

An Intermediate Guide to Estimating Multilevel Models for Categorical Data using SAS® PROC GLIMMIX

Whitney Smiley, Elizabeth Leighton, Zhaoxia Guo, Mihaela Ene, and Bethany A. Bell
University of South Carolina

ABSTRACT

This paper expands upon Ene et al.'s (2015) SAS Global Forum proceeding paper "Multilevel Models for Categorical Data using SAS® PROC GLIMMIX: The Basics" in which the authors presented an overview of estimating two-level models with non-normal outcomes via PROC GLIMMIX. In their paper, the authors focused on how to use GLIMMIX to estimate two-level organizational models; however, they did not include random slopes, address more complex organizational models (e.g., three-level models) or models used to estimate longitudinal data. Hence, the need for the current paper; by building from the examples in Ene et al. (2015), the current paper presents users detailed discussions and illustrations about how to use GLIMMIX to estimate models with random slopes, organizational models in situations with three levels of data, as well as two-level longitudinal data. Consistent with Ene et al.'s paper, we will present the syntax and interpretation of the estimates using a model with a dichotomous outcome as well as a model with a polytomous outcome. Concrete examples will be used to illustrate how PROC GLIMMIX can be used to estimate these models and how key pieces of the output can be used to answer corresponding research questions.

Keywords: MULTILEVEL MODELING, PROC GLIMMIX, GROWTH MODELING, THREE LEVEL MODELS

INTRODUCTION

At the 2015 SAS Global Forum in Dallas, TX, Ene et al. presented the logic behind multilevel models as well as some basic demonstrations on how to use PROC GLIMMIX to estimate two-level organizational models with non-normal outcome data. Although their paper was quite informative for the beginning PROC GLIMMIX user, their examples did not include details on how to use GLIMMIX to include random slopes, estimate longitudinal data (e.g., children with achievement scores across a school year) or data that has more than two levels (e.g., children nested within classrooms nested within schools). Hence, the need for the current paper; by building from the examples in Ene et al. (2015), the current paper presents users detailed discussions and illustrations about how to use GLIMMIX to estimate models in situations with multiple levels or longitudinal data. In order to get the most out of this paper, it is recommended that the reader have prerequisite knowledge regarding the mathematical equations used to estimate the models, including knowledge of the link function, Bernoulli distribution, odds ratios, and predicted probabilities. If the reader does not have knowledge of these concepts, or needs a refresher, it is recommended that the reader starts with the Ene et al. (2015) SAS Global Forum paper prior to reading this paper.

MODEL BUILDING

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). Note that this table is for 3-level models, however if building a two-level model is of interest, the researcher can follow the same table, but stop at model 4. 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 this table are general in nature. Sometimes intermediate models are also examined depending on the model fit indices (e.g., 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). Note that all examples in this paper follow this model building process in order to build a final table with all relevant estimates.

In order to find the best fitting model, we examine improvement in model fit by using Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) as we examine improvement in model fit. 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. 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.

Lastly, when building models we also deemed random effects as significant or not by examining the p-value for the Wald distribution using the covtest option. Note that testing variance components is more complicated than simply examining the p-value of the Wald's test obtained by using the covtest option. For brevity, we used the Wald test p-value as our sample size was large and the p-value associated with the Wald test has been shown to be correct when sample sizes are large (Singer, 1998). For more information on smaller sample sizes are other ways to examine significance for variance components, the reader is encouraged to read Singer (1998) and Bell, Smiley, Ene, and Blue (2014).

Table 1. Model Building Process for 3-Level Generalized Linear Models with Random Intercepts and Random Slopes

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Variables entered into the model	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
Relevant information from output	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. The 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.

DATA SOURCES

For the examples involving three-level organizational logit models with dichotomous and polytomous outcomes, we used a non-publicly available school district data set where students are nested within teachers and teachers 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 categorical measure of math achievement, English proficiency status (LEP, 0 = not proficient, 1 = proficient), and exceptional child status (EC, 0 = not exceptional, 1 = exceptional). The teacher variable is the number of years of teaching experience (TCHEXP), which was grand mean centered. At the school level, there are two variables, the first of which is a dummy coded variable indicating participation in a school-level intervention (TREATMENT, 0 = control, 1 = treatment) and the second is the percent of students at the school that receive free and reduced lunch (SCHL_FRL), which was grand mean centered. In the original data set, math achievement was continuous. However, in order to create a categorical outcome, two variables were created to create a non-linear outcome, including score_d which had a dichotomous outcome, and score_p which had a polytomous outcome. Note that all continuous predictors without a meaningful zero have been grand mean centered.

The two growth model examples (examples 3 and 4) were estimated using the Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999 (ECLS-K) data set, which is publicly available on the 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). For the purposes of our growth model examples, 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). Students were assessed for reading achievement at four time points during their early education experience: twice in kindergarten (fall and spring) and twice in first grade (fall and spring). For this study, time was recorded in months, specifically months since starting kindergarten. A time of "0" was set to the fall of their kindergarten year, followed by 8 months in the spring of their kindergarten year, 12 months in the fall of their first grade year, and 20 months in spring of their first grade year. Variables used in the examples include the student's biological sex (SEX, 0 = male, 1 = female), time measured in months (TIMEMOS), and a measure of reading achievement. The reading achievement measure was

originally continuous, however, for the purposes of this paper, two new variables were created to create a non-linear outcome, including SCORE_D which was a dichotomous outcome and SCORE_P which was a polytomous outcome. Note that all continuous predictors without a meaningful zero have been grand mean centered.

EXAMPLE 1: THREE-LEVEL ORGANIZATIONAL MODELS WITH DICHOTOMOUS OUTCOMES

The use of PROC GLIMMIX will be demonstrated on the three-level non-publicly available data set mentioned previously. The primary interest of the current study is to investigate the impact of certain student-, teacher- and school-level variables on a student's likelihood of passing a math achievement test. Therefore, using student level data (level-1), teacher level data (level-2) and school level data (level-3), we build three-level hierarchical models to investigate the relationship between math achievement and predictor variables at all levels. The specific research questions examined in this example include:

1. How much variance in math test pass rates is attributable to teachers and schools? What is the pass rate for students with a typical teacher at a typical school?
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 teacher experience, controlling for student characteristics?
4. What is the impact of the school-wide intervention program, controlling for student, teacher, and school characteristics?

