

Copyright  
by  
Jason Leon Meyers  
2004

**The Dissertation Committee for Jason Leon Meyers Certifies that this is the  
approved version of the following dissertation:**

**The Impact of the Inappropriate Modeling of Cross-Classified  
Data Structures**

**Committee:**

---

Susan N. Beretvas, Supervisor

---

Laura M. Stapleton

---

Keenan A. Pituch

---

Barbara G. Dodd

---

Daniel A. Powers

**The Impact of the Inappropriate Modeling of Cross-Classified  
Data Structures**

**by**

**Jason Leon Meyers, B.A., M.A.**

**Dissertation**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Doctor of Philosophy**

**The University of Texas at Austin  
May 2004**

## **Acknowledgements**

This research was supported by a grant from the American Educational Research Association which receives funds for its “AERA Grants Program” from the National Science Foundation and the U.S. Department of Education’s National Center for Education Statistics of the Institute of Education Sciences under NSF Grant #REC-9980573. Opinions reflect those of the author(s) and do not necessarily reflect those of the granting agencies.

# **The Impact of the Inappropriate Modeling of Cross-Classified Data Structures**

Publication No. \_\_\_\_\_

Jason Leon Meyers, Ph.D.

The University of Texas at Austin, 2004

Supervisor: Susan Natasha Beretvas

Hierarchical linear modeling (HLM) is typically used in the social sciences to model data from clustered settings, such as students nested within classrooms. However, not all multilevel data are purely hierarchical in nature. For example, students can be nested within the neighborhoods in which they live and within the schools they attend. But, most likely, students from a given neighborhood do not all attend the same school and students from a given school do not all reside within the same neighborhood. Because neighborhoods are not nested within schools nor vice-versa, the two are said to be cross-classified. Cross-classified random effects modeling (CCREM) is used to model data from these non-hierarchical contexts.

While use of CCREM has increased in various disciplines such as medicine, it is seldom used in educational research. CCREM is mentioned in most multilevel modeling textbooks (for example, Raudenbush & Bryk, 2002; Hox, 2002; Snijders & Boskers, 1999). However, it remains infrequently used, most likely because the models are technically sophisticated and can be somewhat difficult to interpret.

Little research has been conducted assessing when it is necessary to use CCREM, so this dissertation involved several studies. A Monte Carlo Simulation Study was conducted in order to investigate potential factors affecting the need to use CCREM as well as the impact of ignoring cross-classification. As a follow-up study, CCREM was applied to a large-scale national data set in order to provide insight into the potential effects of ignoring the cross-classified data structure.

Results of both studies indicated that when using HLM instead of CCREM, the fixed effect estimates were unaffected but the standard error estimates associated with the variables modeled incorrectly were biased. In addition, the estimates of the variance components displayed bias. The observed bias was related to the proportion of the total variance that was between each cross-classified factor, the sample size, and the similarity of the cross-classified factors. Implications and limitations are discussed and suggestions for future research are presented.

## Table of Contents

List of Tables.....	x
List of Figures.....	xii
Chapter 1 Introduction .....	1
Chapter 2 Literature Review .....	4
Introduction to Hierarchical Linear Modeling .....	4
Methods for Hierarchical Data Analysis and their Shortcomings.....	5
Reasons to Use Hierarchical Linear Modeling .....	7
HLM-- An Extension of the Linear Model .....	9
Fully Unconditional Model .....	10
Adding a Level-1 Predictor .....	12
Adding a Level-1 and a Level-2 Predictor .....	13
A Three-Level Model.....	15
Cross-Classified Data .....	17
Cross-Classified Random Effects Modeling (CCREM) in the Literature- Appropriate Use .....	21
Ignoring Cross-Classified Structures in the Literature.....	24
Cross-Classified Random Effects Modeling-Fully Unconditional Model ...	26
Cross-Classified Random Effects Modeling-Adding Predictors .....	28
Statement of the Problem .....	31
Chapter 3 Method.....	34
Study 1.....	34
Monte Carlo Simulation .....	34
Conditions .....	34
Correlation of Residuals at Level 2.....	35
Number of Feeder Schools .....	35

Number of Levels of Cross-Classified Factors .....	36
Average Middle School Size.....	36
Intra-unit Correlation Coefficients (IUCCs) .....	37
Data Generation.....	39
Generating the combinations of middle school and high school ...	40
Analysis.....	45
Study 2.....	49
Data .....	49
Variables.....	49
Dependent Variable.....	50
Predictor Variables.....	50
Analysis.....	51
Chapter 4: Results .....	56
Study 1.....	56
Fixed Effect Estimates .....	56
X, The Student-Level Predictor .....	56
Relative Bias .....	56
Standard Error Bias .....	56
Z, The Middle School-Level Predictor .....	60
Relative Bias .....	60
Standard Error Bias .....	60
W, The High School-Level Predictor.....	65
Relative Bias .....	65
Standard Error Bias .....	65
Random Effect Estimates .....	68
Variance Between Students.....	68
Relative Bias .....	68
Standard Error Bias .....	69



Variance Between High Schools.....	73
Relative Bias .....	73
Standard Error Bias .....	73
Random Effects when the correlation is .40 .....	77
Fit Indices.....	80
Study 2.....	83
Descriptive Statistics .....	83
Fixed Effect Estimates .....	84
Fixed Effect Standard Error Estimates.....	85
Random Effect Estimates .....	86
Random Effect Standard Error Estimates .....	87
Fit Indices.....	87
Chapter 5 Discussion.....	89
Summary .....	89
Limitations and Directions for Future Research .....	94
Educational Importance and Conclusion.....	97
Appendix A .....	101
Appendix B .....	111
References .....	121
Vita .....	125

