGF13 covtest noclprint method = ML;
class L2_id;
model y=x1 w1/solution ddfm = SATTERTHWAITE;
random intercept x1/ sub=L2_id type=vc;
```
On the PROC MIXED statement, in addition to listing the data set to use for the analysis, we have requested covtest, noclprint, and method options; the covtest option is required to generate hypothesis testing output for the variance

and covariance components (aka random effects), the noclprint option is simply requesting SAS to not print the CLASS level identification information, and the method= option is required to tell SAS to estimate the models using maximum likelihood. The default estimation method in PROC MIXED is restricted maximum likelihood so if you want models estimated with REML, there is no need to use the method option.

The CLASS statement is used to indicate the clustering /grouping variable(s); in the example code above, you can see that L2_id is the clustering variable for the analysis. Another way to think about this is that the CLASS statement is where you list the variable in your data that identifies the higher level units.

The MODEL statement is used to specify the fixed effects to be included in the model whereas the RANDOM statement is used to specify the random effects to be estimated. In the example code above, y is the outcome variable and x1 and w1 are the predictors. The solution option on the MODEL statement asks SAS to print the estimates for the fixed effects in the output. The ddfm = option allows the user to specify which degrees of freedom to use when estimating the models. In this example, we have requested the Satterthwaite degrees of freedom be used. There are five degrees of freedom methods available in PROC MIXED – residual, between, containment, Satterthwaite, and Kenward-Roger. A complete discussion of these different methods is beyond the scope of this paper, however, we will mention that both the Satterthwaite and Kenward-Roger DF methods are well-suited for most multilevel models as they both are intended to be used in models with unbalanced designs and both can handle complex covariance structures. Moreover, the Kenward-Roger DF method also adjusts for small sample bias, so, it is ideal to use when estimating models with relatively small samples. For more technical details on the different DF estimation methods, we encourage users to read the SAS documentation for PROC MIXED.

With the intercept and x1 listed on the RANDOM statement, this example code is specifying for both the model intercept and the slope of x1 be allowed to vary between level-2 units. The sub= option on the RANDOM statement corresponds with the CLASS statement. Specifically, in this example sub= L2_id tells SAS to estimate each of the random effects, here the intercept and x1, within in level-2 unit and the resulting variance components are estimates of the variation of the underlying mean responses between level-2 units. The next option type = is used to specify what covariance structure to use when estimating the random effects. Again, although a discussion of the many covariance structure options available in PROC MIXED is beyond the scope of this paper, in the example above, and throughout the examples used in the paper, we use the type = VC option, which is actually the default in PROC MIXED. This is the most simple structure as the covariance matrix for level-2 errors (referred to in SAS as the G-matrix) is estimated to have separate variances but no covariance. The same is used for estimating the covariance matrix for level-1 errors. When specifying random effects, the covariance structures can get quite complicated rather quickly. Thus, it is critical for users to remember to check their log when estimating random intercept and slope models – it is not uncommon to receive a "non-positive definite G-matrix" error. When this happens, the results for the random effects are not valid and the covariance part of the model needs to be simplified.

DATA SOURCE FOR ILLUSTRATED EXAMPLES

For the examples involving two- and three-level organizational models (examples 1 and 3), we used a non-publicly available school district data set where students are nested within classrooms which are nested within schools. This data set has three student-level variables, one teacher-level variable, and two school-level variables. The student variables are a standardized measure of math achievement, English proficiency status, and exceptional child status. The teacher variable is the number of years of teaching experience. The two school variables are participation in a school-level intervention for improving achievement targeted at lower-performing schools (i.e., treatment schools vs. control schools) and the percent of students receiving free-and reduced lunch. As stated above, organizational models need to be in wide format, thus there is only one observation per student.

The two--level growth models (example 2) were estimated using the Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999 (ECLS-K) data set, which is publicly available on NCES' website. The ECLS-K is a study completed by the U.S. Department of Education, National Center for Education Statistics (NCES; www.nces.gov). The focus of this study was to follow a nationally representative sample of children from kindergarten in 1998-1999 through 8th grade in order to understand children's early school experiences. Data were collected at the student, teacher, and school level twice a year. For the purposes of our growth model example, we took a sample of 120 students across the first four waves of data (beginning and end of kindergarten and beginning and end of first grade). Variables used in the examples include the student's biological sex and a measure of reading achievement, where higher numbers equal higher reading achievement. As stated previously, data for growth models need to be in long format so these data consists of multiple observations for each student.