As shown in Table 1, 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 taught by a typical teacher at a typical school, as well as providing information about the variability in pass rates between teachers within schools, the variability among schools, and the variability within schools. The PROC GLIMMIX syntax for the unconditional model is shown below. The key difference between this model and a model that would only have two levels is the use of two RANDOM statements. The first RANDOM statement shows that schools are allowed to have their own intercepts, and that schools are the highest level of the hierarchy. The second RANDOM statement specified that each teacher is allowed to have their own intercept and that the teachers are nested within schools [i.e., / SUBJECT=TEACHERID(SCHOOLID)].

```
PROC GLIMMIX DATA=SESUG2015 METHOD=LAPLACE NOCLPRINT;
CLASS TEACHERID SCHOOLID;
MODEL SCORE_D(EVENT=LAST)=/CL DIST=BINARY LINK=LOGIT SOLUTION ;
RANDOM INTERCEPT / SUBJECT=SCHOOLID TYPE=VC;
RANDOM INTERCEPT / SUBJECT=TEACHERID(SCHOOLID) TYPE=VC;
COVTEST / WALD;
```

The SAS PROC GLIMMIX syntax generated the Solutions for Fixed Effects table shown below. The estimate provided in this table (-0.1759) represents the log odds of passing the math test for students taught by a typical teacher at a typical school and helps us answer part of 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	-0.1759	0.08933	139	-1.97	0.0509	0.05	-0.3526	0.000681

To be more meaningful, these initial results could be transformed into predicted probabilities (PP). This transformation can be computed using the formulas presented below. In conclusion, the probability of success for a student with a typical teacher in a typical school is 0.456.

$$p_{\text{success}} = \phi_{ij} = \frac{e^{n_{ij}}}{1 + e^{n_{ij}}} = \frac{e^{-0.1759}}{1 + e^{-0.1759}} = 0.456 \quad (\text{Eq. 1})$$

$$p_{\text{failure}} = 1 - \phi_{ij} = 0.544 \quad (\text{Eq. 2})$$

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 teachers and schools. As mentioned in the basic paper by Ene et al. (2015), there is assumed to be no error at level-1, therefore, a slight modification to calculate the ICC is needed. 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	schoolid	0.9087	0.1359	6.69	<.0001
Intercept	teacherid(schoolid)	0.6354	0.04043	15.72	<.0001

$$ICC_T = \frac{\tau_T}{\tau_T + \tau_S + 3.29} = \frac{0.6354}{0.6354 + 0.9087 + 3.29} = 0.1314 \quad (\text{Eq. 3})$$

$$ICC_S = \frac{\tau_S}{\tau_T + \tau_S + 3.29} = \frac{0.9087}{0.6354 + 0.9087 + 3.29} = 0.1880 \quad (\text{Eq. 4})$$

This indicates that approximately 13% of the variability in the pass rate for the math test is accounted for by the teachers within schools, and 19% of the variability is accounted for by the schools, leaving 78% of the variability to be accounted for by the students [i.e., $1 - (.1314 + .1880)$]. Based on the p -values for each of the variance estimates, we can also state that the variability between teachers and between schools is statistically significant. Note that this is not a definitive way to check for model fit. PROC GLIMMIX also produces output containing model fit information for all models estimated, including AIC and BIC which will be recorded in the summary table below (Table 2).

Next, we build Model 2 as mentioned in Table 1, by entering our level-1 predictors (LEP and EC) as fixed effects on the model statement. No changes are made to either of the RANDOM statements. The fixed effects information generated from the following syntax is shown below. Here we see that both LEP and EC are significant predictors of student math test success.

```
PROC GLIMMIX DATA = SESUG2015 METHOD=LAPLACE NOCLPRINT;
CLASS TEACHERID SCHOOLID;
MODEL SCORE_D(EVENT=LAST)=LEP EC/CL DIST=BINARY LINK=LOGIT SOLUTION ODDSRATIO
(DIFF=FIRST LABEL);
RANDOM INTERCEPT / SUBJECT=SCHOOLID TYPE=VC;
RANDOM INTERCEPT / SUBJECT=TEACHERID(SCHOOLID) TYPE=VC;
COVTEST / WALD;
```

Solutions for Fixed Effects								
Effect	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
Intercept	-0.01631	0.08756	139	-0.19	0.8525	0.05	-0.1894	0.1568
lep	-0.1414	0.02401	72279	-5.89	<.0001	0.05	-0.1884	-0.09430
ec	-1.0566	0.03361	72279	-31.44	<.0001	0.05	-1.1225	-0.9908

In order to answer question 2, we run Model 3 in Table 1 by allowing the level-1 variables to vary across level-2 and level-3. This can be accomplished by adding LEP and EC to both of the random statements. The PROC GLIMMIX syntax for this model is displayed below.

```
PROC GLIMMIX DATA = SESUG2015 METHOD=LAPLACE NOCLPRINT;
CLASS TEACHERID SCHOOLID;
MODEL SCORE_D(EVENT=LAST)=LEP EC /CL DIST=BINARY LINK=LOGIT SOLUTION ODDSRATIO
(DIFF=FIRST LABEL);
RANDOM INTERCEPT LEP EC/ SUBJECT=SCHOOLID TYPE=VC;
RANDOM INTERCEPT LEP EC/ SUBJECT=TEACHERID(SCHOOLID) TYPE=VC;
COVTEST / WALD;
```

Running this code generated the Covariance Parameter Estimates table below. In this table, we can see that LEP and EC significantly vary at the school and teacher level. Note that as we previously stated, testing variance components can get a little more complicated, but in these illustrations we used this methods for brevity. Again, the reader is encouraged to read Singer (1998) and Bell, Smiley, Ene, and Blue (2014) for more information on this.