## List of Tables

Table 1:	Hierarchical Clustering of Students within Schools, Within Neighborhoods .....	19
Table 2:	Cross-Classified Clustering of Students within Schools, Within Neighborhoods .....	20
Table 3:	Simulation Design Conditions .....	38
Table 4:	Descriptive Statistics for the Fixed Effect Coefficient, $\gamma_{10}$ , associated with the student-level variable .....	58
Table 5:	Bias Statistics for the Fixed Effect Coefficient, $\gamma_{10}$ , associated with the student-level variable .....	59
Table 6:	Descriptive Statistics for the Fixed Effect Coefficient, $\gamma_z$ , associated with the middle school-level variable .....	63
Table 7:	Bias Statistics for the Fixed Effect Coefficient, $\gamma_z$ , associated with the middle school-level variable .....	64
Table 8:	Descriptive Statistics for the Fixed Effect Coefficient, $\gamma_w$ , associated with the high school-level variable .....	66
Table 9:	Bias Statistics for the Fixed Effect Coefficient, $\gamma_w$ , associated with the high school-level variable .....	67
Table 10:	Descriptive Statistics for the Estimates of the Variance Between Students in Conditions when correlation =0 .....	71
Table 11:	Bias Statistics for the Estimates of the Variance Between Students in Conditions when correlation =0 .....	72

Table 12:	Descriptive Statistics for the Estimates of the Variance Between High Schools in Conditions where Correlation =0 .....	75
Table 13:	Bias Statistics for the Estimates of the Variance Between High Schools in Conditions where Correlation =0 .....	76
Table 14:	Descriptive Statistics for the Estimates of the Variance Between Students in Conditions where Correlation = .40 .....	78
Table 15:	Descriptive Statistics for the Estimates of the Variance Between Students in Conditions where Correlation =.40 .....	79
Table 16:	Means (and Standard Deviations) for Goodness of Fit Indices .....	82
Table 17:	Descriptive Statistics of variables included in analysis .....	83
Table 18:	Fixed Effect Estimates .....	85
Table 19:	Fixed Effect Standard Errors.....	86
Table 20:	Random Effect Estimates.....	86
Table 21:	Random Effect Standard Errors .....	87
Table 22:	Fit Indices.....	88

## List of Figures

Figure 1:	Two Feeder School Condition.....	43
Figure 2:	Three Feeder School Condition.....	44

## Chapter 1: Introduction

Hierarchical linear modeling (HLM) techniques are now commonly used to model the clustering inherent in educational research settings (such as students within schools). In addition, the range of possible models has quickly expanded. One of the more complicated additions involves the modeling of what are termed cross-classified random effects. While many data structures involve such cross-classification of units, cross-classified random effects modeling (CCREM) techniques are seldom used.

In a purely nested design, students attending a set of schools all live in the same neighborhood and all students in a given neighborhood attend the same set of schools. In reality, nested data is not always purely hierarchical. For example, it is often the case that students in a given neighborhood attend several different schools, and that students at a given school come from a variety of neighborhoods. In situations such as this, students are nested within schools and within neighborhoods, yet schools can no longer be considered as nested within neighborhoods (nor vice-versa). This data structure provides an example of cross-classified data where students are cross-classified by school and by neighborhood.

Longitudinal data sets frequently provide CREM data structures. For instance, a researcher might be interested in modeling change in students' test

scores over time. Typically, students attend varied combinations of middle schools and high schools. All students at a given middle school do not subsequently attend the same high school, and students at a given high school do not all come from the same middle school. Therefore, students are nested within a cross-classification of middle schools and high schools. Characteristics of both types of schools likely contribute to the variance in high school test scores and, thus, CCREM should be used for analysis. Many other examples are well suited to the use of CCREM.

While the use of CCREM has increased in recent years in fields such as medicine (Rasbash and Browne, 2001) and in sociometric research (Hox, 2002; Snijders & Boskers, 1999), it is still only rarely used in educational research. Use of CCREM is now mentioned in most multilevel modeling textbooks (for example, Raudenbush & Bryk, 2002; Hox, 2002; Snijders & Boskers, 1999) although it is still infrequently used. The technical sophistication of the model seems to discourage its common use. Of the commonly used multilevel modeling programs, only MLwiN (Rasbash et. al, 2000) and aML (Lillard & Panis, 2000) can fit these models, although future versions of HLM (Raudenbush, Bryk, Cheong, & Congdon, 2000) will include this capability. SAS PROC MIXED can also be used to perform CCREM (Yang, 2003). Applied researchers need to

understand better when CCREM data structures should not be ignored and what the resulting impact could be if they are.

Since little research has been conducted assessing when it is necessary to use CCREM, this dissertation involved several studies. A Monte Carlo Simulation Study was conducted in order to investigate potential factors affecting the need to use CCREM as well as the impact of ignoring cross-classification. As a follow-up study, CCREM was applied to a large-scale national data set in order to provide insight into the potential effects of ignoring the cross-classified data structure.

## **Chapter 2: Literature Review**

### **INTRODUCTION TO HIERARCHICAL LINEAR MODELING (HLM)**

Hierarchical Linear Models are alternatively known as multilevel models, mixed effects-models, random-effects models, random-coefficient regression models and covariance-component models in various disciplines (Raudenbush & Bryk, 2002). Although called by many names, these statistical techniques have a common use. They are frequently used in the social sciences to model data that are collected from clustered or nested settings.

In educational research, students who are nested within the same context, such as a classroom or school, should not be considered independent. Students sharing the same setting likely share some characteristics that influence the dependent variable of interest. For example, suppose a researcher is interested in understanding and explaining variation in students' scores on a hypothetical statewide high stakes test (SHST). In addition to being used at various grade levels, satisfactory performance on the SHST at Grade 11 is prerequisite to a high school diploma. To explain the variation in students' performance on this test, the researcher would likely want to explore relationships between test scores and student-level variables, such as socioeconomic status (SES) as well some



variables describing the contexts in which the various students learn, such as the student-teacher ratio. If the data set includes both individual and group-level variables, it can be analyzed in several ways. These methods will be described in the next section.

