

Multilevel Models for Categorical Data Using SAS[®] PROC GLIMMIX: The Basics

Mihaela Ene, Elizabeth A. Leighton, Genine L. Blue, and Bethany A. Bell
University of South Carolina

ABSTRACT

Multilevel models (MLMs) are frequently used in social and health sciences where data are typically hierarchical in nature. However, the commonly used hierarchical linear models (HLMs) are appropriate only when the outcome of interest is normally distributed. When you are dealing with outcomes that are not normally distributed (binary, categorical, ordinal), a transformation and an appropriate error distribution for the response variable needs to be incorporated into the model. Therefore, hierarchical generalized linear models (HGLMs) need to be used. This paper provides an introduction to specifying HGLMs using PROC GLIMMIX, following the structure of the primer for HLMs previously presented by Bell, Ene, Smiley, and Schoeneberger (2013). A brief introduction into the field of multilevel modeling and HGLMs with both dichotomous and polytomous outcomes is followed by a discussion of the model-building process and appropriate ways to assess the fit of these models. Next, the paper provides a discussion of PROC GLIMMIX statements and options as well as concrete examples of how PROC GLIMMIX can be used to estimate (a) two-level organizational models with a dichotomous outcome and (b) two-level organizational models with a polytomous outcome. These examples use data from High School and Beyond (HS&B), a nationally representative longitudinal study of American youth. For each example, narrative explanations accompany annotated examples of the GLIMMIX code and corresponding output.

Keywords: MULTILEVEL MODELING, HIERARCHICAL GENERALIZED LINEAR MODELS, PROC GLIMMIX

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 (e.g., students) or contextual levels (e.g., schools) can lead to erroneous conclusions. More specifically, 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 & Carroll, 1997). Multilevel models have been developed to properly account for the hierarchical (correlated) nesting of data (Heck & Thomas, 2000; Hox, 2002; Klein & Kozlowski, 2000; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999) and are frequently used in social and health sciences where data are typically hierarchical in nature.

Whereas these models are often named generically as multilevel models or hierarchical linear models, there are many types of multilevel models. More specifically, these models could differ in terms of the number of levels [e.g., 2-level (students nested within schools), 3-level (students nested within schools nested within districts)], 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). However, some of these models such as those with normally distributed outcomes are more commonly discussed in the literature than the models with non-normal outcomes. Also, even when considered, models with dichotomous outcomes (e.g., pass/fail) are more often discussed than those with polytomous outcomes (e.g., below basic, basic, proficient), the latter ones being considered a natural extension of the binary version. While this is indeed the case in terms of conceptualizing the models, there are certain particularities of the models with polytomous outcomes (e.g., syntax, output, interpretation) that may pose challenges for the researcher who is not familiar with this type of model. Therefore, this paper aims to introduce the reader to multilevel models with non-normal outcomes (i.e., hierarchical generalized linear models), to explain the differences between the models with dichotomous and polytomous outcomes, and to provide concrete examples of how to estimate and interpret these models using PROC GLIMMIX.

HIERARCHICAL GENERALIZED LINEAR MODELS (HGLMs)

In contrast with HLMs that have continuous, approximately normally distributed outcomes, HGLMs are appropriate to use for categorical, non-normally distributed response variables including binary, proportions, count, or ordinal data. When dealing with this type of model, the assumptions of normally distributed, homoscedastic errors are violated (Hox, 2002; O'Connell, Goldstein, Rogers, & Peng, 2008). Therefore, the transformation of the outcome using a nonlinear link function and the choice of the appropriate non-normal error distribution need to be incorporated into the models so the model building strategies and the interpretations used for HLMs will still be applicable (Luke, 2004).

In the case of multilevel models with dichotomous outcomes, the binomial distribution (i.e., Bernoulli) and the logit link are most commonly used to estimate for example, the odds of success and the impact of various characteristics at different levels on these odds (i.e., odds ratio). Similarly, the models with polytomous, ordinal-type outcomes use a multinomial distribution and a cumulative logit link to compute the cumulative odds for each category of response, or the odds that a response would be at most, in that category (O'Connell et al., 2008). Although a detailed discussion regarding the theoretical reasons behind these choices is beyond the purpose of this paper, we will mention these distributions and link functions throughout the paper as they represent key elements when estimating HGLMs.

Understanding the conceptualization of these models in a hierarchical generalized linear framework also requires knowledge of the mathematical equations used to estimate the models. For example, a researcher might be interested in modeling the math achievement of individual students (level-1) nested within schools (level-2). In the dichotomous case (e.g., pass/fail), the researcher might be interested in exploring the pass rate at a typical school and the influence of student and school characteristics on the chance of passing the math test. The equations necessary for estimating this model are presented below.

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

Equation 1 represents a simple level-1 model with one student-level predictor, where η_{ij} represents the log odds of passing the math test for student i in school j , β_{0j} is the intercept or the average log odds of passing the math test at school j , X_{ij} is a student-level predictor for student i in school j , and β_{1j} represents the slope associated with X_{ij} , showing the relationship between the student-level variable and the log odds of passing the math test. It is important to notice that unlike hierarchical linear models, this model has no error variance at level-1. This is not estimated separately when dealing with hierarchical generalized linear models with binary outcomes because in this case, the variance is a function of the population mean and is directly determined by this mean (Luke, 2004).