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	schoolid	0.9671	0.1426	6.78	<.0001
lep	schoolid	0.1703	0.04785	3.56	0.0002
ec	schoolid	0.2833	0.09073	3.12	0.0009
Intercept	teacherid(schoolid)	0.5986	0.03943	15.18	<.0001
lep	teacherid(schoolid)	0.1456	0.03100	4.70	<.0001
ec	teacherid(schoolid)	0.2433	0.06226	3.91	<.0001

Once level-1 variables are allowed to vary, we add the fixed effects for level-2 (Model 4 in Table 1). In order to do this, TCHEXP is added to the model statement as shown below. The syntax produced the fixed effects table shown below it.

```
PROC GLIMMIX DATA = SESUG2015 METHOD=LAPLACE NOCLPRINT;
CLASS TEACHERID SCHOOLID;
MODEL SCORE_D(EVENT=LAST)=LEP EC TCHEXP/CL DIST=BINARY LINK=LOGIT SOLUTION ODDSRATIO
(DIFF=FIRST LABEL);
RANDOM INTERCEPT LEP EC/ SUBJECT=SCHOOLID TYPE=VC;
RANDOM INTERCEPT LEP EC/ SUBJECT=TEACHERID(SCHOOLID) TYPE=VC;
COVTEST / WALD;
```

Solutions for Fixed Effects								
Effect	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
Intercept	-0.1768	0.09532	139	-1.85	0.0658	0.05	-0.3652	0.01169
Lep	-0.1944	0.06168	129	-3.15	0.0020	0.05	-0.3165	-0.07240
ec	-1.0274	0.07860	136	-13.07	<.0001	0.05	-1.1828	-0.8720
tchexp	0.02115	0.004132	70442	5.12	<.0001	0.05	0.01305	0.02925

Then we continued the model building process by allowing the level-2 variable to vary across level-3 (Model 5 in Table 1), using the following syntax. Below the syntax is the Covariance Parameter Estimates produced by the syntax. In this box, the p-value for teacher experience is not significant, and thus we do not use this variable as a random effect moving forward. Note that the output for the odds ratio has been truncated for brevity.

```
PROC GLIMMIX DATA = SESUG2015 METHOD=LAPLACE NOCLPRINT;
CLASS TEACHERID SCHOOLID;
MODEL SCORE_D(EVENT=LAST)=LEP EC TCHEXP/CL DIST=BINARY LINK=LOGIT SOLUTION ODDSRATIO
(DIFF=FIRST LABEL);
RANDOM INTERCEPT LEP EC TCHEXP/ SUBJECT=SCHOOLID TYPE=VC;
RANDOM INTERCEPT LEP EC/ SUBJECT=TEACHERID(SCHOOLID) TYPE=VC;
COVTEST / WALD;
```

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	schoolid	0.7654	0.1047	7.31	<.0001
lep	schoolid	0.1959	0.05952	3.29	0.0005
ec	schoolid	0.2590	0.08245	3.14	0.0008
tchexp	schoolid	0.000423	0.000304	1.39	0.0820
Intercept	teacherid(schoolid)	0.5866	0.03919	14.97	<.0001
lep	teacherid(schoolid)	0.1559	0.03287	4.74	<.0001
ec	teacherid(schoolid)	0.2642	0.06670	3.96	<.0001

Lastly, we add the level-3 fixed effects to the model (Model 6 in Table 1) by adding TREATMENT and SCHL_FRL on the model statement as shown below. The Solutions for Fixed Effects produced by the model is presented below the syntax. Note that the output for the odds ratio has been truncated for brevity.

```

PROC GLIMMIX DATA = SESUG2015 METHOD=LAPLACE NOCLPRINT;
CLASS TEACHERID SCHOOLID;
MODEL SCORE_D(EVENT=LAST)=LEP EC TCHEXP TREATMENT SCHL_FRL/CL DIST=BINARY LINK=LOGIT
SOLUTION ODDSRATIO (DIFF=FIRST LABEL);
RANDOM INTERCEPT LEP EC/ SUBJECT=SCHOOLID TYPE=VC;
RANDOM INTERCEPT LEP EC/ SUBJECT=TEACHERID(SCHOOLID) TYPE=VC;
COVTEST / WALD;

```

Solutions for Fixed Effects								
Effect	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
Intercept	1.2352	0.1163	139	10.62	<.0001	0.05	1.0053	1.4651
Lep	-0.1701	0.06147	129	-2.77	0.0065	0.05	-0.2917	-0.04843
ec	-1.0155	0.07936	136	-12.80	<.0001	0.05	-1.1724	-0.8585
tchexp	0.02534	0.004108	70440	6.17	<.0001	0.05	0.01728	0.03339
treatment	0.2906	0.05844	70440	4.97	<.0001	0.05	0.1761	0.4052
schl_frl	-2.6262	0.1852	70440	-14.18	<.0001	0.05	-2.9892	-2.2632

After estimating all of the models, we compiled the output into a single summary table (Table 2). From this table, we can see that according to the BIC rules, model 6 is our best fitting model. We will use this model to answer the remaining research questions. Our second research question asks about whether any level-1 or level-2 variables vary among schools. For Model 6, we can see that LEP and EC vary across level-2 and level-3. We also found that TCHEXP did not vary across level-3 (i.e., it wasn't significant in Model 5 and thus was not carried on to Model 6). The third research questions asked what the impact of teacher experience was, controlling for student characteristics. In Model 6, we see that the estimate for teacher experience was 0.03 and it was significant. Thus, we can conclude that there is a relationship between teacher experience and the probability of passing the assessment. For a more intuitive interpretation, we can interpret the odds of passing the test, which is 1.026; for a one-unit increase in teacher experience, we expect to see about a 3% increase in the odds of passing the exam. Our last research question asks if the school-wide intervention had an impact on pass rates. To answer this, we look at the estimate for treatment for Model 6. Here, we see that the log odds of the treatment variable is 0.29 and statistically significant. Thus, it appears the treatment did have an impact on pass rates. Specifically, if we examine the odds ratio of treatment (output not provided), we can see that the odds of passing the exam for students in the treatment group are about 34% higher than the odds for students in the control group.

