

# Journal of the Midwest Association for Information Systems (JMWAIS)

---

Volume 1

Issue 2 Volume 2015, Issue 2, July 2015

Article 4

---

2015

## Estimating Random Effects in Multilevel Structural Equation Models Using Mplus

Andy Luse

Oklahoma State University, [andylose@okstate.edu](mailto:andylose@okstate.edu)

Follow this and additional works at: <http://aisel.aisnet.org/jmwais>

---

### Recommended Citation

Luse, Andy (2015) "Estimating Random Effects in Multilevel Structural Equation Models Using Mplus," *Journal of the Midwest Association for Information Systems (JMWAIS)*: Vol. 1 : Iss. 2 , Article 4.

Available at: <http://aisel.aisnet.org/jmwais/vol1/iss2/4>

This material is brought to you by the Journals at AIS Electronic Library (AISeL). It has been accepted for inclusion in Journal of the Midwest Association for Information Systems (JMWAIS) by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

Date: 7-27-2015

## Estimating Random Effects in Multilevel Structural Equation Models Using Mplus

Andy Luse

Oklahoma State University, [andyluse@okstate.edu](mailto:andyluse@okstate.edu)

### Abstract

This research provides a method for discovering group-level differences for each individual group compared to the overall average for all groups on endogenous variables in a multilevel structural equation modeling context using the Mplus software package. Furthermore, methods for calculating associated significance values of these group-specific differences is described. This builds on current techniques for discovering group-level differences in multilevel regression models by extending this capability to full multilevel structural equation models. The included analysis provides a verification mechanism for the proposed method in a multilevel regression context using Mplus by comparing the output with SAS. This provides confirmation that the Mplus method mirrors SAS output for multilevel regression and therefore can then be extended to multilevel structural equation models using Mplus.

**Keywords:** multilevel structural equation modeling, random effects, multilevel regression, group differences

Copyright © 2015 by Andy Luse

## 1. Introduction

Multilevel modeling (also referred to as multilevel analysis, hierarchical linear modeling, or multilevel regression) is a type of data analysis which allows for a focus on nested sources of variability in data (Snijders & Bosker, 1999). Multilevel modeling provides a statistical method that can analyze nested (e.g. group) data by allowing the researcher to analyze individual-level, group-level, and cross-level effects simultaneously in the same model (Hofmann, 1997; Raudenbush & Bryk, 1992). Many areas of data analytics and science are naturally suited for multilevel analysis including sociological issues such as individuals within neighborhoods (Leventhal & Brooks-Gunn, 2000), family issues of members within households (Teachman & Crowder, 2002), psychological issues such as individual depression within states (Chen, Subramanian, Acevedo-Garcia, & Kawachi, 2005), educational issues such as students within classrooms and/or schools (Koth, Bradshaw, & Leaf, 2008), and organizational issues such as members within teams (Short, Piccoli, Powell, & Ives, 2005) just to name a few.

Traditional multilevel modeling builds on conventional regression analysis to allow for estimation of effects at multiple levels of a regression model. While this adds a much needed tool for multilevel analysis, the use of regression within a multilevel context is limiting for researchers. Structural equation modeling (SEM) builds on traditional regression by allowing for the simultaneous estimation of both a measurement factor-analytic model and a structural model (Gefen, Straub, & Boudreau, 2000). Recently multilevel SEM (MLSEM) has emerged as a viable technique for combining the advantages of multilevel modeling with that of SEM (B. Muthén & Asparouhov, 2011).

MLSEM provides tremendous opportunities for statistical analyses, but MLSEM software tools have still not added some key functionality/output that is available in software packages for multilevel regression analyses as output. One such feature is the ability to calculate individual group deviations, and associated significance values, on endogenous dependent variables from the overarching grand group value. This type of post-hoc analysis allows the researcher to identify the specific groups in a multilevel analysis that significantly differ on a dependent variable, which allows for a fuller more nuanced picture. This research describes methods for calculating such group deviational effects in a MLSEM context using the popular Mplus software package (L. Muthén, Muthén, Asparouhov, & Nguyen, 2011). Given that these methods are new and not provided as a built-in option in the software, a multilevel *regression* model in Mplus will be compared with SAS, which provides such estimates as an option. This comparison provides validation for this technique and allows researchers to extend such techniques to a multilevel *SEM* context using Mplus.

## 2. Factor Scores

Mplus provides factor scores for all latent variables in a multilevel model, including  $\mathbf{v}_{gb}$ ,  $\boldsymbol{\eta}_{gb}$ , and  $\boldsymbol{\eta}_{giw}$ , taken from equations for a two level model.

$$\begin{aligned} y_w &= v_w + \Lambda_w \eta_w + K_w x_w + \varepsilon_w \\ \eta_w &= \alpha_w + \beta_w \eta_w + \Gamma_w x_w + \xi_w \\ y_b &= v_b + \Lambda_b \eta_b + K_b x_b + \varepsilon_b \\ \eta_b &= \alpha_b + \beta_b \eta_b + \Gamma_b x_b + \xi_b \end{aligned}$$

In the above,  $\mathbf{v}$  are the intercepts of group  $g$  at the between level  $b$  and  $\boldsymbol{\eta}$  are the latent endogenous variables for both group  $g$  at the between level  $b$  and individual  $i$  in group  $g$  at the within level  $w$ .<sup>1</sup> These factor scores are the posterior means obtained as in a single-level single-group estimation utilizing E-step and M-step procedures implementing a modified Quasi-Newton EM algorithm (Lange, 1995a, 1995b). At the group level, these factor scores are estimates of the overarching group-based mean for all individuals within a particular group for that latent variable.

Beginning with version 6, Mplus added the functionality of estimating standard errors for factor scores. In the context of multilevel models, a factor score is given for each variable that is regressed on at least one other variable at the between level. From here, a standard error is computed that corresponds to the estimated factor score. Using numerical integration, this standard error is obtained using the formula:

$$\sqrt{\left(\sum p_i * x_i^2\right) - \left(\sum p_i * x_i\right)^2}$$

<sup>1</sup> Please see (Bentler & Liang, 2003; B. Muthén, 1994; L. Muthén, 2012) for an in-depth discussion of all aspects of these equations.

Where  $x_i$  are the integration points and  $p_i$  are the posterior probabilities of  $\eta = \xi$  for all observed data (L. Muthen, personal communication, June 30, 2014).

The standard errors associated with the factor scores are the needed component to allow for the method described in this research to be possible. These standard errors enable the method described to identify differences in group values for specific groups on specific variables at the group level as well as significance values associated with these differences.

### 3. Estimation

Recent advancements in software functionality have allowed for the estimation of multilevel SEM models. The software allows the user to specify two-level models, as in traditional multilevel regression models, while also allowing for concurrent measurement and structural model estimation, as in SEM. This allows the user to identify random intercepts between groups, random slopes, the impact of group level covariates, cross-level interactions, etc. This has provided a much needed step forward in the researcher's methodological toolkit and provides for a much richer analysis for these types of datasets.