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} W_j + u_{0j} \\ \beta_{1j} &= \gamma_{10} \end{aligned} \quad (\text{Eq. 2})$$

Equation 2 represents a simple level-2 model with one school-level predictor, where γ_{00} provides the log odds of passing the math test at a typical school, W_j is a school-level predictor for school j , γ_{01} is the slope associated with this predictor, u_{0j} is the level-2 error term representing a unique effect associated with school j , and γ_{10} is the average effect of the student-level predictor. As the effect of the student-level predictor is modeled as fixed or constant across schools, this represents a random intercept-only model.

$$\eta_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} W_j + u_{0j} \quad (\text{Eq. 3})$$

Next, the combined level-1 and level-2 model (Equation 3) is created by substituting the values of β_{0j} and β_{1j} as shown in Equation 2 into the level-1 equation represented in Equation 1. As presented in this combined model, the log odds of passing the math test for student i in school j (η_{ij}) is determined by the log odds of passing the math test of a typical student at a typical school (γ_{00}), the effect of the student-level ($\gamma_{10} X_{ij}$) and school-level predictor ($\gamma_{01} W_j$), as well as the school-level error [$u_{0j}, u_{0j} \sim N(0, \tau_{00})$]. To be more meaningful for the researcher, the log odds of success could be converted into probabilities of success (Equation 4).

$$\phi_{ij} = \frac{e^{\eta_{ij}}}{1 + e^{\eta_{ij}}} \quad (\text{Eq. 4})$$

where e takes a value of approximately 2.72, η_{ij} represents the log odds of success, ϕ_{ij} is the probability of success, and $1 - \phi_{ij}$ is the probability of failure. In addition, odds ratios could also be calculated to reflect for instance, the predicted change in odds between students with various values of the student-level predictor.

In the polytomous case (e.g., below basic, basic, proficient) the researcher might be interested in the probability of being at or below a proficiency level and similar with the dichotomous model, in the influence of student and school characteristics on this probability for each category. When dealing with polytomous outcomes, multiple logits are simultaneously estimated (M-1 logits, where M = the number of response categories). Therefore, in this situation with three categories of response, there will be two logits and their corresponding intercepts simultaneously estimated, each of them indicating the probability of responding in or below a particular category. The equations necessary for estimating this model are presented below.

$$\begin{aligned} \eta_{1ij} &= \log\left(\frac{P(R_{ij} \leq 1)}{1 - P(R_{ij} \leq 1)}\right) = \beta_{0j} + \beta_{1j} X_{ij} \\ \eta_{2ij} &= \log\left(\frac{P(R_{ij} \leq 2)}{1 - P(R_{ij} \leq 2)}\right) = \beta_{0j} + \beta_{1j} X_{ij} + \delta_j \end{aligned} \quad (\text{Eq. 5})$$

Equation 5 represents the level-1 model with one student-level predictor, where η_{ij} is the log odds of being at or below a proficiency level for student i in school j . Compared to the level-1 model for dichotomous outcomes previously presented, this model consists of two equations instead of one. More specifically, η_{1ij} corresponds to the log odds of being at or below the lowest proficiency level (i.e., below basic) for student i in school j , $P(R_{ij} \leq 1)$ represents the probability of responding at or below this lowest proficiency level, and $1 - P(R_{ij} \leq 1)$ corresponds to the probability of responding above this lowest proficiency level for student i in school j . Next, β_{0j} is the intercept or the average log odds of being at or below the below basic level in math at school j , X_{ij} is a student-level predictor for student i in school j , and β_{1j} represents the slope associated with X_{ij} , showing the relationship between the student-level variable and the log odds of scoring at or below this level in math. Similarly, η_{2ij} represents the log odds of being at or below the next proficiency level (i.e., basic) for student i in school j . Notice that, for the ease of estimation and interpretation, the intercept for this proficiency level is represented by a common intercept (β_{0j}) and an extra term, δ_j , representing the difference between this category and the preceding one. However, there is only one slope (β_{1j}) associated with the student-level predictor (X_{ij}) that remains constant across logits.

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}W_j + u_{0j} & (\text{Eq. 6}) \\ \beta_{1j} &= \gamma_{10} \\ \delta_j &= \delta\end{aligned}$$

In the level-2 model with one school-level predictor (Equation 6), γ_{00} represents the log odds of being at or below the below basic level in math at a typical school, W_j is a school-level predictor for school j , γ_{01} is the slope associated with this predictor, u_{0j} is the level-2 error term representing a unique effect associated with school j , and γ_{10} is the average effect of the student-level predictor. Similar with the dichotomous model, the effect of the student-level predictor is modeled as fixed or constant across schools, therefore this equation represents a random intercept-only model. Also, whereas this model allows the common intercept (β_{0j}) to vary across schools, the difference between the logits (δ) remains fixed across schools.