### **METHODS FOR HIERARCHICAL DATA ANALYSIS AND THEIR SHORTCOMINGS**

Aside from multilevel modeling, two alternative data analysis strategies exist for clustered data, but they present a host of problems. If one has information at both an individual and group level, one might consider either aggregating or disaggregating the data. Aggregating the data involves summarizing across micro-units (e.g. students) within macro-units (e.g. classrooms) and conducting analyses at the macro-unit level.

Aggregation has several shortcomings. The first problem is that the interpretations suffer from a shift of meaning. Conclusions from aggregated data cannot be used to make inferences about the individual-level data. For instance, one might find that classrooms with a high number of students perform more poorly, on average, on the SHST than classrooms with a smaller student-teacher ratio. However, no student-level inferences can be made. Along these same lines, aggregating data can lead to what is known as the “ecological fallacy.” This means, basically, that relationships among macro-units (such as classrooms)

cannot be applied to the individual units (students). For example, the percentage of minority students in a classroom could be related to the average SHST score. This however, gives us no information about the micro-level relation between race and SHST score. The percentage of minority students is a variable that means something for the classroom and this meaning is distinct from the meaning of minority as an individual-level variable. Finally, aggregating data prevents the examination of potential cross-level interactions. For example, the relationship between students' grade point averages (GPAs) and SHST scores might be different for individuals in classrooms with a good supply of educational materials than for individuals in classrooms with sparse supplies. Aggregated data cannot be used to model this type of relationship (Snijders & Boskers, 1999).

Disaggregation is the direct opposite of aggregation. Here, data are decomposed to the individual-level. Suppose for example, 100 classrooms are sampled, and 10 students are sampled from each classroom. Disaggregation involves treating the data as 1000 observations and ignoring the within-classroom clustering. This technique has severe pitfalls. First, disaggregating and treating the observations within classrooms as independent implies that the sample size of the data being analyzed will be dramatically exaggerated. One might disaggregate to the level of the student and estimate the relationship between a characteristic of the classroom (e.g. size) and a dependent variable such as SHST

score. Disaggregating is like assuming there are 1000 independent observations when, in fact, there are only 100 independent observations (the 100 classrooms). When studying between-group differences, this can lead to serious inflation of type I error rates. For example, one might conclude that the smaller the classroom is, the higher the SHST scores are, when in the population there are no differences based on classroom size (Snijders & Bosker, 1999). On the other hand, when studying within-group differences, disaggregation can lead to the use of unnecessarily conservative tests (i.e. low statistical power) (Snijders & Bosker, 1999). In addition, research has repeatedly shown that the larger the contextual effects (for example, the effects of classroom characteristics), the more HLM is needed over traditional ordinary least squares models (Roberts, 2000; Kreft & de Leeuw, 1998; Snijders & Bosker, 1999). As Snijders and Boskers conclude in their book, “if the macro-units have any meaningful relation with the phenomenon under study, analyzing only aggregated and disaggregated data is apt to lead to misleading and erroneous conclusions (Snijders & Boskers, 1999, p. 16).”

Appropriate use of hierarchical linear modeling is not plagued by the same problems from which aggregation and disaggregation suffer.

### **REASONS TO USE HIERARCHICAL LINEAR MODELING**

There are many reasons why HLM is preferable to disaggregating or aggregating in situations with nested data, such as the students-within-classrooms

example. First of all, if there is variability between units at several different levels, such as the student and classroom levels, ignoring the source of the variability could lead one to draw incorrect conclusions from the data regarding the statistical significance of certain variables. In addition, clustered data have inherent dependencies. Within a given cluster, units within that cluster, such as students within classrooms, should not be considered independent. In the example, students within a classroom share many similarities. For instance, students from one classroom might have access to fewer educational materials and be taught by a more inexperienced teacher than students in another classroom in the study. These students likely share common causes not included in the model. HLM models these common causes by allowing the intercept and effect of the explanatory variables to vary across higher-level units. Violation of the assumption of independence often results in inflated effect size estimates and spuriously small standard errors for the parameter estimates because variation at both the student and classroom levels is not included in the model (Snijders & Bosker, 1999). Use of HLM permits the appropriate modeling of such dependencies, compensating for the violation of the assumption of independence (Snijders & Bosker, 1999). Conceptually, HLM is nothing more than a simple extension of the linear model. This extension will be described in the next section.

## HLM-- AN EXTENSION OF THE LINEAR MODEL

HLM extends the linear regression model by incorporating random effects at multiple levels. Suppose that students are *not nested* within a shared context, such as a classroom. If a researcher were simply interested in using GPA to predict SHST score, both variables would be student-level variables and traditional ordinary least squares (OLS) regression would be appropriate. The regression equation would be represented as:

$$y_i = \beta_0 + \beta_1 X_{1i} + r_i \quad (1)$$

where  $y_i$  is the outcome variable SHST score for student  $i$ ,  $X_{1i}$  represents a student's GPA,  $\beta_0$  is the expected SHST score for an individual with a grade point average of zero,  $\beta_1$  is the expected change in SHST score associated with a one point increase in  $X_1$  (GPA), and  $r_i$  is the residual term associated with student  $i$ . The residual term is assumed to be independently and normally distributed with a mean of zero and a constant variance,  $r \sim N, (0, \sigma^2)$ .

Continuing on with our same example, if students *are nested* within classrooms and the SHST scores vary across these classrooms, use of OLS regression would not be appropriate. In addition, the assumption of independence could not be met for multiple students sampled from the same classroom if variation in SHST scores between classrooms exists.

Using multilevel modeling allows the researcher to partition the variance in the outcome variable (here, SHST score) at the group level and individual level. This partitioning allows the regression equation to vary across higher-level clustering units (here, classrooms) rather than being assumed fixed for all macro-level units, as is the case in traditional OLS regression. The basic types of two-level hierarchical linear models will now be described.