EXAMPLE 1: ESTIMATING TWO-LEVEL ORGANIZATIONAL MODELS

The use of PROC MIXED for two-level organizational models will be demonstrated on an example of hierarchical data consisting of students nested within schools. The primary interest of this study is to investigate the impact of a school-level intervention targeted at lower-performing schools. Therefore, using student data (level-1) and school data (level-2), two-level nested models are used to investigate the relationship between math achievement and school intervention as well as other characteristics at the student and school level. More specifically, the variables used in this study include a standardized math achievement score (ma_z) as our outcome of interest, student level binary variables representing Limited English Proficiency (lep = 1) and Exceptional Child status (i.e., students with disabilities; ec = 1); and school level variables regarding participation in the intervention (treatment = 1) and the percent of students in the school receiving Free or Reduced-Price Lunch (schl_frl). To ensure that our continuous predictor, schl_frl, had a meaningful zero, we grand-mean centered it before estimating the models. All other predictors were binary, thus, they already had a meaningful zero.

The specific research questions examined in this example include:

1. How much of the variance in math achievement scores is attributable to students and schools?
2. Does the influence of any student-level predictor vary among schools?
3. What is the impact of a school-wide intervention program on math achievement while controlling for student and school characteristics?

The model building process started by examining the unconditional model with no predictors (Model 1) to assess between-school variation in student math achievement. This model addresses our first research question regarding the variance in math achievement attributable to students (level-1) and schools (level-2). The PROC MIXED syntax for the two-level unconditional model for this study is shown below. Note that for this example, schoolid is our level-2 identification variable so we use it on the CLASS statement and the sub = option on the RANDOM statement. Ma_z is our criterion variable and we do not have any predictor variables in this model.

```
PROC MIXED data=lib.SGF13 covtest noclprint method = ML;
class schoolid;
model ma_z=/solution ddfm = SATTERTHWAITE;
random intercept / sub=schoolid type=vc;
```

This PROC MIXED syntax generated the Covariance Parameter Estimates table shown below. Using the estimates presented in the table, we can compute the Intraclass Correlation Coefficient (ICC) that indicates how much of the total variation in math achievement is accounted for by the schools.

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	schoolid	0.2623	0.06823	3.84	<.0001
Residual		0.7086	0.007015	101.01	<.0001

The general equation for calculating the ICC is provided below. In this equation, σ^2_{school} refers to the covariance estimate for the intercept and σ^2_{error} refers to the covariance estimate for the residual.

$$ICC = \frac{\sigma^2_{school}}{\sigma^2_{school} + \sigma^2_{error}}$$

Thus, based on the output above we calculated the ICC as

$$ICC = \frac{.2623}{.2623 + .7086} = .2702$$

This indicates that 27% of the variability in student math achievement is accounted for by for the schools in our study, leaving 73% of the variability in math achievement to be accounted by for the students. These results answer our first research question and provide support for using a two-level model. The large proportion of the variability in student math achievement explained by the schools emphasizes the importance of accounting for the hierarchical structure of the data.

PROC MIXED syntax also generates model fit information for all models estimated. The output containing the AIC and BIC values for the unconditional model estimated using the syntax above is below.

Fit Statistics	
-2 Res Log Likelihood	51122.5
AIC (smaller is better)	51126.5
AICC (smaller is better)	51126.5
BIC (smaller is better)	51129.4

In order to answer the other research questions addressed by this study, we continued the model building process by first including student variables as fixed effects, then including these as random effects, and ending by including the school variables as fixed effects. More specifically, our next step in the model building process was to include students' Limited English Proficiency (LEP) status and Exceptional Child (EC) status as fixed effects in a level-1 model with random intercept only (Model 2) to examine the relationship between these variables and student math achievement. The PROC MIXED syntax for this model is shown below. Notice that this is the same code as used for Model 1 with the addition of the student-level predictors LEP and EC included on the MODEL statement.

```
PROC MIXED data=lib.SGF13 covtest noclprint method = ML;
class schoolid;
model ma_z = lep ec/solution ddfm = SATTERTHWAITE;
random intercept /sub=schoolid type=vc;
```

The Solution for Fixed Effects table generated by this PROC MIXED syntax is shown below.

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	-0.00108	0.08924	29.9	-0.01	0.9904
lep	-0.09763	0.01609	2E4	-6.07	<.0001
ec	-0.4737	0.01930	2E4	-24.55	<.0001

Then, we continued by adding these student variables as random effects (Model 3) in order to determine if their influence on math achievement varied among schools. The PROC MIXED syntax for this level-1 model with random intercept and slopes is shown below. You should notice that this syntax is similar to the syntax used to estimate Model 2, except now we have LEP and EC listed on the RANDOM statement – this allows the two level-1 variables to vary between schools.

```
PROC MIXED data=lib.SGF13 covtest noclprint method=ML;
class schoolid;
model ma_z = lep ec/solution ddfm = SATTERTHWAITE;
random intercept lep ec/ sub=schoolid type=vc;
```

The Covariance Parameter Estimates table shown below was generated by this PROC MIXED syntax.