Similar with the dichotomous example, the combined level-1 and level-2 model (Equation 7) can be created by substituting the values of β_{0j} , β_{1j} , and δ_j as shown in Equation 6 into the level-1 equation represented in Equation 5. As a result, the log odds of being at or below the lowest proficiency level in math for student i in school j (η_{1ij}) will be determined by the log odds of being at or below that proficiency level of a typical student at a typical school (γ_{00}), the effect of the student-level ($\gamma_{10}X_{ij}$) and school-level predictor ($\gamma_{01}W_j$), as well as the school-level error [u_{0j} , $u_{0j} \sim N(0, \tau_{00})$]. In addition, the log odds of being at or below the next proficiency level (η_{2ij}) will also include the difference between the intercepts corresponding to this category and the preceding one (δ). For the ease of interpretation, the log odds could be converted in odds or probabilities as described for the dichotomous model.

$$\begin{aligned}\eta_{1ij} &= \log\left(\frac{P(R_{ij} \leq 1)}{1 - P(R_{ij} \leq 1)}\right) = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}W_j + u_{0j} & (\text{Eq. 7}) \\ \eta_{2ij} &= \log\left(\frac{P(R_{ij} \leq 2)}{1 - P(R_{ij} \leq 2)}\right) = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}W_j + \delta + u_{0j}\end{aligned}$$

MODEL BUILDING PROCESS

The transformation of the dependent variable in HGLM via a non-linear link function and the choice of an appropriate non-normal error distribution allow researchers to engage in the standard model building strategies used with HLMs (O'Connell et al., 2008). While different sources provide different guidelines on the model building process when estimating multilevel models, the eventual goal is the same – to estimate the most parsimonious models that best fit the data. Other common elements of the different model building strategies include (a) always starting with an unconditional model (i.e., a model containing 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.

As researchers build their models by adding in additional parameters to be estimated model by model, fit statistics/indices are used to assess model fit. In multilevel linear modeling, estimating the model parameters via maximum likelihood (ML) estimation allows the researcher to assess model fit using one of two common procedures: (1) conducting a likelihood ratio test when examining differences in the -2 log likelihood (-2LL) when models are nested (i.e., models that have been fit using the same data and where one model is a subset of the other), also known as a deviance test or (2) investigating the change in Akaike's Information Criterion (AIC) and Bayesian

Information Criterion (BIC) when the models are either nested or not nested and differ in both fixed and random effects. Given the non-normal nature of the outcome variables in HGLMs, the use of ML estimation is not appropriate. However, there are several available estimation techniques for HGLM that are approximate to ML or are “quasi-likelihood” strategies. While a detailed discussion about the various estimation techniques available for HGLMs is beyond the scope of this paper, our examples utilize a common estimation technique available with PROC GLIMMIX, the Laplace estimation. By using Laplace estimation, the researcher is able to assess model fit using the same strategies applied with multilevel linear modeling. In the examples presented in this paper, all models examined are nested models, therefore we will assess model fit by examining the change in the -2LL between models.

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 Table 1). 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 Table 1 are general in nature. It is possible for more complex models to be examined, given the nature of the researcher’s questions. For example, our examples are for random intercept only models. It is also possible for researchers to examine how level-1 variables vary between level-2 units (i.e., the addition of random slopes) or examine relationships at more than two levels. For a more detailed discussion about more complex multilevel models, please see Bell, Ene, Smiley, and Schoeneberger (2013).

Model 1	Model 2	Model 3
No predictors, just random effect for the intercept	Model 1 + level-1 fixed effects	Model 2 + level-2 fixed effects
Output used to calculate ICC – provides information on how much variation in the outcome exists between level-2 units	Results indicate the relationship between level-1 predictors and the outcome	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 2

Table 1. Model Building Process for 2-level Generalized Linear Models with Random Intercepts Only

As you progress through the suggested model building process outlined in Table 1, you also examine model fit as measured by change in the -2LL. Lower deviance implies better fit; however, models with more parameters will always have lower deviance. Researchers can use a likelihood ratio test to investigate whether or not the change in the -2LL is statistically significant. This likelihood ratio test is analogous to a chi-square difference test, where χ^2 is equal to the difference in the -2LL of the more simple model (i.e., the model with fewer parameters to be estimated) minus the -2LL of the more complex model, and with degrees of freedom (*df*) equal to the difference in the number of parameters between the two nested models.

PROC GLIMMIX STATEMENTS AND OPTIONS

Before explaining the PROC GLIMMIX statement and options, we would like to address an important data management issue – the centering of all predictor variables used in a model. 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. For a well-written discussion on the pros and cons of the two approaches to centering predictor variables, we refer readers to Enders and Tofighi (2007). In the examples presented in our paper, we utilized both binary and standardized predictors, therefore, our variables already had a “meaningful zero” value for interpretation of the intercept and other predictors of interest.

PROC GLIMMIX is similar to PROC MIXED (used with multilevel linear models), as well as other modeling procedures in SAS, in that the researcher is required to specify a CLASS, MODEL, and RANDOM statement. In addition, PROC GLIMMIX also requires the researcher to specify a COVTEST statement which will generate hypothesis tests for the variance and covariance components (i.e., the random effects).

Below are two example codes for PROC GLIMMIX. The first example is code used to estimate a simple random intercept only model with one level-1 predictor (x1) and one level-2 predictor (w1) with a dichotomous outcome (y). The second example is code used to estimate a simple random intercept only model with one level-1 predictor (x1) and one level-2 predictor (w1) with a polytomous outcome (y).

```
PROC GLIMMIX DATA=SESUG2014 METHOD=LAPLACE NOCLPRINT;
CLASS L2_ID;
MODEL Y (EVENT=LAST) = X1 W1/ CL DIST=BINARY LINK=LOGIT SOLUTION ODDSRATIO (DIFF=FIRST LABEL);
```

```

RANDOM INTERCEPT / SUBJECT=L2_ID TYPE=VC SOLUTION CL;
COVTEST / WALD;

PROC GLIMMIX DATA=SESUG2014 METHOD=LAPLACE NOCLPRINT;
CLASS L2_ID;
MODEL Y = X1 W1 / CL DIST=MULTI LINK=CLOGIT SOLUTION ODDSRATIO (DIFF=FIRST LABEL);
RANDOM INTERCEPT / SUBJECT=L2_ID TYPE=VC SOLUTION CL;
COVTEST / WALD;