Table 2. Estimates for Three-Level Generalized Linear Dichotomous Models of Math Achievement (N=73,671)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6 ^a
<i>Fixed effects</i>						
Intercept	-0.18 (0.09)	-0.02 (0.09)	-0.03 (0.09)	-0.18 (0.10)	-0.18* (0.09)	1.24* (0.12)
Student LEP		-0.14* (0.02)	-0.20* (0.06)	-0.19* (0.06)	-0.17* (0.06)	-0.17* (0.06)
Student EC		-1.06* (0.03)	-1.02* (0.08)	-1.03* (0.08)	-1.02* (0.08)	-1.02* (0.08)
Teacher experience				0.02* (0.01)	0.02* (0.01)	0.03* (0.01)
Treatment						0.29* (0.06)
School FRL						-2.63* (0.19)
<i>Error Variance</i>						
Level-2 intercept	0.64* (0.04)	0.63* (0.04)	0.60* (0.04)	0.60* (0.03)	0.59* (0.04)	0.60 (0.04)
Level-3 intercept	0.91* (0.14)	0.86* (0.13)	0.97* (0.14)	0.93* (0.14)	0.77* (0.10)	0.20 (0.04)
LEP _T			0.15* (0.03)	0.15* (0.03)	0.15* (0.03)	0.14 (0.03)
EC _T			0.24* (0.06)	0.24* (0.06)	0.25* (0.06)	0.23 (0.06)
LEP _S			0.17* (0.05)	0.17* (0.05)	0.20* (0.06)	0.17 (0.05)
EC _S			0.28* (0.09)	0.28* (0.09)	0.26* (0.08)	0.29 (0.09)
TCHEXP _S					0.00 (0.00)	
<i>Model fit</i>						
AIC	85827.68	84718.57	84269.25	84244.85	84245.95	84127.63
BIC	85836.51**	84733.27**	84295.72**	84274.27**	84278.31**	84162.93**

Note: * $p < .05$; **=significant decrease in BIC; ICC_{Teacher} = 0.131; ICC_{School} = 0.188; Values based on SAS PROC GLIMMIX. Entries show parameter estimates with standard errors in parentheses; Estimation Method = Laplace.

^aBest fitting model

EXAMPLE 2: THREE-LEVEL ORGANIZATIONAL MODELS WITH POLYTOMOUS OUTCOMES

This example will illustrate the use of PROC GLIMMIX in estimating three-level organizational models with polytomous outcomes. This example will use the same data as the three-level dichotomous model but use a different outcome with three categories (i.e., 0 = fail, 1 = pass, 2 = advanced). The purpose of this example is to investigate factors that predict the probability of students being at or below a proficiency level in math. More specifically, the research questions examined in this example include:

1. What is the likelihood of being at or below each proficiency level for a typical student taught by a typical teacher at a typical school?
2. Does this likelihood of being at or below each math proficiency level vary across teachers and across schools?
3. Does the influence of any student-level predictor vary among teachers or schools? Does the influence of the teacher-level predictors vary among schools?
4. What is the impact of teacher experience on a student's likelihood of being at or below a proficiency level in math while controlling for student characteristics?
5. What is the impact of the school-wide intervention program on a student's likelihood of being at or below a proficiency level in math while controlling for student, teacher, and school characteristics?

Similar with the dichotomous example, we used the model building method in Table 1. For brevity, in this section we will mostly focus on highlighting the differences between the dichotomous and the polytomous three-level organizational models in terms of syntax, output, and interpretation. Note that in this section we use a cumulative logit model. Note that before researchers start building models using this method they should test the assumption of equal slopes (O'Connell & McCoach, 2008). For illustrative purposes, this assumption was not checked prior to running these models, but if a researcher were to use our syntax for polytomous models, they would first need to check this assumption. If this assumption does not hold, a multinomial multilevel model would need to be used (O'Connell & McCoach, 2008).

The PROC GLIMMIX syntax for the unconditional model (Model 1) 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=SESUG2015 METHOD=LAPLACE NOCLPRINT;
CLASS TEACHERID SCHOOLID;
MODEL SCORE_P =/CL DIST=MULTI LINK=CLOGIT SOLUTION;
RANDOM INTERCEPT / SUBJECT=SCHOOLID TYPE=VC;
RANDOM INTERCEPT / SUBJECT=TEACHERID (SCHOOLID) TYPE=VC;
COVTEST / WALD;
```

The major difference in terms of the output generated for this unconditional model consists in the Solution for Fixed Effects table shown below. 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., fail and pass) for students with a typical teacher in a typical school.

Solutions for Fixed Effects									
Effect	score_p	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
Intercept	0	0.1818	0.08880	139	2.05	0.0425	0.05	0.006255	0.3574
Intercept	1	1.3112	0.08896	139	14.74	<.0001	0.05	1.1353	1.4871

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 achievement. More specifically, these log odds can be used to calculate the PP of being at or below each math proficiency level, using the same formula presented in the dichotomous example (Eq. 1 and 2). However, now we need to apply this formula to each of the intercept values. As results indicate, the log odds of being at or below the failing math level for students across typical teachers in typical schools is 0.1818, resulting in PP of 0.5453. Similarly, the log odds of being at or below the passing math level is 1.3112, resulting in a cumulative probability of 0.7877. Finally, the cumulative probability of being at or below the advanced 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 failing math level is 0.5453, at the passing level is 0.2424 (0.7877 – 0.5453), and at the advanced level is 0.2123 (1 – 0.7877).

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 of the likelihood of being at or below each math proficiency level across teachers and across schools. Results indicate that this likelihood significantly varies across both teachers [$\tau_{\pi}=0.7040$, $z(139) = 16.86$, $p<.0001$] and schools [$\tau_{\beta}=0.8925$, $z(139) = 6.73$, $p<.0001$]. Also, using both of the intercept variance estimates (0.7040 and 0.8925) provided in this table, the ICCs for this model are calculated in the same way as described for the dichotomous example (Eq. 3 and 4). Again, given our data contains students nested within teachers nested in schools, an ICC is calculated for both teachers and schools. The resulting ICC values of 0.1441 and 0.1826 mean that approximately 14% of the variability in the likelihood of being at or below a certain math proficiency level is accounted for by teachers, and approximately 18% of the variability is accounted for by schools.

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	schoolid	0.8925	0.1326	6.73	<.0001
Intercept	teacherid(schoolid)	0.7040	0.04176	16.86	<.0001

Next, the syntax for estimating the level-1 model (Model 2) 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 the student level variables representing a student's limited English proficiency (LEP) and exceptional child (EC) status as fixed effects on the MODEL statement. In this example, the LEP and EC estimates illustrate the relationship between these student characteristics and the log odds of being at or below a math proficiency level. Also, note in the output that although there are two estimates for the intercepts, only one slope associated with LEP and one slope associated with EC is estimated for this model. This suggests that the values of the LEP and EC estimates remain constant across logits/ intercepts. Also, as shown in the syntax, the user can request ORs as part of the output; however, that portion of the output is not included in this section due to space limitations.