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	schoolid	0.2831	0.07387	3.83	<.0001
lep	schoolid	0.05115	0.01611	3.18	0.0007
ec	schoolid	0.08612	0.02588	3.33	0.0004
Residual		0.6738	0.006681	100.85	<.0001

In order to address the impact of the school-wide intervention program while controlling for student and school characteristics, we continued the model building process by adding the school level variables to our model. More specifically, our last step in the model building process for this study was to include school participation in the intervention and the percent of students in the school receiving free or reduced-price lunch as fixed effects (Model 4). The PROC MIXED syntax for this level-1 and level-2 model with random intercept and slopes and the fixed effects output generated from this last model are shown below.

```
PROC MIXED data=lib.SGF13 covtest noclprint method= ML;
class schoolid;
model ma_z = lep ec treatment schl_frl_c/solution ddfm = SATTERTHWAITE;
random intercept lep ec/ sub=schoolid type=vc;
```

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	0.02681	0.07254	26	0.37	0.7147
lep	-0.09375	0.04513	28.7	-2.08	0.0468
ec	-0.4432	0.05842	29.9	-7.59	<.0001
treatment	0.09139	0.1864	25.9	0.49	0.6280
schl_frl_c	-1.8327	0.3359	25.7	-5.46	<.0001

Now that we have estimated all of our models, we have compiled the output into a single summary table – see Table 3. By having all of the output in one table, including the AIC and BIC values, we are able to easily assess for improvements in model fit across our four models. Then, once we identify the best fitting model, we will use it answer our 2nd and 3rd research questions.

As shown in Table 3, for each additional model we estimated, the AIC and BIC values in the more complex models are smaller than the values for the model directly preceding it. For example, by adding level-1 fixed effects in Model 2, the AIC and BIC values decreased 617.1 and 617.2, respectively, compared to the AIC and BIC values from Model 1. This same pattern in decreasing AIC and BIC values occurs across the four models. Therefore, we will answer the remaining research questions based on the output from the best fitting model, which is Model 4 – the random intercept and slope model with both level-1 and level-2 predictors.

Our second research question asked if the influence of any student-level predictors varied among schools. To answer this question, we will look at both the fixed effect and random effect estimates for the two student-level variables, LEP and EC. For both variables, it appears that not only is the relationship between each variable and math achievement is statistically significant (b = -0.09 & b = -0.44, respectively) but that the relationship between LEP and math achievement and EC and math achievement also vary between schools. We determined that the random slopes were statistically significant by both the *p*-values for the estimated variance components as well as the overall improvement in model fit when the random slopes were added to the model.

Table 3. *Estimates from Two-level Organizational Linear Models Predicting Math Achievement (N =20436)*

	Model 1	Model 2	Model 3	Model 4
Fixed Effects				
Intercept	-0.07	-0.001	-0.01	0.03
	(0.09)	(0.09)	(0.09)	(0.07)
Limited English Proficiency		-0.10*	-0.10*	-0.09*
		(0.02)	(0.04)	(0.04)
Exceptional Child Status		-0.47*	-0.44*	-0.44*
		(0.02)	(0.06)	(0.06)
Treatment				0.09
				(0.19)
School FRL				-1.83*
				(0.33)
Error Variance				
Level-1	0.71*	0.69*	0.67*	0.67*
	(0.01)	(0.01)	(0.01)	(0.01)
Level-2 Intercept	0.26*	0.24*	0.28*	0.08*
	(0.07)	(0.06)	(0.07)	(0.02)
Limited English Proficiency			0.05*	0.05*
			(0.02)	(0.02)
Exceptional Child Status			0.09*	0.09*
			(0.02)	(0.03)
Model Fit				
AIC	51126.5	50509.4	50223.7	50189.3
BIC	51129.4	50512.2	50229.5	50195.0

Note: *Statistically significant, $p<.05$; ICC = .27
Values based on SAS Proc Mixed. Entries show parameter estimates with standard errors in parentheses.
Estimation Method = ML; Satterthwaite degrees of freedom

Last, the third research question examined if a school-wide intervention program for low-performing schools impacted student math achievement while controlling for student and school characteristics. To answer this question, we look at the fixed effect output for the level-2 variable, treatment. As shown in Table 3, the regression coefficient for the treatment variable is 0.09 and it is not statistically significant. Thus, there is not enough evidence to conclude that the school-level treatment impacted math achievement. More specifically, it appears that when controlling for student LEP and EC status and school-level free and reduced lunch, the difference between math achievement for students in the intervention schools is not significantly different than the math achievement for students in control schools.

EXAMPLE 2: ESTIMATING TWO-LEVEL GROWTH MODELS

PROC MIXED can also be used for multilevel growth modeling. For the purposes of this example we will explore how children's reading achievement changed throughout the first two years of school. In order to answer this question, we would have to have student reading achievement data at different time points throughout their kindergarten and 1st grade school years (i.e., time is nested with students). Notice that in this real world situation, some children may be absent on assessment days and assessments may be given at different time points depending on the classroom the student is in. Because of this, we may have a large amount of missing data and differing time points in which children are given assessments creating an unbalanced data set. But, that is ok – growth models do not require observations to have complete data for all time points. They also do not require the time points to be equally spaced out across observations. These are two important benefits of estimating growth models vs. traditional repeated-measures ANOVA. For the purposes of this example, we used 120 children pulled from the Early Childhood Longitudinal Study-Kindergarten (ECLS-K) public use data that contains information about a student sample from Kindergarten until Grade 8. We were interested to see the growth of reading across kindergarten through grade 1 and the effect that student sex has on reading achievement. Specifically, we will answer the following research questions:

1. What is the average reading achievement for kindergarteners? Is there variation in children's reading achievement in kindergarten?