```

On the PROC GLIMMIX statement for both the dichotomous and polytomous outcome examples, in addition to listing the data set to use for the analysis, we have requested METHOD and NOCLPRINT options; the NOCLPRINT option is requesting SAS to not print the CLASS level identification information, and the METHOD=OPTION is required to tell SAS to estimate the models using the Laplace estimation method. The default estimation method in PROC GLIMMIX is METHOD=RSPL, a residual pseudo-likelihood estimation technique. However, the Laplace estimation is approximate to maximum likelihood (ML) estimation and will allow the researcher to compare and select the best fitting model via a deviance test. While a detailed discussion about the various estimation methods available with PROC GLIMMIX is beyond the scope of this paper, we encourage readers to read the SAS documentation for PROC GLIMMIX for more technical information.

The CLASS statement is used to indicate the clustering/grouping variable(s); in the example codes above, you can see that L2_ID is the clustering variable in 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. The EVENT=LAST option provided on the MODEL statement requests that the predicted log of the odds of success (i.e., the variable coded as “1”) be modeled. This is not required when there are more than two response categories; therefore, it has been removed from the example for a polytomous outcome model. For these models, the default option in SAS is to model the log odds of lower levels of the ordinal outcome, in the sequence encountered in the data. The CL option requests SAS to print confidence limits for fixed effect parameters. The DIST=OPTION and LINK=OPTION are required for HGLMs in order to transform the outcome variable into an approximately linear variable with normally distributed errors. In our examples we have chosen two options that are common for both dichotomous outcomes (a binary distribution and the logit link) and polytomous outcomes (the multinomial distribution and the cumulative logit link). The SOLUTION option on the MODEL statement asks SAS to print the estimates for the fixed effects in the output. Also, the ODDSRATIO option on the MODEL statement asks SAS to print the odds ratios (OR) of the predicted value in the output. The DIFF=FIRST option tells SAS to generate odds ratios comparing the first level of the predictor variables.

The example codes above are for random intercept only models, therefore, only the intercept is specified on the RANDOM statement. Please note that if you wanted to allow level-1 variables to vary between level-2 units (i.e., random slopes), you would specify this here by also including any or all of your level-1 variables on the RANDOM statement. The SUBJECT=OPTION on the RANDOM statement corresponds with the CLASS statement. 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 GLIMMIX is beyond the scope of this paper, in the examples above and throughout the examples used in this paper, we use the TYPE=VC option, the default in SAS. 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. 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.

The COVTEST statement is required to generate hypothesis tests for the variance and covariance components (i.e., the random effects). In our examples, we have specified the WALD option, which tells SAS to produce Wald Z tests for the covariance parameters.

DATA SOURCE FOR ILLUSTRATED EXAMPLES

The data for the dichotomous and polytomous logit model examples originate from High School and Beyond (HS&B; n.d.), a nationally-representative dataset in which students are nested within schools. The HS&B study is part of the National Education Longitudinal Studies (NELS) program of the U.S. Department of Education, National Center for Education Statistics (NCES; www.nces.gov), and was established to investigate the life trajectories of the nation’s youth, starting with their high school years and following them into adulthood. HS&B data is publicly-available on the NCES website, and consists of survey results from two cohorts, surveyed bi-annually.

EXAMPLE 1: ESTIMATING TWO-LEVEL ORGANIZATIONAL MODELS WITH DICHOTOMOUS OUTCOMES

The use of PROC GLIMMIX will be demonstrated on a hierarchical dataset with students nested in schools. Our sample comes from the aforementioned HS&B dataset and includes 7185 students nested within 160 schools. To illustrate HGLM with dichotomous outcomes, we use three student level variables and two school level variables. Specifically, to represent student's socioeconomic status (SES), we used a standardized score calculated as a composite of family poverty, parental education, and parental occupation. In addition, two binary variables were used to represent students' minority status (minority = 1) and students' gender (female = 1). Also, for the purpose of this study, we created a dichotomous outcome (i.e., fail = 0, pass = 1) based on a math achievement scaled score ranging from -2.83 to 24.99. Lastly, school level variables used in this study include school SES, calculated as a mean standardized score based on the same variables as the student level measure, and school sector, based on whether the school was classified as public or Catholic (Catholic = 1).

The primary interest of the current study is to investigate the impact of certain student- and school-level variables on a student's likelihood of passing a math achievement test. Therefore, using student level data (level-1) and school level data (level-2), we build two-level hierarchical models to investigate the relationship between math achievement and predictor variables at both levels. The specific research questions examined in this example include:

1. What is the pass rate of the math test for students at a typical school?
2. Do pass rates of the math test vary across schools?
3. What is the relationship between student SES and the likelihood of passing the math test while controlling for student and school characteristics?
4. What is the relationship between school sector and the likelihood of passing the math test while controlling for student and school characteristics?