Traditional multilevel regression techniques have also allowed for the estimation of post-hoc tests of individual group deviations from the overall grand mean of all groups on the dependent variable. For example, when looking at students within schools using a dependent variable of math achievement, these post-hoc analyses allow users to see beyond the fact that schools may differ on average student math achievement, but also to see how each specific school differs in their average student math score from the overall grand mean of math scores. Furthermore, software packages allow the researcher to identify which of these differing schools significantly deviate from this grand mean, providing a method for identifying those schools that may be significantly lower or higher on average student math score. This potentially allows for greater analysis as to why the school is significantly lower/higher on average math achievement and ideally devise a plan for increasing the scores of the students in the lower schools using methods employed by higher schools.

Some multiple regression software packages offer the ability to perform the above post-hoc tests. One popular package, SAS (SAS, 2011), provides the solution in a section titled "Solution for Random Effects." This solution provides the beta estimation of the group difference for each group, as well as the associated standard error, t-statistic, and p-value. To date, MLSEM software does not offer an option for specifying that these group differences and associated t-tests be calculated. Also, no method has yet been explicated to utilize these MLSEM software packages to calculate these deviations of group means from the grand mean with associated significance values.<sup>2</sup>

This research devises a method for estimating group differences and associated significance values in a MLSEM context using the Mplus software package (L. Muthén et al., 2011). Since this is the first attempt at such calculations, the methods used to calculate such differences need to be verified before they can be trusted fully. The SAS PROC MIXED procedure has been used extensively for multilevel regression models by researchers, and this method offers the ability to evaluate group differences in a multilevel regression context. Mplus is a software package which is capable of estimating MLSEMs, but no research to date has utilized this software to find group differences and associated significance values. Mplus is also capable of estimating multilevel regression models using the same basic syntactical approaches as it uses to estimate its multilevel structural equation models. Therefore, to help verify that Mplus is correctly estimating group differences in a MLSEM using our proposed method, its results for a multilevel regression model can be compared to that of SAS. If these numbers align, this can offer verification that the same proposed method can be extended to an MLSEM context using Mplus.

#### 3.1 Model Description

To test the two software packages, the same dataset was utilized for both packages. This data consists of 309 high school students nested within 40 high schools.<sup>3</sup> The research utilizes Lent's Social Cognitive Career Theory (SCCT) (Lent, 2005) to aid in the prediction of a student's intention to major in information technology (IT). For estimation of the multilevel regression model (to allow SAS's PROC MIXED procedure to verify the proposed method using Mplus) the items for each of the student-level independent and dependent variables were averaged to create a

---

<sup>2</sup> As an example, if SAS does not include a specific option for some calculation, users have been known to develop their own solutions using SAS macros to estimate the item of interest. In regard to calculating group-level differences and corresponding significance tests, Mplus does not have a specific option for calculating these estimates as does SAS PROC MIXED, and furthermore, no one has yet devised a "homebrewed" method, as with a SAS macro, for filling in this functionality.

<sup>3</sup> This data is utilized from previous research (Luse, Rursch, & Jacobson, forthcoming), that used this same dataset but only in a single-level model without looking at group differences.

single variable. School size was also utilized as a school-level covariate. The model and proposed hypotheses are displayed in Figure 1. For simplicity, we only look at random intercepts for this example, but the use of this same method for random slopes is discussed later. Since this research is concerned with methodological issues, we will not fully detail the SCCT model or its underlying relationships. The reader is referred to (Lent, 2005; Lent, Brown, & Hackett, 1994) for a review of the SCCT model.

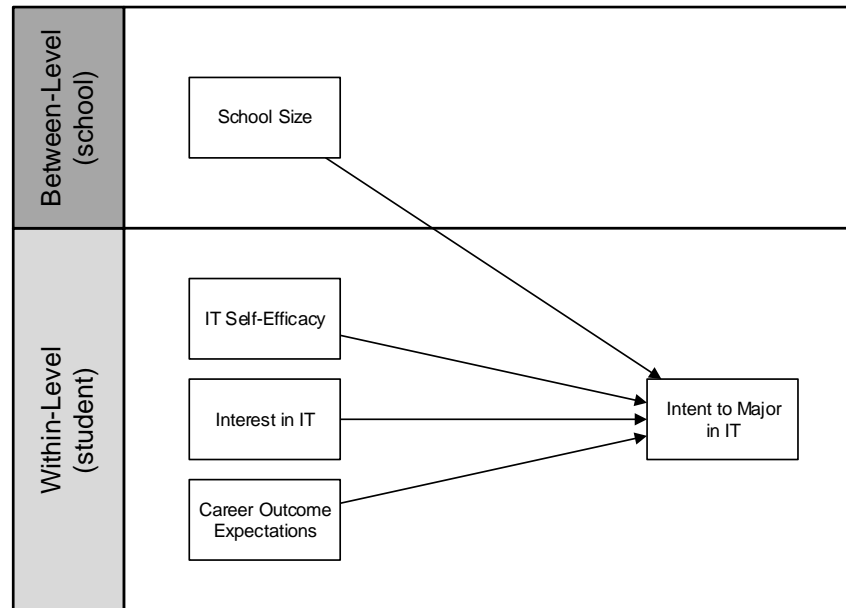


Figure 1. Hypothesized research model.

### 3.2 Syntax Setup for SAS and Mplus

In order to compare model output between SAS and Mplus, we first need to estimate each model. This section will describe each of the syntax files used to estimate the SCCT model. Figure 2 shows the SAS syntax for estimating the multilevel regression SCCT model.

```

DATA mydata;
  INFILE "C:\filename.dat";
  INPUT  ITSE1 ITSE2 ITSE3 ITSE4 ITSE5 ITSE6 ITSE7 ITSE8 ITSE9
        Int1 Int2 Int3 Int4 Int5
        Career1 Career2 Career3 Career4
        Intent
        sch_num sch_size;

RUN;

DATA mydata1;
  SET mydata;
  ITSE = MEAN(ITSE1,ITSE2,ITSE3,ITSE4,ITSE5,ITSE6,ITSE7,ITSE8,ITSE9);
  Interest = MEAN(Int1,Int2,Int3,Int4,Int5);
  Career = MEAN(Career1,Career2,Career3,Career4);

RUN;

PROC SQL;
  CREATE TABLE grpmeanctr AS
  SELECT sch_num, ITSE - MEAN(ITSE) AS grpcITSE, Interest - MEAN(Interest) AS grpcInterest,
        Career - MEAN(Career) AS grpcCareer, Intent, sch_size
  FROM mydata1
  GROUP BY sch_num;

  CREATE TABLE grpANDgrandmeanctr AS
  SELECT *, sch_size - MEAN(sch_size) AS grdcsch_size
  FROM grpmeanctr;

QUIT;

PROC MIXED DATA=grpANDgrandmeanctr METHOD=ML COVTEST;
  CLASS sch_num;
  MODEL Intent = grpcITSE grpcInterest grpcCareer grdcsch_size /SOLUTION DDFM=BW;
  RANDOM INTERCEPT / SOLUTION SUBJECT=sch_num TYPE=UN;

RUN;

```

Figure 2. SAS syntax for multilevel regression model of SCCT.