```
PROC GLIMMIX DATA = SESUG2015 METHOD=LAPLACE NOCLPRINT;
CLASS TEACHERID SCHOOLID;
MODEL SCORE_P = LEP EC /CL DIST=MULTI LINK=CLOGIT SOLUTION ODDSRATIO (DIFF=FIRST LABEL);
RANDOM INTERCEPT / SUBJECT=SCHOOLID TYPE=VC;
RANDOM INTERCEPT / SUBJECT=TEACHERID(SCHOOLID) TYPE=VC;
COVTEST / WALD;
```

Solutions for Fixed Effects									
Effect	SCORE_P	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
Intercept	0	0.02069	0.08683	139	0.24	0.8120	0.05	-0.1510	0.1924
Intercept	1	1.1667	0.08696	139	13.42	<.0001	0.05	0.9948	1.3387
LEP		0.1605	0.02260	72278	7.10	<.0001	0.05	0.1162	0.2048
EC		1.0609	0.03289	72278	32.25	<.0001	0.05	0.9965	1.1254

For brevity, Table 3 shows the rest of the syntax needed to continue the model building process. Also, all output looks similar to the output for Model 2, just with more variables. After running all models, Table 4 was constructed to present a summary of the results obtained for this example, including estimates for all six models considered in the model building process as well as model fit information. As with the previous example, the six models are compared in terms of fit in order to decide on the best fitting model for these data. Based on the changes in BIC, Model 6 was deemed the best fitting model and therefore, was used to answer the rest of the research questions.

Table 3. Model Building Syntax for Models 3-6.

Model	Syntax	Changes from previous model
3	<pre>PROC GLIMMIX DATA = SESUG2015 METHOD=LAPLACE NOCLPRINT; CLASS TEACHERID SCHOOLID; MODEL SCORE_P = LEP EC /CL DIST=MULTI LINK=CLOGIT SOLUTION ODDSRATIO (DIFF=FIRST LABEL); RANDOM INTERCEPT LEP EC/ SUBJECT=SCHOOLID TYPE=VC; RANDOM INTERCEPT LEP EC/ SUBJECT=TEACHERID(SCHOOLID) TYPE=VC; COVTEST / WALD;</pre>	Allowed the level-1 variables to vary across level-2 and level-3 by adding LEP and EC on both RANDOM statements
4	<pre>PROC GLIMMIX DATA = SESUG2015 METHOD=LAPLACE NOCLPRINT; CLASS TEACHERID SCHOOLID; MODEL SCORE_P = LEP EC TCHEXP/CL DIST=MULTI LINK=CLOGIT SOLUTION ODDSRATIO (DIFF=FIRST LABEL); RANDOM INTERCEPT LEP EC/ SUBJECT=SCHOOLID TYPE=VC; RANDOM INTERCEPT LEP EC/ SUBJECT=TEACHERID(SCHOOLID) TYPE=VC; COVTEST / WALD;</pre>	Added TCHEXP as a fixed effect by adding it to the MODEL statement
5	<pre>PROC GLIMMIX DATA = SESUG2015 METHOD=LAPLACE NOCLPRINT; CLASS TEACHERID SCHOOLID; MODEL SCORE_P = LEP EC TCHEXP/CL DIST=MULTI LINK=CLOGIT SOLUTION ODDSRATIO (DIFF=FIRST LABEL); RANDOM INTERCEPT LEP EC TCHEXP/ SUBJECT=SCHOOLID TYPE=VC; RANDOM INTERCEPT LEP EC/ SUBJECT=TEACHERID(SCHOOLID) TYPE=VC; COVTEST / WALD;</pre>	Allowed TCHEXP to vary across level-3 by adding it to the RANDOM line for level-3
6	<pre>PROC GLIMMIX DATA = SESUG2015 METHOD=LAPLACE NOCLPRINT; CLASS TEACHERID SCHOOLID; MODEL SCORE_P = LEP EC TCHEXP TREATMENT SCH_FRL/CL DIST=MULTI LINK=CLOGIT SOLUTION ODDSRATIO (DIFF=FIRST LABEL); RANDOM INTERCEPT LEP EC / SUBJECT=SCHOOLID TYPE=VC; RANDOM INTERCEPT LEP EC/ SUBJECT=TEACHERID(SCHOOLID) TYPE=VC; COVTEST / WALD;</pre>	Added level-3 fixed effects by putting TREATMENT and SCH_FRL on the MODEL statement **Note: TCHEXP was not significant on the random line from Model 5, so it was not carried into Model 6

Table 4. Estimates for Three-Level Generalized Linear Polytomous Models of Math Achievement (N=73,671)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6 ^a
<i>Fixed effects</i>						
Intercept 0 (Fail)	0.18* (0.09)	0.02 (0.09)	0.04 (0.09)	0.24* (0.09)	0.23* (0.10)	-1.04* (0.12)
Intercept 1 (Pass)	1.31* (0.09)	1.17* (0.09)	1.21* (0.09)	1.40* (0.10)	1.40* (0.10)	0.12 (0.12)
Student LEP		0.16* (0.02)	0.16* (0.06)	0.17 (0.06)	0.16* (0.06)	0.15 (0.06)
Student EC		1.06* (0.03)	0.99* (0.08)	0.99 (0.08)	0.99* (0.08)	0.98 (0.08)
Teacher experience				-0.03 (0.01)	-0.03* (0.01)	-0.03 (0.00)
Treatment						-0.29 (0.06)
School FRL						2.39 (0.20)
<i>Error Variance</i>						
Level-2 intercept	0.704* (0.04)	0.72* (0.04)	0.67* (0.04)	0.68 (0.04)	0.68* (0.04)	0.68 (0.04)
Level-3 intercept	0.89* (0.13)	0.84* (0.13)	0.96* (0.14)	0.91 (0.13)	0.88* (0.13)	0.22 (0.05)
LEP _T			0.14* (0.03)	0.14 (0.03)	0.14* (0.03)	0.14 (0.03)
EC _T			0.38* (0.07)	0.38 (0.07)	0.38* (0.07)	0.37 (0.07)
LEP _S			0.19* (0.05)	0.19 (0.05)	0.20* (0.05)	0.19 (0.05)
EC _S			0.28* (0.07)	0.28 (0.09)	0.28* (0.09)	0.30 (0.10)
TCHEXP _S					0.00 (0.00)	
<i>Model fit</i>						
AIC	128633.1	127438.5	126857.6	126817.7	126810.7	126810.7
BIC	128644.9**	127456.2**	126887.0**	126850.1**	126846.0**	126846.0**

Note: * $p < .05$; **=BIC decrease significant; $ICC_{Teacher} = .144$; $ICC_{School} = .183$; 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 6) will be used to answer our final research questions. Our third research question refers to whether or not any of the level-1 or level-2 variables vary at the higher levels. Examining the parameter estimates for Model 6, we see that LEP and EC vary across both level-1 and level-2, however from Model 5 we see that teacher experience did not vary significantly. Our fourth research question asks about the relationship between a teacher's teaching experience and the likelihood of being at or below a proficiency level in math. To answer this question, we examine the parameter estimate for teaching experience. This estimated slope is significant and negative ($b = -0.03$, $p < .05$), indicating that as a teacher's teaching experience increases, a student's likelihood of being at the lowest proficiency level in math decreases. Finally, our fifth research question investigates the relationship between a school's participation in an intervention program and students' likelihood of being at or below a math proficiency level. The parameter estimate for participation in the intervention is significant and negative ($b = -0.304$, $p < .05$), indicating that the likelihood of being at the lowest math proficiency level for students in schools that participated in the intervention is lower compared to students in schools that did not participate in the intervention. As shown in previous examples, to make more meaningful interpretations, the ORs portion of the output could be consulted and/or the log odds could be transformed into PP for a more meaningful interpretation.

EXAMPLE 3: TWO-LEVEL LONGITUDINAL DATA WITH DICHOTOMOUS OUTCOMES

In this example, the use of PROC GLIMMIX will be demonstrated on a longitudinal dataset in which time (level-1) is nested within students (level-2). Specifically, we will explore how students' reading achievement changed throughout the first two years of school (i.e., kindergarten through first grade). Our sample comes from the aforementioned ECLS-K data set.

In this example, we are primarily interested in assessing the impact of time and student level variables on the likelihood of a student passing a reading achievement test. At the student level, we are interested in the impact of the student's gender (1 = female) across time on the likelihood of passing the reading achievement exam. Specifically, we will examine the following research questions:

1. What is the pass rate on the reading assessment for kindergarteners? Is there variation in the pass rate in kindergarten?

2. How does the pass rate of the reading assessment change from the beginning of kindergarten through first grade? Do children vary on the pass rate over time?
3. Is a student's gender related to their pass rate on the reading assessment at the start of kindergarten?

Before starting the model building process to answer our research questions, it is important to note that in order to estimate growth models using PROC GLIMMIX, our data must be in a long format (i.e., multiple observations for each student). For a detailed example and description of this, please see Bell, Ene, Smiley, and Schoeneberger (2013).

For this example, we follow the model building process in Table 1 from Model 1 until Model 4. In this two-level growth model example, time is our level-1 variable and students are the level-2 units. First, we estimated the unconditional model with no predictors (Model 1). This model provides an overall estimate of the pass rate at kindergarten for the typical student, as well as providing information about the variability in pass rates between students in kindergarten (i.e., research question 1). The PROC GLIMMIX syntax and relevant output for the two-level unconditional growth model is shown below.