2. How does reading achievement change from kindergarten through first grade? Do children vary in their reading achievement growth over time?
3. Is a child's sex related to reading achievement in Kindergarten?

Before starting the model building process to answer our three research questions, it is important to note that to estimate growth models using PROC MIXED, our data must be in long format (i.e. multiple observations for each child). Below is a snapshot of the first 2 observations of the data set we are using. Notice in this data set there are

multiple reading scores for each observation, TIMEMOS is a variable that shows when the child was assessed (0 indicating at the beginning of Kindergarten, and 8, 12, and 20 indicating the number of months after starting Kindergarten), and sex is a variable where 0 indicates males and 1 indicates females.

Obs	CHILDID	READING	TIMEMOS	sex
1	15001	20.091	0.000	0
2	15001	26.855	8.000	0
3	15001	28.982	12.000	0
4	15001	54.190	20.000	0
5	15003	20.992	0.000	0
6	15003	22.804	8.000	0
7	15003	25.528	12.000	0
8	15003	48.226	20.000	0

We followed the same model building process as was used in example 1 but instead of students being nested in schools, time is nested in students -- time is our level-1 variable and students are the level-2 units. First, we estimated the unconditional model with no predictors (Model 1) to assess between-student variation in reading achievement. This model also addresses our first research question regarding the average reading achievement for kindergarteners and if children's reading achievement in kindergarten varied between students.

The PROC MIXED syntax and relevant output for the two-level unconditional growth model is shown below. Note that for this example childid is our level-2 identification variable so we use it on the CLASS statement and the sub = option on the RANDOM statement. Reading is our criterion variable and we do not have any predictor variables in this model.

```
proc mixed covtest noclprint data = temp method=ml;
class childid;
model reading = /solution ddfm = SATTERTHWAITE;;
random intercept /sub=childid type=vc;
```

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	CHILDID	47.7706	13.4292	3.56	0.0002
Residual		203.01	15.2423	13.32	<.0001

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	35.7501	0.9083	119	39.36	<.0001

Just as we did in the first example, using output from the Covariance Parameter Estimates table we can calculate the ICC for the unconditional growth curve model by calculating the ratio of the intercept variance to total variance (intercept + residual). Thus, 47.7706/250.7806 = .19. In the growth model context, the ICC of .19 tells us that 19% of the variance in reading achievement exists between children. This box also reveals that intercept variance in the unconditional model is statistically significant, thus, suggesting that children do vary in their reading achievement in kindergarten. This information answers half of the first research question. To answer the other part of the question, we have to use output found in the Solution for Fixed Effects table. In unconditional growth models, the intercept is the average value on the outcome being examined at time = 0. So, in this example, timemos = 0 is fall term of

kindergarten. Thus, the average reading achievement at the beginning of kindergarten for these 120 children is 35.75 and it is statistically significant.

The model fit statistics from the unconditional model are provided in the output below. Although these values are not terribly meaningful alone, they will be compared to the AIC and BIC estimates from Model 2 to determine if adding the time fixed effect to the model improves model fit.

Fit Statistics	
-2 Log Likelihood	3959.0
AIC (smaller is better)	3965.0
AICC (smaller is better)	3965.1
BIC (smaller is better)	3973.4

The PROC MIXED syntax used to estimate the second growth model is below. Here you see that we added the level-1 predictor timemos – the fixed effect parameter estimate for the variable timemos represents the average reading achievement growth. The fixed effect output from this second model is provided below.

```
proc mixed covtest noclprint data = temp method=ml;
class childid;
model reading = timemos/solution ddfm = SATTERTHWAITE;
random intercept /sub=childid type=vc;
.
```

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	19.8037	0.9806	156	20.20	<.0001
TIMEMOS	1.5813	0.03506	355	45.10	<.0001

Now, we have an estimate of the impact of time, which has a fixed effect value of 1.5813 and is statistically significant. Thus, with every one month time period that passes, on average, students increase 1.5813 points in reading achievement.

Next, in order to understand if the average growth in reading achievement varies across students, we need to estimate a model where our time variable (TIMEMOS) is added to our random line as shown below.

```
proc mixed covtest noclprint data = temp method=ml;
class childid;
model reading = timemos/solution ddfm = SATTERTHWAITE;
random intercept timemos/sub=childid type=VC;
```

From this output, we need our AIC and BIC information as well as our random effect parameters from our Covariance Parameter Estimates. Both pieces of the output are provided below.