The first DATA statement is used to bring in the data from the associated file that contains the data needed for the analysis. This method uses a fixed ASCII file (to enable the use of the same exact input file for both SAS and Mplus), but the user can use other methods for importing data. The second DATA statement is used to compute each of the student-level independent variables. Given that this is a regression analysis, we must compute one observed variable by taking the average of each of the items that will be used to compute each of these variables. By taking the average, this also allows for centering of variables, which is required when running this type of multilevel analysis. The PROC SQL statement is used to first center each of the student-level independent variables within their associated school group (also referred to as centering within context) using the first CREATE TABLE statement. The second CREATE TABLE statement then uses the table created by the first CREATE TABLE statement and centers the school-level variable of sch\_size based on the grand mean of all school sizes. An in-depth discussion of SQL is beyond the scope of this manuscript, but this or other methods should be used to center the student-level independent variables *within* school and the school-level independent variables *across* schools before performing the multilevel analysis.

The PROC MIXED statement is used to run the actual multilevel regression analysis. The DATA statement tells SAS which dataset to use, while the METHOD informs SAS to use a maximum likelihood (ML) estimation method and COVTEST tells the program to run a significance test of the student-level and school-level covariance estimates. Restricted maximum likelihood (REML) is typically used as the default estimation method, but ML was used to compare the SAS output with Mplus as Mplus will also be set to use the ML option. Also, ML is the default method used for full SEM models, so this will facilitate the move from regression-based multilevel modeling to SEM-based multilevel modeling in the future.

Next, the actual model to be estimated is specified. The CLASS statement tells the program which variable will be used to group the observations. Given that this is school data, the associated sch\_num will be used to group student-level variables. Next, the model statement tells the program to regress the dependent variable Intent (Intent on Major in IT) on the student-level independent variables of grpcITSE (group-centered IT Self-Efficacy), grpcInterest (group-centered Interest in IT), and grpcCareer (group-centered Career Outcome Expectations), as well as the school-level covariate of grdcsch\_size (grand mean-centered school size). The /SOLUTION option tells the program to give estimates and associated significance values for each of the independent variables in the model and DDFM=BW tells SAS to use the between/within method for computing the denominator degrees of freedom for fixed effect hypothesis tests. Next, the RANDOM statement tells the model to estimate random intercepts (for each school) and also to give estimates and associated significance values for each of these random intercepts. By adding this SOLUTION

statement, SAS will provide the solution for random effects that we need to see the difference between each specific school on average student Intent to Major in IT from the overall grand mean of intent to major for all schools. Finally, the SUBJECT parameter tells SAS which variable will identify the subjects in this analysis (given this is a multilevel analysis, the school will actually be considered the subject level) while the TYPE parameter tells SAS to assume a completely generalized covariance where no restrictions are assumed.

```

TITLE:
  SCCT MLM;

DATA:
  FILE IS filename.dat;

VARIABLE:
  NAMES ARE ITSE1 ITSE2 ITSE3 ITSE4 ITSE5 ITSE6 ITSE7 ITSE8 ITSE9
           Int1 Int2 Int3 Int4 Int5
           Career1 Career2 Career3 Career4
           Intent
           sch_num sch_size;

  USEVARIABLES = Intent sch_size ITSE Interest Career;

  WITHIN = ITSE Interest Career;
  BETWEEN = sch_size;
  CLUSTER = sch_num;

ANALYSIS:
  TYPE = TWOLEVEL;
  ESTIMATOR = ML;

DEFINE:
  ITSE = MEAN(ITSE1 ITSE2 ITSE3 ITSE4 ITSE5 ITSE6 ITSE7 ITSE8 ITSE9);
  Interest = MEAN(Int1 Int2 Int3 Int4 Int5);
  Career = MEAN(Career1 Career2 Career3 Career4);

  CENTER ITSE Interest Career (GROUPMEAN);
  CENTER sch_size (GRANDMEAN);

MODEL:
  %WITHIN%
  Intent ON ITSE Interest Career;

  %BETWEEN%
  Intent ON sch_size;

  f BY;
  Intent ON f@1;
  Intent@0;

OUTPUT:
  SAMP STANDARDIZED;

SAVEDATA:
  FILE = mlm_output.txt;
  SAVE = FS;

```

Figure 3. Mplus syntax for multilevel regression model of SCCT.

Mplus syntax is broken up into separate sections. The TITLE section provides a place for providing a descriptive title of the analysis being performed. The DATA section allows you to specify the file to be used employing the FILE IS statement. The file used here is a fixed format ASCII file without any column identifiers in the file itself (i.e. only data). Next, the VARIABLE section allows you to identify the variables in both the file and those to be used in the analysis. The NAMES ARE statement provides a listing of the various data columns present in the data file. For this analysis, only the variables used in the analysis were included in the data file, but if the data file includes more variables than are used in the analysis, the USEVARIABLES statement allows for the specification of which variables from the file will actually be used in the analysis. Notice that since we will be combining individual variables

into their respective constructs, we only specify the combined variables as those which will be used in the analysis. The WITHIN and BETWEEN statements identify which variables used in the analysis are at the within level (i.e. student

level) and which are at the between level (i.e. school level), while the CLUSTERING statement identifies which variable will be used to identify groups (i.e. schools) within the data. The ANALYSIS section allows for the identification of the types of analysis to be used. Here we are using a TWOLEVEL analysis with ML as the estimation technique. The DEFINE section allows for the computation of other variables to be used in the model. Given that this is a regression model, we have defined combined measures for each student-level variable using the average of the constituent items for each. The CENTER option allows the identification of which variables should be either grand-mean or group-mean centered.

The MODEL section is used to define the model that will be estimated by Mplus. When estimating a multilevel model, the MODEL section contains two subsections, the %WITHIN% (i.e. student-level) subsection and the %BETWEEN% (i.e. school-level) subsection. The within subsection of our model is used to specify the primary regression model. For this research we regress Intent to Major in IT on IT Self-Efficacy, Interest in IT, and Career Outcome Expectations. The between subsection specifies the added regression of Intent to Major on school size.

The final portion of the between section of the model is a novel approach developed specifically for this research as a way of estimating the group differences that mirror the solution for random effects estimated by SAS. Mplus does not have a method for easily estimating these parameters, but these developed statements provide a method for doing so. First, we create a latent variable *f* which has no indicators. We set the variance of this latent variable to 1 and then regress Intent to Major on this latent variable. Finally, we set the variance of Intent to Major at 0. This moves the residual variance term of Intent to Major to the latent *f* variable with mean zero. By having a separate variable *f* to hold the residual of the dependent variable, our analysis will be able to provide an estimate (factor score) and associated standard error for this variable for each individual group. We then use these group estimates and associated standard errors to derive an estimation of a t-statistic for each group residual on Intent to Major as well as an associated significance value using the output from the SAVEDATA command (described below). The OUTPUT subsection specifies that we would like to get sample and standardized statistics related to the fixed effects in the model.<sup>4</sup>

The SAVEDATA command allows for a separate data file apart from the standard Mplus output to be saved along with the analysis. The user can choose where the file should be saved as well as what items should be saved in the file beyond the standard variable values. We have specified that factor scores (FS) should be saved as these will provide the deviational estimates for endogenous variables from the grand mean on that endogenous variable (using the factor score of the created *f* variable) and associated standard errors. The combination of the added *f* commands in the model statement as well as the FS option in the SAVEDATA section provides the ability to estimate group differences and significance values.