```
PROC GLIMMIX DATA=ONE METHOD=LAPLACE NOCLPRINT;
CLASS CHILDDID;
MODEL SCORE_D (EVENT=LAST)=/CL DIST=BINARY LINK=LOGIT SOLUTION;
RANDOM INTERCEPT / SUBJECT=CHILDDID 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 (-0.09849) represents the log odds of passing the reading exam at time = 0, or the fall semester/beginning of the student's kindergarten year 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	-0.09849	0.2776	119	-0.35	0.7234	0.05	-0.6482	0.4513

To be more meaningful, these initial results could be transformed into predicted probabilities (PP) using the same approach as described in example 1.

The PROC GLIMMIX syntax also provides a Covariance Parameter Estimates table (shown below) and will help us answer the second part of our first research question. Using the estimates presented in the table below, we can compute the intraclass correlation coefficient (ICC) that indicates how much of the total variation in the probability of passing the reading test is accounted for by the students. As mentioned previously, 3.29 will be used as our level-1 error variance in calculating the ICC. Note that since this is only a two-level example, only one ICC is computed.

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	CHILDDID	6.3522	1.7397	3.65	0.0001

$$ICC = \frac{\tau_{00}}{\tau_{00} + 3.29} \quad (\text{Eq. 8})$$

$$ICC = \frac{6.3522}{6.3522 + 3.29} = 0.6588$$

In the growth model context, the ICC of approximately .66 tells us that 66% of the variance in the pass rate of the reading exam exists between students. The Covariance Parameter Estimates output also reveals that the intercept variance in the unconditional model is statistically significant [$\tau_{00} = 6.3522$; $z(119) = 3.65$, $p < .0001$]. In summary, the unconditional model results revealed the probability of passing the reading exam at the beginning of kindergarten is .475, and the probability of passing the reading exam (or pass rate) varies between students. Model fit information (i.e., AIC and BIC) is also provided by PROC GLIMMIX for growth models and is used to find the best fitting model.

In order to answer the other research questions addressed by this study, we continued the model building process by first including time as a level-1 predictor. The fixed effect parameter estimate for the variable TIMEMOS represents the average change over time in the pass rate of the reading exam. The PROC GLIMMIX syntax for the second model is provided below. Notice with the syntax that this is the same code used for Model 1 with the addition of the time variable (TIMEMOS), added to the MODEL statement as well as the ODDSRATIO and (DIFF = FIRST LABEL).

```
PROC GLIMMIX DATA=ONE METHOD=LAPLACE NOCLPRINT;
CLASS CHILDDID;
MODEL SCORE_D (EVENT=LAST) = TIMEMOS/CL DIST=BINARY LINK=LOGIT SOLUTION ODDSRATIO
(DIFF=FIRST LABEL);
RANDOM INTERCEPT / SUBJECT=CHILDDID TYPE=VC;
COVTEST / WALD;
```

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

Solutions for Fixed Effects								
Effect	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
Intercept	0.003678	0.3305	119	0.01	0.9911	0.05	-0.6507	0.6580
TIMEMOS	-0.01022	0.01790	359	-0.57	0.5681	0.05	-0.04542	0.02497

Next, we add the level-1 predictors as a random intercept in the model by adding TIMEMOS to the random line using the following syntax:

```
PROC GLIMMIX DATA=ONE METHOD=LAPLACE NOCLPRINT;
CLASS CHILDDID;
MODEL SCORE_D (EVENT=LAST) = TIMEMOS/CL DIST=BINARY LINK=LOGIT SOLUTION ODDSRATIO
(DIFF=FIRST LABEL);
RANDOM INTERCEPT TIMEMOS/ SUBJECT=CHILDDID TYPE=VC;
COVTEST / WALD;
```

The above syntax produces the Covariance Parameter Estimates output below. Here we see that pass rates do not significantly vary across time.