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	CHILDID	73.2412	11.2629	6.50	<.0001
TIMEMOS	CHILDID	0.1115	0.02898	3.85	<.0001
Residual		23.6180	2.0253	11.66	<.0001

Fit Statistics	
-2 Res Log Likelihood	3256.7
AIC (smaller is better)	3262.7
AICC (smaller is better)	3262.8
BIC (smaller is better)	3271.1

Based on the results in the Covariance Parameter output table above, it seems that growth in reading achievement does vary between children.

Lastly, we are interested to see if there are differences in baseline (i.e., timemos = 0) reading achievement for girls and boys. In order to examine this, we must add sex as a level-2 predictor to our previously estimated random intercept and slope mode. The PROC MIXED syntax for this fourth and final model is below.

```
proc mixed covtest noclprint data = temp method=ml;
class childid;
model reading = sex timemos/solution ddfm = SATTERTHWAITE;
random intercept timemos/sub=childid type=VC;
```

Since we only changed the model by adding a fixed effect, our primary focus is the parameter estimate for the level-2 predictor, sex. The fixed effect output table from Model 4 is below.

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	18.6220	1.3367	136	13.93	<.0001
TIMEMOS	1.5812	0.03506	355	45.09	<.0001
sex	2.3655	1.8239	118	1.30	0.1972

Notice that sex does not seem to be a statistically significant level-2 predictor. However, before we answer all of the research questions for this example, we need to examine changes in AIC and BIC. Doing so will allow us to determine the most parsimonious and best fitting model. Then, after the best fitting model is selected, we will answer our research questions from that model. As we did in the first example, we have compiled all of the output for our four models into a single summary table – see Table 4 below.

Again, before answering our three research questions for this example, first, we need to examine model fit as we move from the unconditional model to the most complex random intercept and slope model with level-1 and level-2 predictors. As show in Table 4, as the models became more complex, the AIC and BIC values decreased, thus indicating better model fit throughout the progression of models. Although the improvement in model fit between Model 3 and 4 is not very big, the 4.5 decrease is in Raftery's (1995 as cited in O'Connell & McCoach, 2008) range that suggests favoring the more complex model. Therefore, both Model 1, the unconditional model, and Model 4 will be used to answer our research questions.

For research question 1, we wanted to know what the average reading achievement for kindergartners was and if there was variation in children's reading achievement at kindergarten entry. As explained previously, we find the answers to these questions in the results from the unconditional model (Model 1). Here we see that the fixed effect for the intercept is 35.7501. The fixed effect for the intercept is mathematically equivalent to the grand mean for all children. Thus, the average reading achievement of kindergarteners is 35.7501. In order to understand if there is variation in children's level of reading at kindergarten entry, we interpret the results for the intercept's variance estimate. Although the magnitude of the intercept variance estimate is rarely interpreted in a substantive way, the p-value for that variance parameter indicates that reading achievement at entry into kindergarten does vary between children.

Table 4. *Estimates from Two-level Linear Growth Models Examining Growth in Reading Achievement (N =120)*

	Model 1	Model 2	Model 3	Model 4
Fixed Effects				
Intercept	35.75*	19.80*	19.77*	18.68*
	(0.91)	(0.98)	(0.87)	(1.21)
Time (in months)		1.58*	1.58*	1.58*
		(0.0.04)	(0.04)	(0.04)
Biological sex				2.18
				(1.68)
Error Variance				
Level-1	203.01*	30.20*	23.62*	23.64*
	(15.24)	(2.27)	(2.03)	(2.03)
Intercept	47.77*	92.75*	73.24*	72.74*
	(13.43)	(13.04)	(11.26)	(11.24)
Time			0.11*	0.11*
			(0.03)	(0.03)
Model Fit				
AIC	3965.0	3287.5	3262.7	3258.2
BIC	3973.4	3293.1	3271.1	3266.6

Note: *Statistically significant, p<.05; ICC = .19
Values based on SAS PROC Mixed. Entries show parameter estimates with standard errors in parentheses. Estimation Method = ML; Satterthwaite degrees of freedom

Next, for research question 2, we wanted to know how reading achievement changed from kindergarten through first grade and if children varied in their reading achievement growth. In order to answer this question, we want to examine the fixed and random effect results for the predictor time using output from Model 4. As shown in Table 4, both of these effects are statistically significant; children have positive growth in reading achievement over time such that with every one month increase in time, reading achievement increased 1.5813 units. Likewise, the statistically significant variance estimate for time indicates that children are not all changing at the same rate.

The third research question inquired about the relationship between child sex and reading achievement in kindergarten. Based on the output from Model 4 in Table 4, it appears that there is not enough evidence to suggest that boys and girls differ in their reading achievement upon entry into kindergarten (b = 2.18, p> 05).