### **3.3 Model Statistics Output Comparison between SAS and Mplus**

The model above was estimated using both SAS PROC MIXED and Mplus. Before discussing the group differences, we will first compare standard output from SAS and Mplus to help verify that these two software packages are arriving at the same results, and thereby provide support for the proposed method using Mplus. First, looking at the basic model descriptors (Tables 1 and 2), we see that both software packages are showing 309 observations at the student level and 40 observations at the group level (referred to as subjects and clusters in SAS and Mplus, respectively). Also notice that both models have the same number of independent variables (SAS counts the intercept in its "Columns in X," so it is one greater than Mplus).

---

<sup>4</sup> It should be noted that the method described in this paragraph was developed as a novel method for finding group differences in multilevel SEM and is the primary contribution of this research.



Table 1. SAS basic model descriptors.

Dimensions	
Covariance Parameters	2
Columns in X	5
Columns in Z Per Subject	1
Subjects	40
Max Obs Per Subject	24

Number of Observations	
Number of Observations Read	309
Number of Observations Used	309
Number of Observations Not Used	0

Table 2. Mplus basic model descriptors.

Number of groups	1
Number of observations	309
Number of dependent variables	1
Number of independent variables	4
Number of continuous latent variables	1
.....	
SUMMARY OF DATA	
Number of clusters	40
Average cluster size	7.725

SAS offers very few fit statistics (Table 3), but it does offer some focused on the information criteria method.

Table 3. SAS model fit statistics.

Fit Statistics	
-2 Log Likelihood	1137.6
AIC (smaller is better)	1151.6
AICC (smaller is better)	1151.9
BIC (smaller is better)	1163.4

By comparison, Mplus offers a much larger number and variety and of fit statistics (Table 4). This will become important when Mplus is used in a MLSEM context, but here we will just focus on those fit statistics that align with SAS.

Table 4. Mplus model fit statistics.

TESTS OF MODEL FIT	
Chi-Square Test of Model Fit	
Value	0.000
Degrees of Freedom	0
P-Value	1.0000
Chi-Square Test of Model Fit for the Baseline Model	
Value	76.652
Degrees of Freedom	4
P-Value	0.0000
CFI/TLI	
CFI	1.000
TLI	1.000
Loglikelihood	
H0 Value	-568.777
H1 Value	-568.777
Information Criteria	
Number of Free Parameters	7
Akaike (AIC)	1151.554
Bayesian (BIC)	1177.687
Sample-Size Adjusted BIC	1155.486
(n* = (n + 2) / 24)	
RMSEA (Root Mean Square Error Of Approximation)	
Estimate	0.000
SRMR (Standardized Root Mean Square Residual)	
Value for Within	0.000
Value for Between	0.000

First, as you can see, the AIC measure aligns perfectly with SAS at 1151.6. Mplus provides a Log Likelihood value (-568.777) whereas SAS offers a -2 Log Likelihood value (1137.6), but if you take the value provided by Mplus times -2, this value matches up with SAS perfectly (-568.777 \* -2 = 1137.6). The only other value displayed by both SAS and Mplus is the BIC value. The corresponding values are similar, but not the same (1163.4 vs. 1177.687). This is a notable difference that we will discuss further in the Discussion section.

Next, we look at the statistical output for both programs including both covariance parameters and fixed effects (Tables 5, 6, and 7).

Table 5. SAS covariance parameter estimates.

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
UN(1,1)	sch_num	0.3036	0.1391	2.18	0.0146
Residual		2.1354	0.1814	11.77	<.0001

Table 6. SAS model fixed effects.

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	4.9953	0.1314	38	38.02	<.0001
grpcITSE	0.09415	0.07760	266	1.21	0.2261
grpcInterest	0.5176	0.09836	266	5.26	<.0001
grpcCareer	0.3268	0.1089	266	3.00	0.0030
grdesch_size	0.000641	0.000257	38	2.50	0.0169

Table 7. Mplus model statistics (including covariance parameters and fixed effects).

MODEL RESULTS				
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level				
INTENT ON				
ITSE	0.094	0.078	1.213	0.225
INTEREST	0.518	0.098	5.262	0.000
CAREER	0.327	0.109	3.001	0.003
Residual Variances				
INTENT	2.135	0.181	11.773	0.000
Between Level				
INTENT ON				
F	1.000	0.000	999.000	999.000
INTENT ON				
SCH_SIZE	0.001	0.000	2.489	0.013
Intercepts				
INTENT	4.995	0.132	37.769	0.000
Variances				
F	0.304	0.139	2.181	0.029
Residual Variances				
INTENT	0.000	0.000	999.000	999.000

The first thing that becomes apparent is the difference in placement of several of the output measures between the two programs, specifically the covariance parameters. The group-level covariance parameter (also referred to as  $\tau_{11}$ ) in the SAS output is located in Mplus under the Variances section for F ( $\tau_{11} = 0.304$ ) while the individual-level covariance parameter (also referred to as  $\sigma^2$ ) is located in Mplus under the Residual Variances section for INTENT ( $\sigma^2 = 2.135$ ). Conversely, the fixed effects are fairly easy to identify in the output between the two models with the most notable difference being that Mplus provides a separate Between Level section for group-level variables. There are two differences to note before we proceed. First, is the inclusion in the Mplus Between Level effects of the variable F. This is used to estimate the post-hoc group differences in Mplus. Given that Mplus does not offer a specific option for estimating these between group differences, separate syntax was added to allow for calculation of these group-level differences (described above). Secondly, while the estimates match up exactly, the t-values and associated significance values differ very slightly between SAS and Mplus for those values analyzed at the between level in the Mplus output (i.e.  $\tau_{11}$ , Intercept, and SCH\_SIZE). This slight difference is due to the default maximum likelihood estimation algorithms and the associated method for analyzing group differences with regard to variances for each. This will be discussed in more detail in the Discussion section.

One additional piece of information offered by Mplus is the  $R^2$  value for the dependent variable of choice to major in IT (Table 8).<sup>5</sup>

<sup>5</sup> While you can use either the Raudenbush and Bryk method (1992) or the Snijders and Bosker method (1999) for estimating  $R^2$

Table 8. Mplus R-square output

R-SQUARE				
Within Level				
Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
INTENT	0.207	0.041	5.017	0.000

After analyzing the above model using both SAS and Mplus we find that both software products estimate the multilevel regression model producing the same output in each. Figure 4 shows a graphical display of the model findings with the associated estimates and significance values which were found in both SAS and Mplus.