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	CHILDDID	6.8110	2.0660	3.30	0.0005
TIMEMOS	CHILDDID	0.004751	0.007728	0.61	0.2693

Lastly, we estimate Model 4, where we add the level-2 variables to the MODEL statement as shown below. The Fixed Effects estimates table is shown below the syntax. Since TIMEMOS was not a significant random effect, this effect was dropped from the model.

```
PROC GLIMMIX DATA=ONE METHOD=LAPLACE NOCLPRINT;
CLASS CHILDDID;
MODEL SCORE_D (EVENT=LAST) = TIMEMOS SEX/CL DIST=BINARY LINK=LOGIT SOLUTION ODDSRATIO
(DIFF=FIRST LABEL);
RANDOM INTERCEPT / SUBJECT=CHILDDID CL TYPE=VC;
COVTEST / WALD;
```

Solutions for Fixed Effects								
Effect	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
Intercept	-0.3548	0.4331	118	-0.82	0.4143	0.05	-1.2126	0.5029
TIMEMOS	-0.01022	0.01789	359	-0.57	0.5682	0.05	-0.04541	0.02497
sex	0.7034	0.5548	359	1.27	0.2057	0.05	-0.3876	1.7943

Table 5 presents a summary of the results obtained for this example, including estimates for all four models considered in the model building process as well as model fit information. As with the three-level organizational model examples, the three models are compared in terms of fit in order to decide on the best fitting model for these data. Chi-squared difference tests show that Model 1 was the best fitting models, and this model will be used to answer the remaining research questions. Note that typically a growth model wouldn't be continued if time was not significant, but the rest of this example was carried out for illustrative purposes.

Table 5. Estimates from a Two-Level Growth Model Predicting Probability of Passing a Reading Test (N=120)

	Model 1 ^a	Model 2	Model 3	Model 4
<i>Fixed Effects</i>				
Intercept	-0.10 (0.28)	0.004 (0.33)	0.01 (0.34)	-0.35 (0.43)
Time		-0.01 (0.02)	-0.01 (0.02)	-0.02 (0.02)
Gender				0.70 (0.55)
<i>Error Variance</i>				
Level-2 Intercept	6.35* (1.74)	6.37* (1.74)	6.25* (1.72)	6.25* (1.72)
Time			0.00 (0.01)	
<i>Model Fit</i>				
AIC	548.03	549.71	551.09	550.08
BIC	553.61	558.07	562.24	561.23

Note: * $p < .05$; ICC = .659; Values based on SAS PROC GLIMMIX. Entries show parameter estimates with standard errors in parentheses; Estimation Method = Laplace.

^aBest fitting model

From this table, we can see that adding time and sex to the model did not significantly improve the fit above and beyond the null model. Because of this, we can conclude that the pass rate on the reading assessment did not significantly change from the beginning of kindergarten through first grade. Additionally, it seems that a student's gender is not related to their pass rates.

EXAMPLE 4: TWO-LEVEL LONGITUDINAL DATA WITH POLYTOMOUS OUTCOMES

In order to exemplify the use of PROC GLIMMIX in estimating two-level longitudinal models with polytomous outcomes, we created a polytomous version of the reading achievement score, by splitting the sample in three categories (i.e., 0 = fail, 1 = pass, 2 = advanced). The purpose of this example is to investigate the probability of students being at or below a proficiency level in reading and the changes in this probability throughout the first two years of school (i.e., kindergarten and 1st grade). More specifically, the research questions examined in this example include:

1. What is the likelihood of being at or below each proficiency level in reading for kindergarten students? Does this likelihood of being at or below each reading proficiency level vary across kindergarten students?
2. How does the likelihood of being at or below each proficiency level in reading change from kindergarten to first grade? Do children vary on their achievement levels over time?
3. Is a student's gender related to their achievement level on the reading assessment at the start of kindergarten?