### **FULLY UNCONDITIONAL MODEL**

Typically, the first step in fitting a hierarchical linear model is to run a fully unconditional model to determine whether or not it is necessary to partition the variance in the outcome variable into the two levels. The fully unconditional model is one in which no individual or group-level variables are included as predictors.

In the fully unconditional model, the student-level regression equation, known as the level-1 model, is defined as follows. (It should be noted that the notation that follows matches that of growth curve modeling with HLM. This is done to ease the transition from the 2-level to the 3-level model and to correspond with the notation of Raudenbush & Bryk, 2002). The level-1 unconditional model is as follows:

$$y_{ij} = \pi_{0j} + e_{ij} \quad (2)$$

where  $y_{ij}$  is the SHST score for student  $i$  in classroom  $j$ ,  $\pi_{0j}$  is the mean SHST score for classroom  $j$ , and  $e_{ij}$  is the level-1 residual representing the deviation of student  $i$ 's score from classroom  $j$ 's mean SHST score,  $e_i \sim NI(0, \sigma^2)$ . The classroom-level model, or level-2 model would be as follows:

$$\pi_{0j} = \beta_{00} + r_{0j} \quad (3)$$

where  $\beta_{00}$  is the grand mean SHST score in the population and  $r_{0j}$  is the random effect of classroom  $j$  that is assumed to be normally and independently distributed with a mean of zero and a constant variance,  $\tau_{00}$ . The intraclass correlation coefficient (ICC) is often used to assess how much of the variance in the outcome variable is between level-2 units. The intraclass correlation coefficient,  $\rho$ , is defined as the ratio of between-group variance to total variance in the outcome measure. Specifically,  $\rho = \tau_{00}/(\tau_{00} + \sigma^2)$ , where  $\tau_{00}$  is the between-group variability and  $\sigma^2$  is the within-group variability. The larger  $\rho$  is, the stronger the dependency between groups and the more important it is to take into consideration the hierarchical nature of the data. A  $\rho$  value of zero indicates that none of the observed variability is between level-2 units, and it is not necessary to use HLM. In such a scenario, use of OLS regression would be appropriate. The size of  $\tau_{00}$ , the variance of the random effect, indicates to the researcher whether or not there is substantial variability across classrooms.

The level-1 and level-2 models can be combined into a singular regression equation as follows:

$$y_{ij} = \beta_{00} + r_{0j} + e_{ij} \quad (4)$$

### **ADDING A LEVEL-1 PREDICTOR**

Usually, the next step in modeling an example like this is to add person-level (level-1) explanatory variables to help explain variability between individuals. The variables, as in standard OLS regression, can be either categorical dummy coded variables, or continuous variables. In the current example, grade point average (GPA) could be added to the model. Equation 2 then becomes:

$$y_{ij} = \pi_{0j} + \pi_{1j}X_{1ij} + e_{ij} \quad (5)$$

where  $\pi_{0j}$  is now the predicted SHST score for classroom  $j$  when  $X_1$  (here, GPA) is zero,  $\pi_{1j}$  is the expected change in SHST score associated with a one point increase in GPA,  $X_{1ij}$  is the GPA for person  $i$  in classroom  $j$ , and  $e_{ij}$  is the level-1 residual. The level-2 equations can be modeled without predictors, such that Equation 3 becomes :

$$\begin{cases} \pi_{0j} = \beta_{00} + r_{0j} \\ \pi_{1j} = \beta_{10} + r_{1j} \end{cases} \quad (6)$$



where the intercept,  $\beta_{00}$ , is now the average predicted SHST score across classrooms when GPA = 0,  $r_{0j}$  is the classroom-level intercept residual,  $\beta_{10}$  is the average coefficient representing the relationship between GPA and SHST score (the GPA “slope”) across classrooms, and  $r_{1j}$  is the classroom-level slope residual. In this model, we are allowing the intercepts (predicted SHST scores when GPA = 0) and slopes (relationships between GPA and scores) to vary across classrooms. Therefore, the effects of  $\pi_{0j}$  and  $\pi_{1j}$  are considered *random effects*, since they are allowed to vary across level-2 units. It is possible under HLM to model some or all effects as fixed, thereby restricting the intercept and/or slope to be constant across groups.

The symbols representing the relevant elements of the variance/covariance matrix for the level-2 residuals are as follows:

$$Var \begin{bmatrix} r_{0j} \\ r_{1j} \end{bmatrix} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} = \mathbf{T} \quad (7)$$

### **ADDING A LEVEL-1 AND A LEVEL-2 PREDICTOR**

To explain any variability between classrooms in the intercept and/or slope parameters, predictors can be added to the level-2 model. Several predictors can be added to a given model, but for the sake of simplicity, we will include only one predictor at each level. The level-1 equation will remain exactly as it was in

the previous model that only included a level-1 predictor (see Equation 5). The level-2 classroom equations will now include an explanatory variable to predict SHST score. In our example, we will add the teacher's number of years teaching experience,  $W_{1j}$ , as the level-2 predictor. The level-2 equations become:

$$\begin{cases} \pi_{0j} = \beta_{00} + \beta_{01}W_{1j} + r_{0j} \\ \pi_{1j} = \beta_{10} + \beta_{11}W_{1j} + r_{1j} \end{cases} \quad (8)$$

where  $W_{1j}$  is the value associated with the predictor variable for classroom  $j$ ,  $\beta_{00}$  becomes the predicted SHST score for a classroom when GPA =0 and with the teacher having zero years of experience,  $\beta_{01}$  is the amount of change in SHST score with a one point increase in years of teaching experience,  $\beta_{10}$  represents the predicted GPA –SHST score slope when  $W_{1j}=0$ , and  $\beta_{11}$  is the amount of change in the GPA-SHST score slope associated with a one point increase in years of teaching experience with  $r_{0j}$  and  $r_{1j}$  as the residuals for the intercept and slope for classroom  $j$ .