The model building process begins with the empty, unconditional model with no predictors. This model provides an overall estimate of the pass rate for students at a typical school, as well as providing information about the variability in pass rates between schools. The PROC GLIMMIX syntax for the unconditional model is shown below.

```
PROC GLIMMIX DATA=SESUG2014 METHOD=LAPLACE NOCLPRINT;
CLASS SCHOOL_ID;
MODEL MATHPASS (EVENT=LAST)=/CL DIST=BINARY LINK=LOGIT SOLUTION;
RANDOM INTERCEPT / SUBJECT=SCHOOL_ID S CL TYPE=VC;
COVTEST /WALD;
```

The SAS PROC GLIMMIX syntax generated the Solutions for Fixed Effects table shown below. The estimate provided in this table (-0.6459) represents the log odds of passing the math test at a typical school (i.e., a school where $u_{0j} = 0$) and helps us answer the first research question addressed in this study.

Solutions for Fixed Effects								
Effect	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
Intercept	-0.6459	0.06818	159	-9.47	<.0001	0.05	-0.7805	-0.5112

To be more meaningful, these initial results could be transformed into predicted probabilities (PP). This transformation can be computed using the formula previously provided in Equation 4. To exemplify, the probabilities of success/failure at a typical school from our sample are calculated below.

$$p_{success} = \phi_{ij} = \frac{e^{\eta_{ij}}}{1 + e^{\eta_{ij}}} = \frac{e^{-0.6459}}{1 + e^{-0.6459}} = .344$$

$$p_{failure} = 1 - \phi_{ij} = 1 - .344 = .656$$

The PROC GLIMMIX syntax also provides a Covariance Parameter Estimates table, as 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 the probability of passing the math test is accounted for by the schools. In HGLMs, there is assumed to be no error at level-1, therefore, a slight modification is needed to calculate the ICC. This modification assumes the dichotomous outcome comes from an unknown latent continuous variable with a level-1 residual that follows a logistic distribution with a mean of 0 and a variance of 3.29 (Snijders & Bosker, 1999 as cited in O'Connell et al., 2008). Therefore, 3.29 will be used as our level-1 error variance in calculating the ICC.

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	SCHOOL_ID	0.6165	0.08870	6.95	<.0001

$$ICC = \frac{\tau_{00}}{\tau_{00} + 3.29}$$

$$ICC = \frac{0.6165}{0.6165 + 3.29} = 0.1578$$

This indicates that approximately 16% of the variability in the pass rate for the math test is accounted for by the schools in our study, leaving 84% of the variability to be accounted for by the students or other unknown factors. The output also indicates that there is a statistically significant amount of variability in the log odds of passing the math test between the schools in our sample [$\tau_{00} = .6165$; $z(159) = 6.95$, $p < .0001$], providing response to our second research question. In summary, the unconditional model results revealed that the probability of passing the math exam at a typical school is .344. However, the probability of passing (or the pass rate), varies considerably across schools.

PROC GLIMMIX also produces output containing model fit information for all models estimated. For illustration purposes, partial output containing the -2LL value for the unconditional model is shown below.

Fit Statistics	
-2 Log Likelihood	8906.74

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. More specifically, our next step in the model building process was to include students' SES, Minority status (minority=1), and Gender (female=1), as fixed effects in a level-1 model with random intercept only (Model 2). The PROC GLIMMIX 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 SES, MINORITY, and FEMALE included on the MODEL statement as well as the ODDS RATIO and (DIFF = FIRST LABEL).

```
PROC GLIMMIX DATA=LIB.SESUG2014 METHOD=LAPLACE NOCLPRINT;
CLASS SCHOOL_ID;
MODEL MATHPASS (EVENT=LAST) = SES MINORITY FEMALE / CL DIST=BINARY LINK=LOGIT SOLUTION
ODDSRATIO (DIFF=FIRST LABEL);
RANDOM INTERCEPT / SUBJECT=SCHOOL_ID TYPE=VC SOLUTION CL;
COVTEST / WALD;
```

Partial Solutions from the Fixed Effects table generated by this PROC GLIMMIX code is shown below. Note that due to space limitation we have not included the portion of the output that contains the ORs for each predictor.

Solutions for Fixed Effects								
Effect	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
Intercept	-0.2537	0.06124	159	-4.14	<.0001	0.05	-0.3746	-0.1327
SES	0.6431	0.04139	7022	15.54	<.0001	0.05	0.5620	0.7243
MINORITY	-0.8753	0.08009	7022	-10.93	<.0001	0.05	-1.0324	-0.7183
FEMALE	-0.3739	0.06013	7022	-6.22	<.0001	0.05	-0.4918	-0.2561

Next, In order to address the impact of the school variables, mean school SES and School Sector (Catholic=1), we continued the model building process by adding the school level variables to our model. The PROC GLIMMIX syntax and fixed effects output for this level-1 and level-2 model with random intercept only is shown below. Again, we have not included the OR portion of the output due to space limitations.

```
PROC GLIMMIX DATA=LIB.SESUG2014 METHOD=LAPLACE NOCLPRINT;
CLASS SCHOOL_ID;
MODEL MATHPASS (EVENT=LAST) = SES MINORITY FEMALE MEANSES SECTOR / CL DIST=BINARY
LINK=LOGIT SOLUTION ODDS RATIO (DIFF=FIRST LABEL);
RANDOM INTERCEPT / SUBJECT=SCHOOL_ID TYPE=VC SOLUTION CL;
COVTEST / WALD;
```

Solutions for Fixed Effects								
Effect	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
Intercept	-0.4567	0.06553	157	-6.97	<.0001	0.05	-0.5862	-0.3273
SES	0.5675	0.04236	7022	13.40	<.0001	0.05	0.4844	0.6505
MINORITY	-0.8361	0.07853	7022	-10.65	<.0001	0.05	-0.9901	-0.6822
FEMALE	-0.3833	0.05887	7022	-6.51	<.0001	0.05	-0.4987	-0.2679
MEANSES	0.6010	0.1188	7022	5.06	<.0001	0.05	0.3681	0.8340
SECTOR	0.4197	0.08787	7022	4.78	<.0001	0.05	0.2475	0.5920

Now that we have estimated all of our models, we have compiled the output into a single summary table – see Table 2. We are now ready to evaluate the models and select the best fitting model by conducting a deviance test. This is a chi-square difference test in which we will compare the difference in the -2LL values between two models, with degrees of freedom equal to the difference in the number of parameters estimated between the models. A deviance test can only be conducted when models are nested; therefore, we will first compare Model 1 to Model 2 and then Model 2 to Model 3 to assess model fit.