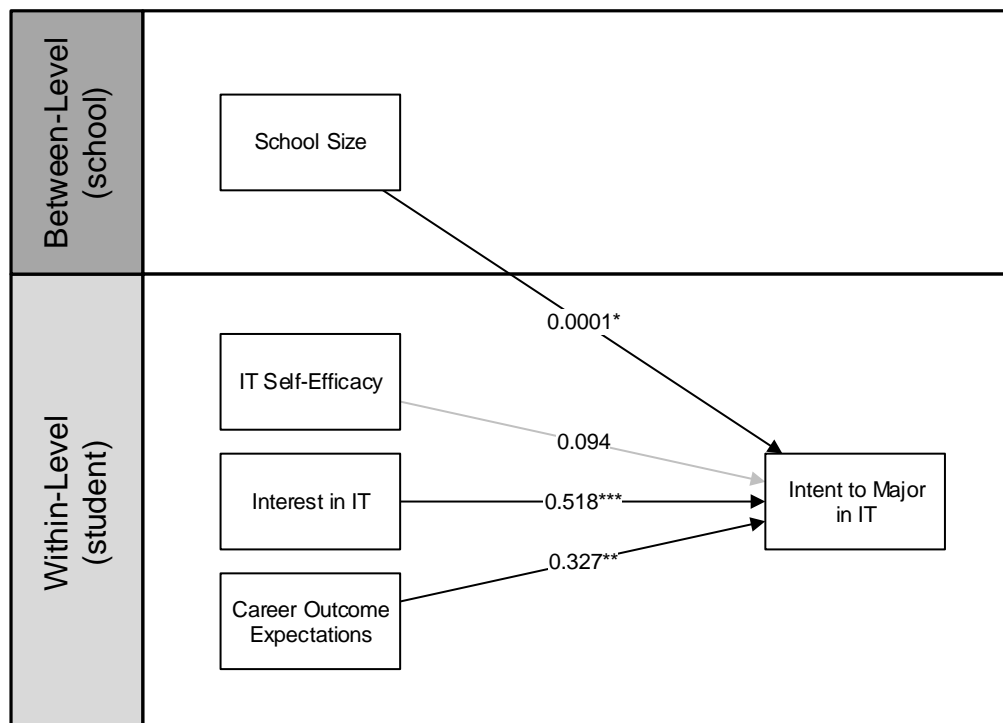


Figure 4. Multilevel regression model (hierarchical linear model) with estimates (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ).

### 3.4 Model Post-Hoc Group Differences Calculation and Comparison between SAS and Mplus

The ability to estimate group-level deviations on a dependent variable from the overall group grand mean is of great usefulness for researchers in identifying those groups who may need further analysis and possible implementation of change programs. SAS offers this output in a table titled Solution for Random Effects which is provided by adding the SOLUTION parameter to the RANDOM portion of the PROC MIXED command. Table 9 shows the output from this command.

values for a multilevel model, Mplus offers this as a standard portion of the output.

Table 9. SAS output for group differences from overall grand group mean.

Solution for Random Effects						
Effect	sch_num	Estimate	Std Err	Pred DF	t Value	Pr >  t
Intercept	19	0.03786	0.3589	304	0.11	0.9161
Intercept	20	0.3184	0.3116	304	1.02	0.3077
Intercept	22	-0.6024	0.4082	304	-1.48	0.1410
Intercept	24	-0.01115	0.4293	304	-0.03	0.9793
Intercept	26	-0.5339	0.3682	304	-1.45	0.1481
Intercept	29	0.9487	0.3474	304	2.73	0.0067
Intercept	30	0.7926	0.4432	304	1.79	0.0747
Intercept	33	0.4836	0.4643	304	1.04	0.2985
Intercept	36	0.1869	0.4268	304	0.44	0.6618
Intercept	37	0.1939	0.3864	304	0.50	0.6162
Intercept	38	-0.07913	0.3129	304	-0.25	0.8005
Intercept	40	-0.2152	0.3032	304	-0.71	0.4784
Intercept	44	-0.03580	0.4024	304	-0.09	0.9292
Intercept	46	0.3194	0.3732	304	0.86	0.3927
Intercept	56	-0.9565	0.3674	304	-2.60	0.0097
Intercept	65	0.1786	0.4878	304	0.37	0.7145
Intercept	94	0.2275	0.3031	304	0.75	0.4536
Intercept	96	-0.04197	0.4910	304	-0.09	0.9319
Intercept	97	-0.5456	0.3006	304	-1.82	0.0705
Intercept	100	0.07200	0.3633	304	0.20	0.8430
Intercept	101	-0.01016	0.3990	304	-0.03	0.9797
Intercept	104	0.04109	0.3515	304	0.12	0.9070
Intercept	107	0.1000	0.5159	304	0.19	0.8464
Intercept	109	-0.7405	0.3029	304	-2.44	0.0151
Intercept	111	-0.01218	0.4101	304	-0.03	0.9763
Intercept	114	-0.1920	0.4872	304	-0.39	0.6938
Intercept	115	0.04750	0.5168	304	0.09	0.9268
Intercept	116	-0.1384	0.4428	304	-0.31	0.7548
Intercept	117	-0.3646	0.4438	304	-0.82	0.4120
Intercept	118	0.1830	0.3978	304	0.46	0.6459
Intercept	140	-0.07716	0.4872	304	-0.16	0.8743
Intercept	141	0.1664	0.4875	304	0.34	0.7331
Intercept	142	0.3368	0.4261	304	0.79	0.4299
Intercept	144	0.2583	0.4446	304	0.58	0.5618
Intercept	152	0.1005	0.5159	304	0.19	0.8457
Intercept	157	-0.08113	0.5161	304	-0.16	0.8752
Intercept	158	0.1843	0.4879	304	0.38	0.7059
Intercept	163	-0.4648	0.3693	304	-1.26	0.2091
Intercept	166	-0.1575	0.4875	304	-0.32	0.7469
Intercept	171	0.08309	0.5161	304	0.16	0.8722

The SAS output provides the group identifier (here sch\_num) the regression estimate for the group's deviation from the grand mean on the dependent variable of Intent to Major in IT, the associated standard error, degrees of freedom, as well as t-statistic and significance value for the estimate. This provides very useful information that allows

the researcher to identify those schools who are significantly higher (i.e. significant positive t-value) or lower (i.e. significant negative t-value) on average student Intent to Major in the school.

The output above is very easy to retrieve in SAS by adding the single SOLUTION parameter. Conversely, Mplus does not offer a built-in method to obtain the estimated school-level deviations on Intent to Major. As described in the syntax section above, a separate file must be saved in Mplus to provide the ability to estimate these values. When the SAVEDATA section is specified in Mplus, a separate file is saved without a very noticeable prompt to the user. The one portion of the standard Mplus output which indicates that a separate file is saved is at the bottom of the output entitled SAVEDATA INFORMATION (see Table 10). This section shows that a file was saved called mlm\_output.txt and the columns of the file are in the order and format specified. Since Mplus does not save column names to the output file, this selected output becomes very important when using this file.

Table 10. Mplus output showing format of accompanying supplemental output file.

SAVEDATA INFORMATION	
Order and format of variables	
INTENT	F10.3
ITSE	F10.3
INTEREST	F10.3
CAREER	F10.3
SCH_SIZE	F10.3
F	F10.3
B_INTENT	F10.3
B_INTENT_SE	F10.3
SCH_NUM	I4
Save file	
mlm_output.txt	

The saved mlm\_output.txt file consists of a fixed ASCII file that contains columns for each of the variables specified in the SAVEDATA information section of the standard Mplus output. The file is composed of as many lines as there are subjects in the data (e.g. in this analysis we had 309 individuals). Each line has the associated values for that individual for each of the variables specified. For each student-level variable (i.e. INTENT, ITSE, INTEREST, CAREER) the values will be specific to that individual. Conversely, for each school-level variable (i.e. SCH\_SIZE, F, B\_INTENT, B\_INTENT\_SE, SCH\_NUM) the value will be the same for each individual from that specific school; therefore, the school-level variables will repeat for each individual from that specific school.