Again, the above equations can be combined into a single equation as follows:

$$Y_{ij} = \beta_{00} + \beta_{01}W_{1j} + \beta_{10}X_{1ij} + \beta_{11}W_{1j}X_{1ij} + r_{0j} + r_{1j}X_{1ij} + e_{ij} \quad (9)$$

Next, the addition of a third level of clustering to the model will be described.

## A THREE-LEVEL MODEL

Multilevel models, as their name implies, can have any number of levels. In addition to the two-level model, three-level models are also commonly used. In our example, students are nested within classrooms. To extend the modeling further, we could model the classrooms as nested within schools. Now, we have students at level 1, classrooms at level 2, and schools at level 3. Given there is considerable variability between both classrooms and schools, this is a viable modeling option. The process is the same with three-level models as it is with two-level models. To continue with our example, let us consider the case with a predictor variable at each of the three levels including: GPA at level 1, number of years of teaching experience at level 2, and percentage of students per school in the free or reduced lunch program at level 3. The level-1 and level-2 equations would be the same as in the previous example although including the level-3 (representing, here, schools) subscript  $k$ :

The Level-1 Equation with GPA,  $X_I$  included, becomes:

$$y_{ijk} = \pi_{0jk} + \pi_{1jk}X_{Iijk} + e_{ijk} \quad (10)$$

and at level 2 with teaching experience,  $W$ , added as a classroom-level predictor,

Equation 8 becomes:

$$\begin{cases} \pi_{0jk} = \beta_{00k} + \beta_{01k}W_{1jk} + r_{0jk} \\ \pi_{1jk} = \beta_{10k} + \beta_{11k}W_{1jk} + r_{1jk} \end{cases} \quad (11)$$

The level-3 equations take on the following form:

$$\begin{cases} \beta_{00k} = \gamma_{000} + \gamma_{001}Z_{1k} + u_{00k} \\ \beta_{01k} = \gamma_{010} + \gamma_{011}Z_{1k} + u_{01k} \\ \beta_{10k} = \gamma_{100} + \gamma_{101}Z_{1k} + u_{10k} \\ \beta_{11k} = \gamma_{110} + \gamma_{111}Z_{1k} + u_{11k} \end{cases} \quad (12)$$

where  $Z_{1k}$  is the value of the predictor variable, free lunch for school  $k$ ,  $\gamma_{000}$  is the predicted SHST score for a school with 0% of the students in the free lunch program given  $W_{1jk} = 0$  and  $X_{ij} = 0$ ,  $\gamma_{001}$  is the change in a school's predicted SHST score associated with a one-point increase in the percentage of students on the free lunch program when  $W_{1jk} = 0$  and  $X_{ij} = 0$ ,  $\gamma_{010}$  is the expected teaching experience slope for a school with 0% of the students in the free lunch program and  $X_{ij} = 0$ ,  $\gamma_{011}$  is the change in this slope associated with a one-point increase in the percentage of students on the free lunch program,  $\gamma_{100}$  is the predicted slope representing the relationship between GPA and SHST score when the values of  $Z$  and  $W$  are zero,  $\gamma_{101}$  is the predicted change in this slope associated with a one point increase in GPA,  $\gamma_{110}$  is the predicted coefficient representing the relationship between teaching experience and the slope for GPA, and  $\gamma_{111}$  is the change in that slope given a one point increase in free lunch. The school-level descriptor,  $Z_{1k}$  has been included in all of the level-3 equations to help explain

variation in the intercepts and slopes at level 2. Again, this model can be written as a single equation:

$$\begin{aligned}
 Y_{ijk} = & \gamma_{000} + \gamma_{001}Z_{1k} + u_{00k} + \gamma_{010}W_{1jk} + \gamma_{011}Z_{1k}W_{1jk} + W_{1jk}u_{01k} + \\
 & r_{0jk} + \gamma_{100}X_{1ijk} + \gamma_{101}Z_{1k}X_{1ijk} + X_{1ijk}u_{10k} + \gamma_{110}W_{1jk}X_{1ijk} + \gamma_{111}Z_{1k}W_{1jk}X_{1ijk} + \\
 & u_{11k}W_{1jk} + r_{1jk}X_{1ijk} + e_{ijk}
 \end{aligned} \tag{13}$$

Thus far, the data from clustered settings being described have been hierarchical. However, the structure of real data is not always so simple. Often, data from clustered settings are not hierarchical, but *cross-classified*. These data structures are described in the next section.

## CROSS-CLASSIFIED DATA

In the previous examples, the data were purely hierarchical. Students were nested within one and only one classroom, which was nested within one and only one school. In practice, the hierarchical structure of the data is not always so simple. For example, students might be nested within the schools they attend and within the neighborhood in which they live and by and large, students attend schools in their neighborhood. However, there are exceptions. Students often attend schools in different neighborhoods than those in which they reside, and schools often attract students from several different neighborhoods. This is especially true with schools that are near neighborhood borders (Hox, 2002). In

an example like this, students are nested within neighborhoods and they are nested within schools. But, neighborhoods and schools are said to be *crossed* with each other, because the nesting structure is ambiguous. In a purely nested design, students attending a set of schools all live in the same neighborhood. For instance, Table 1 depicts the scenario in which 16 students are sampled from each of four neighborhoods. Here, students who attend the same school inhabit the same neighborhood. The structure in Table 1 contains data in which schools are nested within neighborhoods: a three-level hierarchy. Yet, the true nestedness depicted here is unrealistic.

The structure depicted in Table 2 also has 16 students sampled from each of four neighborhoods, but students at the same school do not necessarily live in the same neighborhood. In the structure in Table 2, students are nested within schools and within neighborhoods, yet schools can no longer be considered as nested within neighborhoods (nor vice-versa). If both neighborhood and school are considered to be important sources of variance in, say, SHST scores, a cross-classified multilevel model is appropriate.