The calculations for conducting the deviance test between Model 1 and Model 2 is provided below.

$$X_{diff}^2 = -2LL_{Model\ 1} - -2LL_{Model\ 2}$$

$$X_{diff}^2 = 8906.74 - 8432.39 = 474.35$$

After determining that Model 2 was indeed a better fitting model than Model 1, next we compared Model 3 to Model 2. Thus, we were examining if the addition of school level variables improved model fit. Through this process, we determined that Model 3, a model containing both level-1 and level-2 fixed effects was the best fitting model, therefore we will use the parameter estimates generated by this model to answer our remaining research questions.

Table 2. Estimates for Two-level Generalized Linear Dichotomous Models of Math Achievement (N=7,185)

	Model 1	Model 2	Model 3 ^a
<i>Fixed Effects</i>			
Intercept	-0.65* (0.07)	-0.25* (0.06)	-0.46* (0.06)
Student SES		0.64* (0.04)	0.57* (0.04)
Minority Status		-0.87* (0.08)	-0.84* (0.08)
Gender		-0.37* (0.06)	-0.38* (0.06)
Mean School SES			0.60* (0.12)
Sector			0.42* (0.09)
<i>Error Variance</i>			
Level-2 Intercept	0.62* (0.09)	0.26* (0.05)	0.14* (0.03)
<i>Model Fit</i>			
-2LL	8906.74	8432.39**	8368.59**

Note: *p< .05; **=likelihood ratio test significant; ICC = .158; Values based on SAS PROC GLIMMIX. Entries show parameter estimates with standard errors in parentheses; Estimation Method = Laplace.

^aBest fitting model

Our third research question asks about the relationship between student SES and the likelihood of passing the math test, controlling for other student and school level characteristics. To answer this question, we examine the parameter estimate for Student SES generated by Model 3. This relationship is both significant and positive ($b = 0.57, p < .05$), indicating that as student SES increases, the predicted log odds of the student passing the math test also increases. For a more intuitive interpretation, it is most common to report the ORs and/or PP when describing the relationship between predictors and the outcome. From the OR output (details not shown) it appears that for every one unit increase in student SES the odds of passing the exam increase by 1.76. That is, for students with personal SES one standard deviation above the mean, their odds of passing the math test is 1.76 times greater than students with average SES. Or, using the same equations described with the first model, we can also calculate the PP of passing the exam for students with different SES levels.

$$p_{success,SES=0} = \frac{e^{\eta_{ij}}}{1 + e^{\eta_{ij}}} = \frac{0.633}{1 + 0.633} = .388$$

$$p_{success,SES=1} = \frac{e^{\eta_{ij}}}{1 + e^{\eta_{ij}}} = \frac{1.12}{1 + 1.12} = .528$$

$$p_{success,SES=2} = \frac{e^{\eta_{ij}}}{1 + e^{\eta_{ij}}} = \frac{1.97}{1 + 1.97} = .663$$

It is important to remember that all of the output, whether it is the raw regression coefficients, ORs, or PP, is interpreted based on all variables in the model being equal to 0. For example, the PP for a student with average SES, who is both male and White, and attends a public school with average school-wide SES is .388 whereas the PP for a white male student who attends a public school with average school-wide SES but who has a slightly above average personal SES (e.g., SES = 1) is .528.

Our fourth research question asked about the relationship between school sector (public versus Catholic school), and the likelihood of passing the math test. To answer this question, we examine the fixed effect coefficient for Sector in Table 2. This coefficient is both significant and positive ($b = 0.42$, $p < .05$), indicating that students who attend Catholic school (i.e., where Sector = 1) have a higher predicted log odds of passing the math test compared to students who attend public school. Again, to provide more meaningful results, one would examine the OR output and/or calculate the PP for this relationship. Using the same PP formulas as before results indicate that holding all other student and school level variables constant, the predicted probability that a student attending a Catholic school will pass the math test is .491, compared to .388 for a student who attends public school, controlling for all other student and school level variables in the model.

EXAMPLE 2: ESTIMATING TWO-LEVEL ORGANIZATIONAL MODELS WITH POLYTOMOUS OUTCOMES

In order to exemplify the use of PROC GLIMMIX in estimating two-level organizational models with polytomous outcomes, we created a polytomous version of the math achievement score, by splitting the sample in three categories (i.e., 1 = below basic, 2 = basic, 3 = proficient). The purpose of this example is to investigate the probability of students being at or below a proficiency level and similar with the dichotomous model, the influence of student and school characteristics on this probability for each category. However, for simplicity, we only explored the impact of one student-level variable (SES) and one school-level variable (MEANSES) on the likelihood of students being at or below a math proficiency level.