To use the values in this file, the data should be imported into a spreadsheet program (for this example we utilized Microsoft Excel for our analysis). Since we were only interested in group level effects, we erased all duplicate group records until we had one record for each school (i.e. 40 lines for the 40 groups in this analysis). The F variable was estimated by Mplus and represents the estimate of the group deviation from the grand group mean using the proposed method. Given that we specified the variance of INTENT in the model to be zero, the B\_INTENT\_SE variable becomes the estimated standard error of the F variable as opposed to the B\_INTENT variable. By dividing the estimate (i.e. F) by its associated standard error (i.e. B\_INTENT\_SE) we are able to derive a t statistic for each of the group estimates in the file. Using this t value, we can use the two-tailed t distribution formula in Excel (T.DIST.2T) to derive a significance value for the derived t-statistic. The degrees of freedom for this formula are the number of students in the sample minus the number of variables used in the analysis, or  $309 - 5 = 304$ . Also, the absolute value of the t-statistic should be used, as the significance formula cannot utilize negative values.

Table 11 shows the calculated output along with the formulas used for derived columns included in the column name.

Table 11. Excel file showing estimate of each group (from the Mplus supplemental save file) and accompanying calculated t-statistic and p-value for each.

INTENT	ITSE	INTEREST	CAREER	SCH_SIZE	F	B_INTENT	B_INTENT_SE	SCH_NUM	t statistic (F/B_INTENT_SE)	p-value (T.DIST.2T(ABS(t statistic),df)
4	0.365	-1.414	0.554	809.948	0.038	5.552	0.319	19	0.12	0.905
7	2.7	1.05	1.052	664.948	0.318	5.74	0.262	20	1.21	0.226
5	-0.032	1.2	0.5	776.948	-0.602	4.891	0.39	22	-1.54	0.124
7	-0.014	0.8	0.812	1404.948	-0.011	5.885	0.377	24	-0.03	0.977
5	0.833	1.24	0.95	-446.052	-0.534	4.176	0.354	26	-1.51	0.132
7	0.231	0.3	-1.042	-260.052	0.948	5.777	0.335	29	2.83	0.005
7	1.528	0.95	-0.562	-283.052	0.792	5.606	0.44	30	1.80	0.073
7	-0.111	-0.267	-1	-467.052	0.483	5.18	0.461	33	1.05	0.296
7	-0.178	0.24	-0.05	-401.052	0.187	4.925	0.421	36	0.44	0.657
5	0.056	-0.425	0.969	-366.052	0.194	4.955	0.377	37	0.51	0.607
4	0.066	-0.667	0.347	-341.052	-0.079	4.698	0.292	38	-0.27	0.787
4	2.456	0.926	0.184	133.948	-0.215	4.866	0.286	40	-0.75	0.453
5	1.349	0.971	-0.429	564.948	-0.036	5.322	0.39	44	-0.09	0.927
6	-0.679	0.822	0.222	-187.052	0.319	5.195	0.365	46	0.87	0.383
4	1.211	1.16	-1.025	-415.052	-0.956	3.773	0.354	56	-2.70	0.007
4	0.333	0.2	0.375	-471.052	0.179	4.872	0.486	65	0.37	0.713
7	-1.479	1.4	-0.05	-321.052	0.227	5.017	0.281	94	0.81	0.420
5	-0.222	-0.9	0.375	1082.948	-0.042	5.648	0.486	96	-0.09	0.931
5	-0.015	-0.182	-0.33	-438.052	-0.545	4.169	0.271	97	-2.01	0.045
3	0.444	0.86	0.875	-184.052	0.072	4.949	0.354	100	0.20	0.839
4	-0.454	0.073	-0.964	-420.052	-0.01	4.716	0.39	101	-0.03	0.980
6	0.437	-1.743	-0.411	690.948	0.041	5.479	0.319	104	0.13	0.898
6	0	0	0	313.948	0.1	5.297	0.516	107	0.19	0.846
3	0.044	-0.758	0	-134.052	-0.74	4.169	0.286	109	-2.59	0.010
4	0.204	-1.167	-1.708	-227.052	-0.012	4.838	0.405	111	-0.03	0.976
4	-0.5	-0.7	0.5	-200.052	-0.192	4.675	0.486	114	-0.40	0.693
6	0	0	0	971.948	0.047	5.666	0.516	115	0.09	0.927
6	0.611	-0.2	0.562	-177.052	-0.138	4.743	0.44	116	-0.31	0.754
7	0.745	0.165	1.188	-374.052	-0.364	4.391	0.44	117	-0.83	0.409
4	0.127	-0.343	-1.321	-342.052	0.183	4.959	0.39	118	0.47	0.639
2	0.389	-0.5	-0.25	-229.052	-0.077	4.771	0.486	140	-0.16	0.874
6	-0.444	-0.4	-0.375	-385.052	0.166	4.915	0.486	141	0.34	0.733
6	2.111	1.16	-0.1	-321.052	0.337	5.126	0.421	142	0.80	0.424
7	-0.944	-0.6	0.188	455.948	0.258	5.546	0.44	144	0.59	0.558
6	0	0	0	307.948	0.1	5.293	0.516	152	0.19	0.846
4	0	0	0	-536.052	-0.081	4.571	0.516	157	-0.16	0.875
7	1.681	0.9	0.25	-511.052	0.184	4.852	0.486	158	0.38	0.705
2	0.111	-0.211	-0.25	-483.052	-0.465	4.221	0.354	163	-1.31	0.190
6	0.278	0.5	0.375	336.948	-0.157	5.054	0.486	166	-0.32	0.747
6	0	0	0	525.948	0.083	5.416	0.516	171	0.16	0.872

Upon inspection of both the SAS output for the solution for random effects in

Table 9 as well as the Mplus output with additional calculations by Excel in

Table 11, we find that schools 29, 56, and 109 are significantly different in their mean student Intent to Major in IT. Additionally, Mplus identifies an added school, 97, as significantly different. The estimates from SAS and Mplus match exactly while the standard errors and associated t-statistics are somewhat different. This difference will be examined in the discussion section. While slightly different, the results are highly consistent regarding the deviation of schools from the grand mean of Intent to Major in IT. Furthermore, the t-statistics provide directionality such that while school 29 is significantly higher in mean Intent to Major, schools 56, 97, and 109 are significantly lower.

#### 4. Discussion

The above analysis is meant as a tutorial and verification mechanism for a new method for finding individual group differences and associated significance in a multilevel model in Mplus. The estimation shows that standard output for both SAS and Mplus match up almost exactly. Furthermore, the newly devised method for estimation of individual group differences in Mplus is also highly consistent with the results from SAS's "Solution for Random Effect." Given this verification mechanism, researchers can feel confident in extending these multilevel group mean difference analyses to a MLSEM context (an example figure and syntax for a full MLSEM model using the above data is given in Appendix A).

This research uses a simple example of a random intercept model on a single dependent variable, which was needed to compare the output with SAS (as a multiple regression model only allows for one dependent variable) and to

provide simplicity in the example. However, the same proposed method can be used for random slopes as well as multiple endogenous variables, which is possible using MLSEM using Mplus. Separate “f” variables to isolate the residual are necessary for each random intercept or slope hypothesized to differ between groups in the model, which would allow for a factor score and standard error to be estimated for each residual. Once this is done, individual group differences on each estimated residual can be calculated using a spreadsheet program. This provides an abundance of opportunities to explore differences between individual groups on a number of different aspects.