*Table 1. Hierarchical Clustering of Students Within Schools, Within Neighborhoods*

School	Neighborhood			
	A	B	C	D
1	XXXXXX			
2	XXX			
3	XXXXX			
4	XX			
5		XXXXXXXXXX		
6		XXXXXXX		
7			XXX	
8			XXXXXXXX	
9			XXXXXX	
10				XXXXXXXX
11				XXXXXXXX

*Note.* “X” represents a student. There are 16 students sampled from each neighborhood and students from the same school live in the same neighborhood.

*Table 2. Cross-Classified Clustering of Students Within Schools, Within Neighborhoods*

School	Neighborhood			
	A	B	C	D
1	XXXX			XXX
2	XX	XX		
3	XXXXX			XX
4	XXX	X	XX	
5		XXXX		
6		XXX		
7			XXX	
8		XX	XXX	X
9		XXX	XXXX	XX
10	X		X	XXX
11	X	X	XXX	XXXXX

*Note.* “X” represents a student. There are 16 students sampled from each neighborhood and students from the same school do not necessarily live in the same neighborhood.



Another typical example of cross-classified multilevel models involves a longitudinal or growth curve design. A researcher might be interested in tracking students' SHST score over time. The SHST is first administered in third grade and at intervals through eleventh grade. Typically, students attend one elementary school, a different junior high or middle school, and a different high school. To model change in test scores over time, either classroom or school-level variables would be needed. Since all students from one middle school do not attend the same high school and because all students at a given high school do not come from the same middle school, a cross-classified data structure exists (Goldstein, 1995). Students are nested within a cross-classification of middle school and high schools. It is likely that characteristics of both types of schools contribute to the variance in test scores and, thus, a cross-classified random effects model (CCREM) should be used for analysis. The next section will describe examples using CCREM.

#### **CROSS-CLASSIFIED RANDOM EFFECTS MODELING (CCREM) IN THE LITERATURE – APPROPRIATE USE**

Cross-classified random effects modeling is relatively uncommon in educational research, perhaps due to its complex nature. A search of recent educational research (1990-2003) using the ERIC database indicated that only a

small number of studies have employed cross-classified multilevel modeling techniques.

One of the better known studies was conducted by Raudenbush (1993) who re-analyzed the data from a previous study he had conducted (Garner and Raudenbush, 1991). In these studies, the authors tested for neighborhood and school effects on educational attainment for students living in one school district in Scotland. Because students from one neighborhood attended a host of schools and because students at a school came from multiple neighborhoods, students were cross-classified by schools and neighborhoods. In the original study (Garner and Raudenbush, 1991), the authors ignored the variance between schools. However, in the re-analysis using a cross-classified multilevel model, the author estimated both the between-school and between-neighborhood variance and both studies indicated that deprivation in the neighborhood setting had an effect on educational attainment.

Goldstein and Sammons (1997) studied the influence of secondary and junior schools on exam performance. Because the “junior schools” do not directly feed into secondary schools, students are cross-classified by junior and secondary schools. Not everyone at a particular secondary school came from the same junior schools, and not everyone at a particular junior school subsequently attended the same secondary schools. Goldstein and Sammons found that junior

schools with high average test scores for students when they leave junior school also tend to have high average scores for their students at age sixteen.

In O’Muircheartaigh and Campanelli’s study (1999), the authors attempted to separate the variability in survey response rate due to interviewers from that attributable to primary sampling unit (PSU), a contextual “area” variable. A given household was nested within a combination of neighborhood and interviewer. However, because interviewers worked in several PSU’s and because each PSU was serviced by several interviewers, households were cross-classified by interviewer and primary sampling unit (PSU). There was not an unambiguous nesting of interviewers and PSU’s. The results indicated that the response rate was due more to the influence of the interviewer than the influence of the area, but since both were of interest to the researcher, a cross-classified multilevel model was used.

Lastly, Teitler and Weiss (2000) used cross-classified models to investigate between-neighborhood and between-school variation in the timing of a child’s transition to first sexual intercourse. Again, students were nested within a cross-classification of neighborhood and school. They found that “acceptable transition ages tend to be lower among African American schools than among white schools” (Teitler and Weiss, 2000, p. 127). The authors also found that when school or neighborhood variability existed, it was only among

predominantly white schools. Their data indicated that sexual initiation behaviors did not significantly vary among schools or neighborhoods that were predominantly African American even when comparing extremely poor neighborhoods to middle class neighborhoods (Teitler and Weiss, 2000). Because the researchers were interested in both school and neighborhood-level variables and because of the unambiguous nesting structure of the data, cross-classified multilevel modeling was necessary.

Several studies were found which might have benefited from the use of CCREM, and they will be discussed in the next section.

#### **IGNORING CROSS-CLASSIFIED STRUCTURES IN THE LITERATURE**

The studies described above provide insight into the potential uses of, and need for, methods to analyze cross-classified data. As these methods are relatively new, there are many instances in which researchers have conducted analyses ignoring the cross-classified structure of the data. One such example was provided by Ainsworth (2002). In this study, Ainsworth addressed the impact of neighborhood characteristics on educational achievement by using the National Educational Longitudinal Study (NELS: 88). In assessing change over time, Ainsworth performed “analyses on those students who have persisted in the same neighborhood over time (1988-1992)” (Ainsworth, 2002, p.22). This

analysis strategy results in a loss of information and generalizability. Any students who changed neighborhoods were excluded from the analysis. These students could have provided additional insight into the phenomenon of interest. Ainsworth could have used a cross-classified random effects model allowing the researcher to keep the information about students who changed neighborhoods in the analysis. Observations could have been modeled as nested within individuals who were, in turn, nested within a cross-classification of neighborhoods over time. Use of a cross-classified multilevel model could have led to results that were more informative and perhaps less spurious.