EXAMPLE 3: ESTIMATING THREE-LEVEL ORGANIZATIONAL MODELS

In order to showcase the utility of PROC MIXED for fitting three-level organization models, we expanded the two-level organization model by also accounting for the clustering of students in classrooms and using years of teaching experience as a teacher-level predictor variable. So, in the three-level model we now have students nested in classrooms/teachers nested in schools. The data utilized in this example is identical to those used in the first example, with the addition of a classroom/teacher identification variable and a continuous teacher-level variable denoting the number of years teaching experience (tchexp), which was grand-mean centered for ease of interpretation.

Research questions that could be explored using these three-level models include:

1. How much of the variance in math achievement scores is attributable to students, teachers, and schools?
2. Does the influence of any student-level predictor vary among teachers or schools? Does the influence of the teacher-level predictor vary among schools?
3. What is the impact of a school-wide intervention program, controlling for student, teacher, and school characteristics?

However, for brevity, we do not walk through all of the results again. Instead, we include the PROC MIXED syntax for the models that were estimated, highlighting differences in comparison to the syntax for the two-level organizational model example. We also provide some model output, again, to highlight differences from the output from the 2-level models.

14

First, we estimated an unconditional model so we could calculate the relevant ICC values. Note that in a 3-level model, there will actually be two ICC values to calculate. In this example, we calculated the ICC for teachers and the ICC for schools. The PROC MIXED syntax for our unconditional model is listed below.

```
proc mixed data=sgf13_x3 noclprint method=ml;
class teacherid schoolid;
model ma_z=/solution;
random intercept/sub=schoolid type=vc;
random intercept/sub=teacherid(schoolid) type=vc;
```

The key difference between this syntax and the syntax used to estimate the unconditional 2-level model is the use of two RANDOM statements. The first RANDOM statement is analogous to the RANDOM statement used in example 1 but instead of specifying school as the subject at level-2, we are specifying school as the subject at level-3. Then, the second RANDOM statement specifies the subject as teachers nested within schools at level-2 [i..e., sub = teacherid(schoolid)].

The Covariance Parameter Estimates table resulting from this unconditional model appears below.

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Intercept	teacherid(schoolid)	0.1287
Intercept	schoolid	0.2009
Residual		0.6070

We then used the three variance estimates to calculate teacher ICC and school ICC. The formula for calculating each of these ICCs is similar to the formula used for two-level ICC calculations, however, the denominator now has three elements (residual variance, school variance, and teacher variance) which form the total variance in the model. The two equations for the teacher ICC and school ICC are below.

$$\text{ICC}_T = \frac{\sigma_T^2}{\sigma_e^2 + \sigma_T^2 + \sigma_S^2}$$

$$\text{ICC}_S = \frac{\sigma_S^2}{\sigma_e^2 + \sigma_T^2 + \sigma_S^2}$$

ICC_T expresses the similarity (correlation between) of students instructed by the same teacher (within the same school); or alternatively informs us of how much of the total variation in math achievement exists between teachers. ICC_S expresses the similarity of teachers within the same school, ignoring within-class variation (Hox, 2010).

Plugging the values from the Covariance output table above into our ICC formulas from above, we learn that 13.7% of the variation in math achievement exists between teachers and 21.4% exists between schools, leaving 64.8% of the variance in math achievement existing within students. Thus, a practically meaningful proportion of the variance in math achievement exists at the teacher and school levels, providing support for the use of a three-level analytical model.

The PROC MIXED syntax for estimating the second model, the random intercept model with level-1 predictors is below.

```
proc mixed data=sgf13_x3 noclprint covtest method = ml;
class studentid teacherid schoolid;
model ma_z=lep ec/solution;
random intercept/sub=teacherid(schoolid) type=vc;
random intercept/sub=schoolid type=vc;
```

Next, the PROC MIXED syntax for estimating the third model, the random intercept and slope model with level-1 predictors, is below. The covariance output from this model is also included to show users how many variance

components get estimated in a 3-level model with just two level-1 random slopes. Note here we list the level-1 predictors on both of the RANDOM statements – doing so allowed our level-1 variables to vary between teachers and between schools. The output also clearly shows the estimates for the random slopes between teachers and between schools. In this example, all of the random slopes are significant, however, often times, one or more slopes in these more complex models turn out to not be significant. If that were to have happened, we could have removed the non-significant random slopes and re-estimated the model.

```
proc mixed data=sgf13_x3 noclprint covtest method = ml;
class teacherid schoolid;
model ma_z=lep ec/solution;
random intercept lep ec/sub=teacherid(schoolid) type=vc;
random intercept lep ec/sub=schoolid type=vc;
```

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	teacherid(schoolid)	0.1151	0.01082	10.64	<.0001
lep	teacherid(schoolid)	0.03068	0.009110	3.37	0.0004
ec	teacherid(schoolid)	0.08098	0.01754	4.62	<.0001
Intercept	schoolid	0.2267	0.06429	3.53	0.0002
lep	schoolid	0.03847	0.01324	2.91	0.0018
ec	schoolid	0.03557	0.01581	2.25	0.0122
Residual		0.5728	0.005794	98.87	<.0001

The syntax for the fourth model – a random intercept and slope model with student-level predictors and a teacher-level predictor – is below.