More specifically, the research questions examined in this example include:

1. What is the likelihood of being at or below each proficiency level in math for students at a typical school?
2. Does this likelihood of being at or below each math proficiency level vary across schools?
3. What is the relationship between students' socioeconomic status and their likelihood of being at or below a proficiency level in math?
4. What is the relationship between schools' socioeconomic status and the likelihood of students being at or below a proficiency level in math?

Similar with the dichotomous example, the model building process started by examining the unconditional model with no predictors (Model 1), followed by the level-1 model including student SES (Model 2), and lastly, by the combined level-1 and level-2 model which added school-level SES (Model 3). All models estimated were random intercept only, so similar with the dichotomous example, student- and school-level characteristics were included in the models as fixed effects. For brevity, in this section we will mostly focus on highlighting the differences between the dichotomous and the polytomous models in terms of syntax, output, and interpretation.

The PROC GLIMMIX syntax for the unconditional model estimated for this example is presented below. The key difference between this syntax and the syntax used to estimate the unconditional dichotomous model consists in the options used to specify the link and the distribution type on the MODEL statement. More specifically, this example uses MULTI as the DIST option and CLOGIT as the LINK option to appropriately specify a multinomial distribution and a cumulative logit link necessary for the polytomous model.

```
PROC GLIMMIX DATA=LIB.POLYTOMOUS METHOD=LAPLACE NOCLPRINT;
CLASS SCHOOLID;
MODEL MATHORDINAL= / DIST=MULTI LINK=CLOGIT SOLUTION CL;
RANDOM INTERCEPT/ SUBJECT=SCHOOLID TYPE=VC SOLUTION CL;
COVTEST/WALD;
```

The major difference in terms of the output generated for this unconditional model consists in the Solution for Fixed Effects table. Compared to the output generated for the dichotomous example where there is only one intercept estimated, the output below contains estimates for two intercepts. These are simultaneously estimated and in this

case, represent the log odds of being at or below the first two math proficiency levels (i.e., below basic and basic) of students at a typical school.

These fixed effects estimates are useful for answering our first research question regarding the likelihood of being at or below each proficiency level in math for students at a typical school. More specifically, these log odds can be used to calculate the PP of being at or below each proficiency level in math, using the same formula as in the dichotomous example. The only difference is that now, the formula is applied for each of the intercept values. As results indicate, the log odds of being at or below the below basic math proficiency level at a typical school is -0.4244, resulting in PP of 0.3955. Similarly, the log odds of being at or below the basic math proficiency level is 0.9902, resulting in a cumulative probability of 0.7291. Finally, the cumulative probability of being at or below the proficient level in math adds to 1. Notice that in the polytomous case, these are cumulative probabilities. To calculate the actual probabilities of being at each proficiency level, we need to take a step further and subtract the cumulative probabilities of adjacent categories from one another. As a result, the PP of being at the below basic math proficiency level at a typical school is .3955, at the basic level is .3336 (.7291 - .3955), and at the proficiency level is .2709 (1 - .7291).

Solutions for Fixed Effects									
Effect	MATHORDINAL	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
Intercept	1	-0.4244	0.06829	159	-6.22	<.0001	0.05	-0.5593	-0.2896
Intercept	2	0.9902	0.06916	159	14.32	<.0001	0.05	0.8536	1.1268

Another important piece of the output generated for this model is the Covariance Parameter Estimates table shown below. Similar with the dichotomous example, results presented in this table help us answer the second research question of our study regarding the variability across schools of the likelihood of being at or below each math proficiency level. Results indicate that these relationships significantly vary across schools [$\tau_{00} = .6334$, $z(159) = 7.50$, $p < .0001$]. Also, using the intercept variance estimate provided in this table (.6334), the ICC for this model could be calculated in the same way as described for the dichotomous example. However, for brevity, the ICC calculations are not included in this section.

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	SCHOOLID	0.6334	0.08449	7.50	<.0001

Next, the syntax for estimating the level-1 model and its corresponding output for fixed effect estimates are presented below. As with the dichotomous example, the level-1 model uses the same code as previously used for the unconditional model, with the addition of student SES as a fixed effect on the MODEL statement. However, in this case, the SES estimate illustrates to relationship between this student characteristic and the log odds of being at or below a math proficiency level. To be noted in the output that although there are two estimates for the intercepts, only one slope associated with student SES is estimated for this model. This suggests that the value of the SES estimate remains constant across logits/ intercepts. Also, as shown in the syntax, the user can request ORs as part of the output. But, again, space limitations prevent us from including that portion of the output.

```
PROC GLIMMIX DATA=LIB.POLYTOMOUS METHOD=LAPLACE NOCLPRINT;
CLASS SCHOOLID;
MODEL MATHORDINAL=SES/ DIST=MULTI LINK=CLOGIT SOLUTION CL ODDSATIO(DIFF=FIRST LABEL);
RANDOM INTERCEPT/ SUBJECT=SCHOOLID TYPE=VC SOLUTION CL;
COVTEST/WALD;
```

Solutions for Fixed Effects									
Effect	MATHORDINAL	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
Intercept	1	-0.4447	0.05392	159	-8.25	<.0001	0.05	-0.5512	-0.3382
Intercept	2	1.0249	0.05509	159	18.60	<.0001	0.05	0.9161	1.1337
SES		-0.7152	0.03535	7023	-20.23	<.0001	0.05	-0.7845	-0.6459