While this research shows extremely similar results between SAS and Mplus, there are some slight differences between the outputs of the two programs at the group level. These differences are not with the estimated betas in the model (which match exactly) but with the standard errors associated with these estimates, which are used in constructing t-statistics for significance tests. The primary reason for these differences is in the estimation algorithm used by both SAS and Mplus. While the above analyses have specified that both software products utilize ML estimation techniques, the algorithms utilized for the ML estimation differ. SAS utilizes the Newton-Raphson (NR) algorithm while Mplus uses the Expectation Maximization (EM) algorithm. While both methods offer robust mechanisms for estimating standard errors in multilevel models, the NR algorithm has been shown to provide better estimates of these standard errors by accounting for the variance in parameter estimates (Lindstrom & Bates, 1988). The effect of the variance estimates is generally not present when the number of groups is larger, which is why research has suggested that the number of groups in a multilevel analysis should exceed 50 to combat potential bias in standard error estimates (Maas & Hox, 2005). While only 40 groups were used above, only one group deviatonal estimate (school 97) was found to differ between the two methods, and this estimate was still quite close to the traditional cutoff significance value using both SAS and Mplus ( $p = 0.07$  and  $p = 0.05$  respectively).

The proposed method of finding significant group differences within a MLSEM context using Mplus is a much needed step forward in statistical estimation and adds to the arsenal of behavioral statisticians. This builds on previous multilevel regression techniques by allowing for the simultaneous estimation of measurement and structural models as well as adding the ability to discover individual group differences. This type of analysis provides impetus for further research both statistically and in practice. First, future research can look at further verifying this method using more complex models and other statistical packages (LISRESL, R, etc.). Second, practitioners can use this research to better understand outlying groups by discovering reasons for the nature of the differences in these groups. This can provide practical benefits by allowing for intervention programs for lower-than-average groups as well as analysis of above-average groups to aid in understanding how these groups can be used to help other groups succeed.

### **Acknowledgements**

I would like to thank both Linda K. Muthen and Bengt O. Muthen for their correspondence and help with developing the above method. I would also like to thank David Peters for spurring interest in this problem.



## References

- Bentler, P. M., & Liang, J. (2003). Two-level Mean and Covariance Structures: Maximum Likelihood via an EM Algorithm. In S. P. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues, and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Chen, Y., Subramanian, S. V., Acevedo-Garcia, D., & Kawachi, I. (2005). Women's status and depressive symptoms: A multilevel analysis. *Social Science and Medicine*, *60*(1), 49-60.
- Gefen, D., Straub, D. W., & Boudreau, M. C. (2000). Structural Equation Modeling and Regression: Guidelines for Research Practice. *Communications of the Association for Information Systems*, *4*(1), 7.
- Hofmann, D. A. (1997). An Overview of the Logic and Rationale of Hierarchical Linear Models. *Journal of Management*, *23*(6), 723-744. doi: 10.1177/014920639702300602
- Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2008). A multilevel study of predictors of student perceptions of school climate: The effect of classroom-level factors. *Journal of Educational Psychology*, *100*(1), 96-104.
- Lange, K. (1995a). A gradient algorithm locally equivalent to the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 425-437.
- Lange, K. (1995b). A quasi-Newton acceleration of the EM algorithm. *Statistica sinica*, *5*(1), 1-18.
- Lent, R. W. (2005). A social cognitive view of career development and counseling. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling: Putting theory and research to work*. New York: Wiley.
- Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, *45*, 79-122.
- Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin*, *126*(2), 309-337.
- Lindstrom, M. J., & Bates, D. M. (1988). Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data. *Journal of the American Statistical Association*, *83*(404), 1014-1022.
- Luse, A., Rursch, J. A., & Jacobson, D. (forthcoming). Utilizing Structural Equation Modeling and Social Cognitive Career Theory to Identify Factors in Choice of IT as a Major. *ACM Transactions on Computing Education*.
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology*, *1*(3), 86-92.
- Muthén, B. (1994). Multilevel covariance structure analysis. *Sociological methods & research*, *22*(3), 376-398.
- Muthén, B., & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. Hox & J. K. Roberts (Eds.), *Handbook of Advanced Multilevel Analysis* (pp. 15-40). New York: Taylor and Francis.
- Muthén, L. (2012). FIML Technical Report (pp. 34).
- Muthén, L., Muthén, B., Asparouhov, T., & Nguyen, T. (2011). Mplus (Version 6.11). Retrieved from <http://www.statmodel.com/>
- Raudenbush, S. W., & Bryk, A. S. (1992). *Hierarchical Linear Models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.
- SAS. (2011). *Base SAS® 9.3 Procedures Guide*. Cary, NC: SAS Institute Inc.

Short, J. C., Piccoli, G., Powell, A., & Ives, B. (2005). Investigating Multilevel Relationships in Information Systems Research: An Application to Virtual Teams Research Using Hierarchical Linear Modeling. *Journal of Information Technology Theory and Application*, 7(3), 1-26.

Snijders, T., & Bosker, R. (1999). *Multilevel Analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage Publications.

Teachman, J., & Crowder, K. (2002). Multilevel Models in Family Research: Some Conceptual and Methodological Issues. *Journal of Marriage and Family*, 64(2), 280-294. doi: 10.1111/j.1741-3737.2002.00280.x

Appendix

Mplus Syntax for Multilevel Structural Equation Model of SCCT

The research in this paper provides a proof of concept to help verify the new method proposed for finding random effects using the developed method. The paper uses a regression example to verify the method by comparing the output to that of SAS, but the reason this method was developed was because current software does not allow these current random effects to be estimated in a multilevel structural equation model. This appendix provides both a picture of an example multilevel SEM (Figure 5) and the corresponding syntax to estimate the model, including the newly developed method for estimating random effects (Figure 6). Furthermore, the correlation matrix for the data included in this example is included to allow the user to test this method (Table 12).<sup>6</sup>