```
proc mixed data=sgf13_x3 noclprint covtest method = ml;
class teacherid schoolid;
model ma_z=lep ec tchexp /solution;
random intercept lep ec/sub=teacherid(schoolid) type=vc;
random intercept lep ec/sub=schoolid type=vc;
```

Last, the syntax for the fifth model – a random intercept and slope model with student, teacher, and school predictors is below.

```
proc mixed data=sgf13_x3 noclprint covtest method = ml;
class teacherid schoolid;
model ma_z=lep ec tchexp treatment schl_frl_c /solution;
random intercept lep ec/sub=teacherid(schoolid) type=vc;
random intercept lep ec/sub=schoolid type=vc;
```

Table 5 below provides the summary information for the example 3 model-building process. Model 1 represents the initial unconditional model to estimate the amount of variance in math achievement attributable to students, teachers and schools. Model 2 estimates were based on the inclusion of student-level predictors including student LEP and EC status. Model 3 expanded on Model 2 by specifying the LEP and EC predictors as random slopes at both the teacher and school level. Model 4 expanded on Model 3 and included teacher experience at level-2, and Model 5 included the program treatment variable and the percent of students at schools that received Free or Reduced-Price Lunch (FRL). Examining the AIC and BIC values at the bottom of Table 5, we can see that each progressive model exhibited better fit to the data, except for Model 4 which included a single teacher-level predictor. Note that the teacher-level predictor was dropped from Model 5 therefore Model 3 is nested in Model 5, thus, Model 5 fit values

were compared to the values from Model 3. We did this because Model 4 was not a better fitting model than Model 3, therefore Model 4 should not have been used to estimate more complex models.

Beginning with the fixed effects from Model 5, we can see that student EC status was the only significant student-level predictor, suggesting that students with disabilities tend to have significantly lower math achievement than students without disabilities. Student LEP status was a significant predictor in Model 2, but was no longer a significant fixed effect once specified as a random effect at the teacher and school levels in Model 3. Examining the random effects section, we see that significant variability in school and teacher intercepts, as well as the coefficients associated with LEP and EC, existed even after controlling for these student-level fixed effects. Thus, the association between student LEP and EC statuses and math achievement varies significantly among teachers and among schools (as denoted by the sub-T and sub-S, respectively). Thus, coefficients associated with these variables may be stronger/weaker from teacher to teacher within a school or depending on the school attended by a student.

We can also see that the school-level intervention had a positive, but non-significant impact on math achievement, while school FRL had a significant, negative impact on math achievement. Note also the drastic reduction in the random intercept variance associated with schools in Model 5 compared to Model 3. This informs us that the % of students receiving FRL, an indicator of Socio-Economic Status (SES) within the school, explains a large share of variability in mean math achievement among schools.

Lastly, it is important to note that in this example, there could have actually been a sixth model estimated; had the teacher-level variable in Model 4 been a significant predictor and had improved model fit, the next model that would have been estimated would have been a model in which the teacher-level predictor would have been allowed to vary across schools (i.e., we would have estimated a model that had the teacher-level variable specified as a random effect on school RANDOM statement). So, although we have tried to provide users with a thorough explanation of how to use PROC MIXED to estimate a variety of multilevel models, most often, slight nuances, such as this issue of not estimating a model with a random slope for years of teaching experience, come into play.   Thus, we recommend that researchers use this paper to help guide them in using PROC MIXED to estimate multilevel models, in conjunction with a multilevel modeling textbook.