Ma and Wilkins (2002) analyzed data from the National Survey of American Youth to determine the effect of school-level variables, such as school size and school mean SES, on the growth of student science achievement. The study began when the students were in seventh grade and concluded when the students were in twelfth grade. In this study, the authors ignored the fact that students most likely change schools between seventh and twelfth grade, as most American students transition from middle school to high school between 8<sup>th</sup> and 9<sup>th</sup> grades. The authors could have modeled the test score as being nested within individuals, which were in turn nested within a cross-classification of middle school and high school. Ma and Wilkins, in ignoring the cross-classified data structure, also neglected that school-level characteristics, such as school size,

might change when the individual changes school. By not modeling the cross-classification of middle school and high school, the authors might have reached spurious conclusions. The next section will describe cross-classified multilevel modeling.

### **CROSS-CLASSIFIED RANDOM EFFECTS MODELING -- FULLY UNCONDITIONAL MODEL**

Modeling of cross-classified hierarchical structures closely parallels the modeling of purely hierarchical structures. The neighborhood and school cross-classified example will be used to demonstrate a typical model. The unconditional model includes no predictor variables and would be used for estimating the variance between neighborhoods, between schools, and within “cells.” Cells are combinations of the crossed factors (here, schools and neighborhoods). The within-cell model is represented as:

$$Y_{i(jk)} = \beta_{0(jk)} + e_{i(jk)} \tag{14}$$

where  $Y_{i(jk)}$  is the SHST score for child  $i$  in neighborhood  $j$  and school  $k$ ,  $\beta_{0(jk)}$  is the mean SHST score for students who live in neighborhood  $j$  and attend school  $k$ , and  $e_{i(jk)}$  is the random student effect, or the deviation of the student’s test score from the cell mean. These effects,  $e_{i(jk)}$ , are assumed to be normally and independently distributed with a mean of 0 and a constant variance,  $\sigma^2$ . It should

be noted that the parentheses are used for the subscripts ( $jk$ ) to indicate they conceptually describe cross-classified factors at the same level (see Hox, 2002, p. 125) The level-2 model partitions the variation between cells into the components due to neighborhood effects, school effects, and the interaction between the two factors:

$$\beta_{0(jk)} = \gamma_{00} + b_{00j} + c_{00k} + d_{0(jk)} \quad (15)$$

where  $\gamma_{00}$  is the average SHST score of all students,  $b_{00j}$  is the random effect of neighborhood  $j$ , which is assumed to be normally distributed with a mean of 0 and a constant variance  $\tau_{b00}$ ,  $c_{00k}$  is the random effect of school  $k$ , which is assumed to be normally distributed with a mean of zero and a constant variance,  $\tau_{c00}$ , and  $d_{0(jk)}$  is the random interaction effect, which is assumed to be normally distributed with mean of 0 and a constant variance  $\tau_{d00}$ . Often the within-cell sample sizes are too small to distinguish the variance associated with the interaction effect,  $\tau_{d00}$ , from the within-cell variance,  $\sigma^2$ . Therefore, the interaction effect is often dropped from the model (Raudenbush & Bryk, 2002).

Similar to the intraclass correlation used in traditional HLM, intra-unit correlation coefficients (IUCCs) can be calculated to determine if there is substantial variability in the outcome (SHST score) between each of the cross-classified units (such as between neighborhoods, between schools, and between

cells - the interaction between schools and neighborhoods - after taking school and neighborhood into account). The formulas for IUCCs are as follows for students living in the same neighborhood and attending the same school:

$$\rho_{bcd} = \frac{\tau_{b00} + \tau_{c00} + \tau_{d00}}{\tau_{b00} + \tau_{c00} + \tau_{d00} + \sigma^2} \quad ; \quad (16)$$

for students living in same neighborhood and attending different schools:

$$\rho_b = \frac{\tau_{b00}}{\tau_{b00} + \tau_{c00} + \tau_{d00} + \sigma^2} \quad ; \quad (17)$$

and, for students attending the same schools and living in different neighborhoods:

$$\rho_c = \frac{\tau_{c00}}{\tau_{b00} + \tau_{c00} + \tau_{d00} + \sigma^2} \quad . \quad (18)$$

Note that  $\tau_{d00}$ , as mentioned previously, is often not estimated.

### **CROSS-CLASSIFIED RANDOM EFFECTS MODELING--ADDING PREDICTORS**

Just like in standard HLM, predictors can be added to any level in order to explain variation. For the sake of simplicity, two level-2 variables will be modeled including a neighborhood characteristic and a school characteristic. In addition, one student-level variable, gender, will be included in the model. The level-1 model would take on the following form:



$$Y_{i(jk)} = \beta_{0(jk)} + \beta_{1(jk)}Gender_{i(jk)} + e_{i(jk)} \quad (19)$$

where  $Y_{i(jk)}$  is the SHST score and  $Gender_{i(jk)}$  is the gender of student  $i$  in neighborhood  $j$  and school  $k$ ,  $\beta_{0(jk)}$  is the intercept, or predicted SHST score when gender = 0 for students in cell  $(jk)$ ,  $\beta_{1(jk)}$  is the regression coefficient, or the predicted change in the SHST score of a student in cell  $(jk)$  associated with a one point increase in gender (or the gender difference), and  $e_{i(jk)}$  is the within-cell random effect, or the deviation of student  $i$  in neighborhood  $j$  and school  $k$ 's SHST score from the predicted score based on gender. The random effect,  $e_{i(jk)}$  is assumed to be normally and independently distributed, with a mean of zero and a constant variance,  $\sigma^2$ .