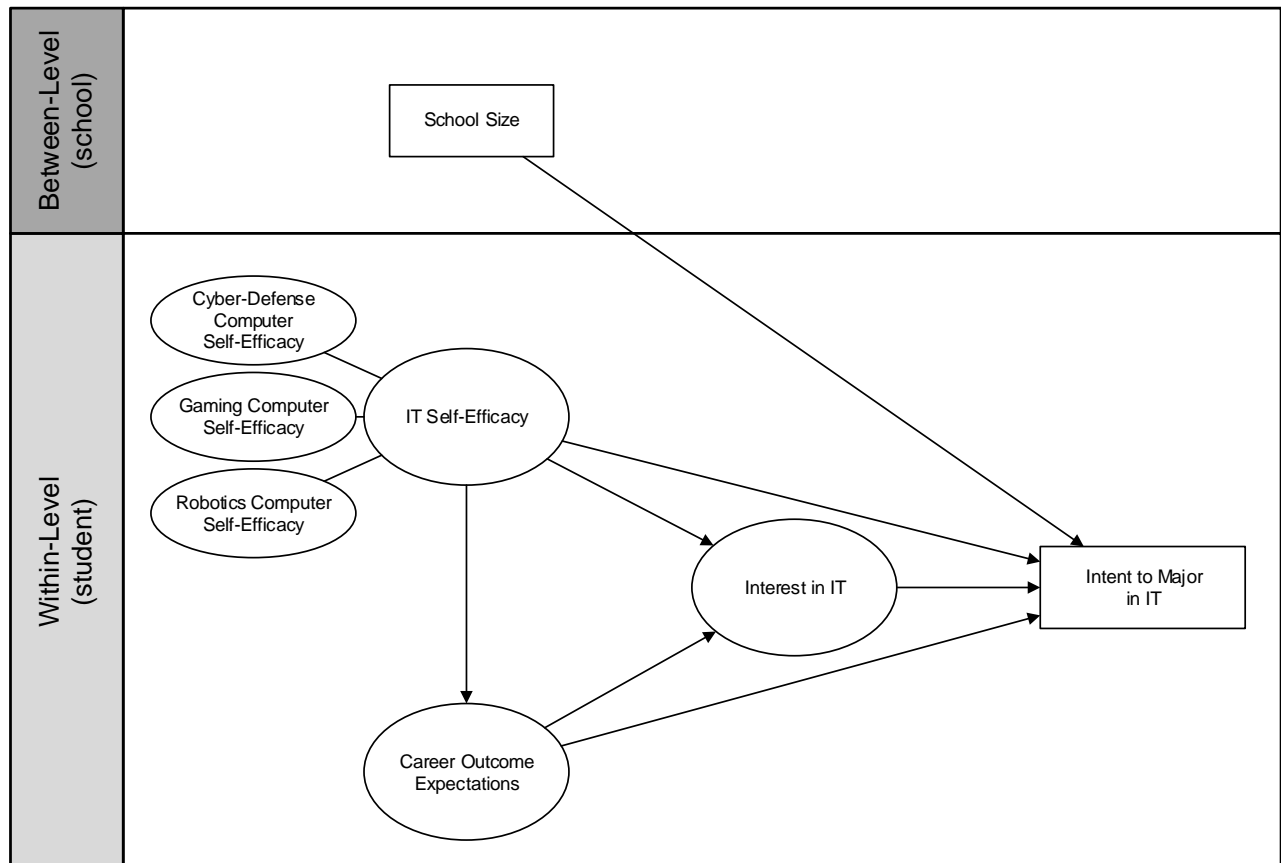


Figure 5. Hypothesized MLSEM version of SCCT model.

<sup>6</sup> It should be noted that, with a multilevel model, there are actually correlation matrices for both the within and between levels, but the variables at the within level only correlate with other within variables and between with other between variables. For this reason, only a single row has been included for the between level as there are only two variables (Intent and school size) which are used at the between level.

```

TITLE:
  SCCT MLSEM;

DATA:
  FILE IS filename.dat;

VARIABLE:
  NAMES ARE ITSE1 ITSE2 ITSE3 ITSE4 ITSE5 ITSE6 ITSE7 ITSE8 ITSE9
    Int1 Int2 Int3 Int4 Int5
    Career1 Career2 Career3 Career4
    Intent
    sch_num sch_size;

  USEVARIABLES = ITSE1 ITSE2 ITSE3 ITSE4 ITSE5 ITSE6 ITSE7 ITSE8 ITSE9
    Int1 Int2 Int3 Int4 Int5
    Career1 Career2 Career3 Career4
    Intent
    sch_size;

  WITHIN = ITSE1 ITSE2 ITSE3 ITSE4 ITSE5 ITSE6 ITSE7 ITSE8 ITSE9
    Int1 Int2 Int3 Int4 Int5
    Career1 Career2 Career3 Career4;
  BETWEEN = sch_size;
  CLUSTER = sch_num;

ANALYSIS:
  TYPE = TWOLEVEL;
  ESTIMATOR = ML;

DEFINE:
  CENTER ITSE1 ITSE2 ITSE3 ITSE4 ITSE5 ITSE6 ITSE7 ITSE8 ITSE9
    Career1 Career2 Career3 Career4
    Int1 Int2 Int3 Int4 Int5(GROUPMEAN);
  CENTER sch_size(GRANDMEAN);

MODEL:
  %WITHIN%
  ITSE_cy BY ITSE1 ITSE2 ITSE3;
  ITSE_ga BY ITSE4 ITSE5 ITSE6;
  ITSE_ro BY ITSE7 ITSE8 ITSE9;
  ITSE BY ITSE_cy ITSE_ga ITSE_ro;

  Career BY Career1 Career2 Career3 Career4;

  Interest BY Int1 Int2 Int3 Int4 Int5;

  Career ON ITSE;
  Interest ON ITSE Career;
  Intent ON ITSE Interest Career;

  %BETWEEN%
  Intent ON sch_size;

  f BY;
  Intent ON f@1;
  Intent@0;

OUTPUT:
  SAMP STANDARDIZED;

SAVEDATA:
  FILE = mlsem_output.txt;
  SAVE = FS;

```

Figure 6. Mplus syntax for multilevel SEM model of SCCT

Table 12. Correlation matrix for sample data.

	ITSE1	ITSE2	ITSE3	ITSE4	ITSE5	ITSE6	ITSE7	ITSE8	ITSE9	CAREER1	CAREER2	CAREER3	CAREER4	INT1	INT2	INT3	INT4	INT5	INTENT	SCH_SIZE
ITSE1	1																			
ITSE2	0.756	1																		
ITSE3	0.675	0.696	1																	
ITSE4	0.636	0.617	0.7	1																
ITSE5	0.604	0.592	0.657	0.892	1															
ITSE6	0.455	0.493	0.495	0.699	0.713	1														
ITSE7	0.395	0.443	0.416	0.404	0.45	0.364	1													
ITSE8	0.37	0.356	0.428	0.421	0.453	0.396	0.746	1												
ITSE9	0.288	0.32	0.357	0.363	0.404	0.336	0.671	0.629	1											
CAREER1	0.07	0.069	0.134	0.119	0.142	0.082	0.164	0.119	0.205	1										
CAREER2	0.057	0.074	0.12	0.128	0.145	0.105	0.18	0.113	0.27	0.682	1									
CAREER3	0.168	0.163	0.171	0.133	0.149	0.115	0.144	0.04	0.126	0.53	0.55	1								
CAREER4	0.099	0.129	0.184	0.156	0.164	0.065	0.178	0.082	0.208	0.565	0.548	0.551	1							
INT1	0.379	0.408	0.433	0.372	0.338	0.235	0.269	0.178	0.296	0.229	0.244	0.313	0.321	1						
INT2	0.308	0.335	0.381	0.338	0.326	0.213	0.207	0.132	0.248	0.215	0.274	0.338	0.235	0.591	1					
INT3	0.308	0.346	0.402	0.323	0.323	0.222	0.403	0.296	0.297	0.282	0.233	0.237	0.278	0.545	0.501	1				
INT4	0.328	0.283	0.365	0.337	0.311	0.183	0.316	0.23	0.3	0.21	0.214	0.243	0.271	0.603	0.577	0.565	1			
INT5	0.342	0.401	0.445	0.365	0.376	0.271	0.299	0.204	0.292	0.228	0.243	0.228	0.254	0.653	0.562	0.629	0.615	1		
INTENT	0.181	0.219	0.303	0.268	0.271	0.212	0.156	0.071	0.163	0.21	0.228	0.255	0.302	0.411	0.302	0.309	0.315	0.4	1	
SCH_SIZE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.502	1

**Author Biography**



Andy Luse is an assistant professor of management science and information systems at Oklahoma State University. He received his doctoral degrees from Iowa State University. Andy performs research in the areas of computer and network security, quantitative methods, and technology adoption.

Page intentionally left blank