Last, the syntax for estimating the combined level-1 and level-2 model would be identical to the syntax for Model 2 with the addition of MEANSES on the MODEL statement. The output generated for this model is not included; however, similar with the output presented for the level-1 model, the output generated for this combined model only includes one slope estimate for MEANSES, thus being constant across both logits.

```
PROC GLIMMIX DATA=LIB.POLYTOMOUS METHOD=LAPLACE NOCLPRINT;
CLASS SCHOOLID;
MODEL MATHORDINAL=SES MEANSES/ DIST=MULTI LINK=CLOGIT SOLUTION CL ODDSRATIO
(DIFF=FIRST LABEL);
RANDOM INTERCEPT/ SUBJECT=SCHOOLID TYPE=VC SOLUTION CL;
COVTEST/WALD;
```

Table 3 presents a summary of the results obtained for this example, including estimates for all three models considered in the model building process as well as model fit information. As with the dichotomous example, the three models are compared in terms of fit in order to decide on the best fitting model for these data. Specifically, based on the changes in the -2LL between nested models, the level-1 model (Model 2) appears to fit significantly better than the unconditional model (Model 1) ($\chi^2(1) = 419.89, p < .0001$). Similarly, the combined level-1 and level-2 model (Model 3) fits significantly better than the level-1 model (Model 2) ($\chi^2(1) = 74.55, p < .0001$), therefore, this appears to be the best fitting model for our data.

Table 3. Estimates for Two-level Generalized Linear Polytomous Models of Math Achievement (N = 7,185)

	Model 1	Model 2	Model 3 ^a
<i>Fixed Effects</i>			
Intercept 1 (Below Basic)	-0.42* (0.07)	-0.44* (0.05)	-0.44* (0.04)
Intercept 2 (Basic)	0.99* (0.07)	1.02* (0.06)	1.03* (0.05)
Student SES		-0.72* (0.04)	-0.64* (0.04)
Mean School SES			-1.03* (0.11)
<i>Error Variance</i>			
Intercept	0.63* (0.08)	0.35* (0.05)	0.20* (0.03)
<i>Model Fit</i>			
-2LL	14890.88	14470.99**	14396.44**

Note: * $p < .05$; **=likelihood ratio test significant; ICC = .161; Values based on SAS PROC GLIMMIX. Entries show parameter estimates with standard errors in parentheses; Estimation Method = Laplace.

^aBest fitting model

Now, similar with the dichotomous example, parameter estimates from the best fitting model (Model 3), will be used to answer our final research questions. More specifically, our third research question refers to the relationship between students' socioeconomic status and their likelihood of being at or below a proficiency level in math. To answer this question, we use the parameter estimate for student SES ($b = -0.64, p < .05$) indicating that there is a negative, statistically significant relationship between students' SES and their probability of being at or below a math proficiency level. More specifically, as students' SES increases, their likelihood of being at the lowest math performance level (i.e., below basic) decreases. Similarly, to answer our last research question, we use the estimate for school SES ($b = -1.03, p < .05$) which also indicates a negative, statistically significant relationship between school average SES and the probability of students being at or below a math proficiency level. Specifically, as school average SES increases, the likelihood of students being at the below basic math level decreases. To make more meaningful interpretations, the ORs portion of the output could be consulted and/or the log odds could be transformed into PP as shown for the dichotomous example.

CONCLUSION

This paper provides researchers with an introduction to estimating hierarchical generalized linear models with both dichotomous and polytomous outcomes via PROC GLIMMIX. Focused on relatively simple and straightforward models to exemplify the use of the GLIMMIX procedure, this paper also presents key elements and issues to consider when dealing with these models. These include but are not limited to the distribution of errors and link function appropriate for both dichotomous and polytomous cases, estimation methods preferred, model building, appropriate evaluation of model fit, and interpretation of GLIMMIX output. Although more advanced aspects regarding these models (e.g., random slopes, cross-level interactions, various estimation methods, options for more complex or more stable variance structures) are not included in this paper, we truly hope that applied researchers who are new to multilevel modeling and in particular to hierarchical generalized linear models, will find this introduction helpful!

REFERENCES

- Bell, B.A., Ene, M., Smiley, W., & Shonenberger, J.A. (2013). A multilevel primer using SAS[®] PROC MIXED. SAS Global Forum 2013 Proceedings. <http://support.sas.com/resources/papers/proceedings13/433-2013.pdf>
- Donner, A. & Klar, N. (2000). *Design and analysis of cluster randomization trials in health research*. London: Edward Arnold.
- 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.
- High School and Beyond [HS&B]*. (n.d.). Retrieved from the National Center for Educational Statistics website, <http://nces.ed.gov/surveys/hsb/>
- Hox, J.J. (2002). *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
- Klein, K.J. & Kozlowski, S.W. (2000). *Multilevel theory, research, and methods in organizations: Foundations, extensions and new directions*. San Francisco: Jossey-Bass.
- Luke, D.A. (2004). *Multilevel modeling*. Thousand Oaks, CA: Sage.
- 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., Goldstein, J., Rogers, H.J., & Peng, C.Y.J. (2008). Multilevel logistic models for dichotomous and ordinal data. In A.A. O'Connell & D.B. McCoach (Eds.), *Multilevel modeling of educational data* (199-242). Charlotte, NC: Information Age Publishing, Inc.
- 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 last author at:

Bethany A. Bell
University of South Carolina
Work Phone: 803-777-2387
E-mail: babell@sc.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.