Table 5. *Estimates from Three-level Organizational Linear Models Predicting Math Achievement (N =20436)*

	Model 1	Model 2	Model 3	Model 4	Model 5
Fixed Effects					
Intercept	-0.11	-0.04	-0.05	-0.04	-0.04
	(-0.08)	(0.08)	(0.09)	(0.09)	(0.06)
LEP		-0.07*	-0.06	-0.06	-0.06
		(0.02)	(0.04)	(0.04)	(0.04)
EC		-0.45*	-0.44*	-0.43*	-0.44*
		(0.02)	(0.05)	(0.05)	(0.05)
Teacher Experience				0.00	
				(0.00)	
Treatment					0.14
					(0.15)
School FRL					-1.78*
					(0.27)
Error Variance					
Level-1	0.61*	0.59*	0.57*	0.57*	0.57*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Intercept (Teacher)	0.13*	0.12*	0.12*	0.12*	0.12*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Intercept (School)	0.21*	0.19*	0.23*	0.23*	0.04*
	(.06)	(.06)	(.06)	(.06)	(.02)
Slope (LEP$_T$)			0.03*	0.03*	0.03*
			(.01)	(.01)	(.01)
Slope (EC$_T$)			0.08*	0.08*	0.08*
			(.02)	(.02)	(.02)
Slope (LEP$_S$)			0.04*	0.04*	0.04*

			(.01)	(.01)	(.01)
Slope (EC$_S$)			0.04*	0.04*	0.04*
			(.02)	(.02)	(.02)
Model Fit					
AIC	48723.9	48171.6	47874.4	47885.3	47836.5
BIC	48728.2	48175.9	47884.5	47895.4	47846.5

Note: *Statistically significant, p<.05; School ICC = .21; Teacher ICC = .14. T = random effect at teacher level; S = random effect at school level. Values based on SAS PROC Mixed. Entries show parameter estimates with standard errors in parentheses. Estimation Method = ML; Satterthwaite degrees of freedom

EXAMPLE 4: ESTIMATING THREE-LEVEL GROWTH MODELS

Just as example 1 was expanded from a two-level organizational model to a three-level organizational model, the same process could be done with example 2. For example, one could estimate a growth model with time nested in students nested in schools. Doing so would allow a researcher to examine the amount of variance in reading achievement that was between students as well as the amount of variance that was between elementary schools. A three-level growth model would look similar to the models estimated in example 3 (i.e., two random statements – one to indicate time nested in students and another to indicate students nested in schools). However, for brevity, we have not included details on estimating three-level growth models. We feel that users can easily adapt the syntax from the three-level organizational models.

CONCLUSIONS

This paper provides researchers with an extensive introduction to estimating a variety of multilevel linear models via PROC MIXED. Likewise, researchers who are relatively new to world of multilevel modeling, we have also provided a comprehensive review of key issues to consider when working with nested data structures. Yet, given the details included in this paper, it is still very much an overview of estimating relatively simple and straightforward models. There are many more advanced areas within the world of multilevel modeling (e.g., issues surrounding sample size, cross-level interactions, checking distributional assumptions, including quadratic terms in growth models). Thus, we want to reminder users that this paper is indeed an introduction to multilevel modeling and PROC MIXED. We suggest that researchers use this paper in conjunction with a multilevel modeling textbook that can offer them more details on the theories and nuances of the multilevel models. Lastly, in conclusion, we hope that you, the applied researcher, find this primer helpful as you navigate the great world of multilevel models!

REFERENCES

Diez Roux, A.V. (2002). A glossary for multilevel analysis. *Journal of Epidemiology and Community Health, 56,* 588-594. doi:10.1136/jech.56.8.588.

Donner, A. & Klar, N. (2000). *Design and analysis of cluster randomization trials in health research.* London: Edward Arnold.

Klein, K.J. & Kozlowski, S.W. (2000). *Multilevel theory, research, and methods in organizations: Foundations, extensions and new directions.* San Francisco: Jossey-Bass.

Early Childhood Longitudinal Study - Kindergarten [ECLS-K]. (n.d.). Retrieved from the National Center for Educational Statistics website, http://nces.ed.gov/ecls/

Enders, C.K. & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12,* 121-138. doi:10.1037/1082-989X.12.2.121

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed). London: Edward Arnold.

Heck, R.H. & Thomas, S.L. (2000). *An introduction to multilevel modeling techniques.* Mahwah, NJ: Erlbaum.

Hofmann, D.A. & Gavin, M.B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management, 24,* 623-641.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Lawrence Erlbaum.

Julian, M. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling, 8,* 325-352. doi: 10.1207/S15328007SEM0803_1

McCoach, D.B. & Black, A.C. (2008). Evaluation of model fit and adequacy. In A. A. O'Connell & D. B. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 245-272). Charlotte, NC: Information Age Publishing.

Moerbeek, M. (2004). The consequences of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research, 39*, 129-149. doi: 10.1207/s15327906mbr3901_5

Murray, D.M. (1998). *Design and analysis of group-randomized trials.* New York, NY: Oxford University Press.

Nich, C. & Carroll, K. (1997). Now you see it, now you don't: A comparison of traditional versus random-effects regression models in the analysis of longitudinal follow-up data from a clinical trial. *Journal of Consulting and Clinical Psychology, 65,* 252-261. doi: 10.1037//0022-006X.65.2.252

O'Connell, A.A. & McCoach, D.B. (2008). *Multilevel modeling of educational data*. Charlotte, NC: Information Age Publishing.

Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Shadish, W., Cook, T. & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MD: Houghton Mifflin*.*

Snijders, T. A.B. & Bosker, R.J. (1999). *Multilevel analysis: An Introduction to basic and advanced multilevel modeling.* Thousand Oaks, CA: Sage.

Wampold, B.E. & Serlin, R.C. (2000). The consequences of ignoring a nested factor on measures of effect size in analysis of variance. *Psychological Methods, 5,* 425-433. doi:10.1037//1082-989X.5.4.425

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Bethany A. Bell
University of South Carolina
College of Education, Wardlaw 133
Columbia, SC 29208
Work Phone: 803-777-2387
E-mail:babell@sc.edu