Note that in this section we use a cumulative logit model. Note that before researchers start building models using this method they should test the assumption of equal slopes (O'Connell & McCoach, 2008), which was not presented here for brevity. Similar with the dichotomous example, we will only be estimating the first 4 models in Table 1 and 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 (Model 1) estimated for this example is presented below. As with the organizational models, 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= ONE METHOD=LAPLACE NOCLPRINT;
CLASS CHILID;
MODEL SCORE_P= / DIST=MULTI LINK=CLOGIT SOLUTION CL;
RANDOM INTERCEPT/ SUBJECT= CHILID TYPE=VC;
COVTEST/WALD;
```

The major difference in terms of the output generated for this unconditional model consists in the Solution for Fixed Effects table (shown below). 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 reading proficiency levels (i.e., fail and pass) of kindergarten students.

Solutions for Fixed Effects									
Effect	SCORE_P	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
Intercept	0	0.05600	0.3108	119	0.18	0.8573	0.05	-0.5593	0.6713
Intercept	1	2.3644	0.3483	119	6.79	<.0001	0.05	1.6747	3.0542

These fixed effects estimates are useful for answering our first research question regarding the likelihood of being at or below each proficiency level in reading for kindergarten students. More specifically, these log odds can be used to calculate the PP of being at or below each reading proficiency level, 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 failing reading level for kindergarten students is 0.0560, resulting in PP of 0.5139. Similarly, the log odds of being at or below the passing reading level is 2.3644, resulting in a cumulative probability of 0.9140. Finally, the cumulative probability of being at or below the advanced level in reading adds to 1. As with the organizational models, 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 failing reading level in kindergarten is .5139, at the passing level is .4001 (.9140 - .5139), and at the advanced level is .0860 (1 - .9140).

Another important piece of the output generated for this model is the Covariance Parameter Estimates table. Similar with the dichotomous example, results presented in this table help us answer the second part of the first research question of our study regarding the variability across kindergarten students in terms of the likelihood of being at or below each reading proficiency level. Results indicate that this likelihood significantly varies across kindergarten students [τ_{00} = 8.3787, $z(119) = 4.32$, $p < .0001$]. Also, using the intercept variance estimate provided in this table (8.3787), the ICC for this model is calculated in the same way as described for the previous examples, resulting in an ICC value of .718. This means that approximately 72% of the variability in the likelihood of being at or below a certain reading proficiency level is accounted for by student characteristics.

Covariance Parameter Estimates						
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z	
Intercept	CHILDDID	8.3787	1.9400	4.32	<.0001	

To continue the model building process, we need to estimate Models 2-4. Table 6 shows the syntax and changes between the syntax from the previous model. All of the output looks similar to the null models (and are analogous to the previous examples), and will not be included for brevity. Table 7 presents a summary of the results obtained for this example, including estimates for all four models considered in the model building process as well as model fit information. As with the previous examples, the four models are compared in terms of fit in order to decide on the best fitting model for these data and that model will be used to answer the remaining research questions.

Table 6. Model Building Syntax for Models 2-4

Model	Syntax	Changes from previous model
2	PROC GLIMMIX DATA=ONE METHOD=LAPLACE NOCLPRINT; CLASS CHILDDID; MODEL SCORE_P=TIMEMOS / DIST=MULTI LINK=CLOGIT SOLUTION CL; RANDOM INTERCEPT/ SUBJECT= CHILDDID TYPE=VC; COVTEST/WALD;	TIMEMOS (a level-1 variable) is entered as a fixed effect by adding it to the MODEL statement
3	PROC GLIMMIX DATA= ONE METHOD=LAPLACE NOCLPRINT; CLASS CHILDDID; MODEL SCORE_P=TIMEMOS / DIST=MULTI LINK=CLOGIT SOLUTION CL; RANDOM INTERCEPT TIMEMOS/ SUBJECT= CHILDDID TYPE=VC; COVTEST/WALD;	TIMEMOS is allowed to be a random slope by including it on the RANDOM statement
4	PROC GLIMMIX DATA= ONE METHOD=LAPLACE NOCLPRINT; CLASS CHILDDID; MODEL SCORE_P=TIMEMOS SEX / DIST=MULTI LINK=CLOGIT SOLUTION CL; RANDOM INTERCEPT / SUBJECT= CHILDDID TYPE=VC; COVTEST/WALD;	SEX (a level-2 variable) is entered as a fixed effect by adding it to the MODEL statement **Note: Since TIMEMOS was not a significant random effect, it was not carried forward.

Table 7. Estimates for Two-level Longitudinal Polytomous Models of Reading Achievement (N = 120)

	Model 1 ^a	Model 2	Model 3	Model 4
<i>Fixed Effects</i>				
Intercept 0 (Fail)	0.06 (0.31)	0.01 (0.35)	-0.07 (0.39)	0.40 (0.47)
Intercept 1 (Pass)	2.36* (0.35)	2.32* (0.38)	2.68* (0.47)	2.71* (0.50)
Time		0.005 (0.02)	0.02 (0.02)	0.004 (0.02)
Student gender				-0.77 (0.60)
<i>Error Variance</i>				
Intercept	8.38* (1.94)	8.38* (1.94)	10.76* (3.04)	8.28 (1.92)
Time			0.02 (0.01)	
<i>Model Fit</i>				
AIC	798.23	800.13	795.31	794.84
BIC	806.59	811.28	809.25	811.56

Note: * $p < .05$; **=significant change in BIC; ICC = .718; Values based on SAS PROC GLIMMIX. Entries show parameter estimates with standard errors in parentheses; Estimation Method = Laplace.

^aBest fitting model

Although the models including time and student gender do not appear to be significantly better models, we can still interpret the parameter estimates of interest to answer our remaining research questions. More specifically, our third research question refers to the change in the log odds of being at or below a reading proficiency level over time. To answer this question, we use the parameter estimate for time ($b = 0.005$, $p > .05$) indicating that although there is a small positive change in students probability of being at or below a reading proficiency level over time, this change is not statistically significant. In other words, there is not enough evidence to suggest that students' likelihood of being at or below a certain reading level changes over time. Similarly, to answer our last research question, we use the estimate for student gender ($b = -0.77$, $p > .05$) which indicates a negative, non-significant relationship between students' gender and their probability of being at or below a reading proficiency level in kindergarten. Therefore, it appears that there is not enough evidence to suggest that boys and girls differ in their likelihood of being at or below a certain reading level upon entry into kindergarten. 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 previous examples.

CONCLUSIONS

In addition to modeling two-level organization models with fixed effects models for HGLMs, this paper shows that PROC GLIMMIX is useful to estimate many different models. Applied researchers can use this paper to learn how to model random slopes, three-level models, and growth models with both binary and polytomous outcomes. In addition, in this paper we showed how to interpret the estimates provided from PROC GLIMMIX in order to answer research questions of interest.

REFERENCES

- Bell, B.A., Ene, M., Smiley, W., & Schoeneberger, 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>
- Ene, M., Leighton, E. A., Blue, G. L., & Bell, B. A. (2015). Multilevel models for categorical data using SAS® PROC GLIMMIX: The basics. SAS Global Forum 2015 Proceedings. https://sasglobalforum.activeevents.com/2015/connect/fileDownload/session/4564BA7822098A5E66412C44FBC24251/3430_Ene-SGF2015FinalPaper_MultilevelModelsForCategoricalDataUsingSASPROCGLIMMIX_TheBasics.pdf
- O'Connell, A.A. & McCoach, D.B. (2008). *Multilevel modeling of educational data*. Charlotte, NC: Information Age Publishing.
- 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.
- Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 23 (4), 323-355.

CONTACT INFORMATION

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

Name: Whitney Smiley
Enterprise: University of South Carolina
Address: 8788 Cypress Grove Run
City, State ZIP: Raleigh, NC 27612
Work Phone: (919) 531-1479
E-mail: whitknee48@gmail.com

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 trademarks of their respective companies.