Now, a school characteristic (*SchoolSize*) and a neighborhood variable (*SES*) will be included to help explain variability in SHST scores. Using the level-2 cross-classified factors' predictors to explain variability in the intercept ( $\beta_{0(jk)}$ ):

$$\begin{cases} \beta_{0(jk)} = \gamma_{00} + \gamma_{01k}SES_j + \gamma_{02j}SchoolSize_k + b_{0j} + c_{0k} \\ \beta_{1(jk)} = \gamma_{10} + b_{1j} + c_{1k} \end{cases} \quad (20)$$

In Equation 20, the neighborhood *SES* effect on the intercept ( $\gamma_{01k}$ ) could be modeled to vary across schools, such that:

$$\gamma_{01k} = \gamma_{01} + b_{01k} \quad (21)$$

and the corresponding *SchoolSize* effect ( $\gamma_{02j}$ ) could be modeled to vary across neighborhoods:

$$\gamma_{02j} = \gamma_{02} + c_{02j} \quad (22)$$

School-level descriptors can then be added (though not done here) to Equation 21 to explain variability between neighborhoods in  $\gamma_{01k}$  and neighborhood-level descriptors can be added to Equation 22 to explain variability between schools in  $\gamma_{02j}$ . These models can become particularly complicated. For the formulation of models that allow the effects of one cross-classified factor to vary randomly across the other cross-classified factor and of models including interaction terms, the interested reader is referred to the chapter on cross-classified random effects in Raudenbush and Bryk (2002).

In our example, combining Equations 21 and 22 with Equation 20 results in the following:

$$\begin{cases} \beta_{0(jk)} = \gamma_{00} + (\gamma_{01} + b_{01k})SES_j + (\gamma_{02} + c_{02j})SchoolSize_k + b_{0j} + c_{0k} \\ \beta_{1(jk)} = \gamma_{10} + b_{1j} + c_{1k} \end{cases} \quad (23)$$

Then, substituting and creating a single regression equation becomes:

$$Y_{i(jk)} = \gamma_{00} + (\gamma_{01} + c_{01k})SES_j + (\gamma_{02} + b_{02j})SchoolSize_k + b_{0j} + c_{0k} + (\gamma_{10} + b_{1j} + c_{1k})Gender_{i(jk)} + e_{i(jk)} \quad (24)$$

In the above cross-classified examples, the crossing occurred at the second level of the two-level model. It should be noted that cross-classifications can occur at *any* level in the data hierarchy.

### **STATEMENT OF THE PROBLEM**

CCREM is still used infrequently in educational research despite the fact that educational contexts often involve cross-classified factors. These methods are relatively new and can be quite complicated. In order to conduct CCREM, standard multilevel software must actually be “tricked” by dummy coding one of the cross-classified factors via the addition of a pseudo-level. For example, with students nested within middle schools and high schools, either high school or middle school must be specified as a third “dummy” level (Hox, 2002). Additionally, each of the high school’s (or middle school’s) variances must be constrained to be equal and the covariances must be set to zero in order for the model to be identified and the parameters estimated. Estimation of CCREM parameters is quite complex and, currently, of the commonly used multilevel modeling programs, only MLwiN (Rasbash et. al, 2000) and aML (Lillard & Panis, 2000) perform the calculations, although future versions of HLM (Raudenbush, Bryk, Cheong,& Congdon, 2000) will include this capability. SAS PROC MIXED (SAS Institute, 2002) is also able to perform CCREM (Yang, 2003).

Most methodological CCREM studies to date discuss the derivation of the estimation algorithms used (Clayton & Rasbash, 1999; Browne, Goldstein, & Rasbash, 2001), and few studies have investigated conditions that might impact parameter estimation. Researchers have compared results using real data when cross-classification was and was not ignored (Fielding, 2002; Raudenbush & Bryk, 2002). They found that estimation of fixed effects seemed relatively unaffected by the use of CCREM. However, the standard error estimates appeared to be underestimated when the cross-classified structure was ignored, impacting the statistical significance inferences associated with the fixed-effect parameter estimates. The extent of this potential problem has not been further elucidated. For instance, it would seem important to evaluate what proportion of variance attributable to each cross-classified factor necessitates the use of CCREM.

Sample size provides another concern. In Goldstein and Sammon's (1997) study, they cautioned that their sample size (758 students in 48 junior and 116 secondary schools) might be somewhat small. Garner and Raudenbush (1991) mentioned that with only an average sample size of six per level-2 unit, they were wary of the precision of random effect estimation when using CCREM. In the CCREM studies surveyed for this study, average cell sizes ranged from 4.4 to 135.88 with overall sample sizes ranging from 758 to 2,310 and number of

levels of cross-classified factors (e.g. neighborhoods and schools) ranging from 17 to 524. In addition, many cells in the cross-classifications were empty. The degree of missingness, represented by the number of empty cells, might impact the need to use CCREM. Further research should help identify the sample size necessary for accurate parameter estimation in CCREM.

Use of CCREM is now mentioned in most HLM textbooks (for example, Raudenbush & Bryk, 2002; Hox, 2002; Snijders & Boskers, 1999) although it is still infrequently used. The technical sophistication of the model seems to discourage its common use. However, applied researchers need to understand better when CCREM data structures should not be ignored and what the resulting impact could be if they are.

Because little research has been done on when it is necessary to use CCREM, several studies seem necessary. Application of the procedure to a large-scale national data set will provide insight into the potential effects of ignoring the cross-classified data structure. In addition, under the controlled environment of a simulation study, potential factors affecting the need to use CCREM will be investigated.

## Chapter 3: Method

This investigation was comprised of two studies. The first study was a Monte Carlo simulation study designed to investigate the impact of correctly modeling versus neglecting to model cross-classified data under a variety of conditions. As a follow-up to the simulation study, cross-classified random effects modeling (CCREM) was applied to a large-scale national data set. This study examined the effects of ignoring the data's cross-classified structure. The National Educational Longitudinal Study (NELS:88) was used to provide the real data analyzed in Study 2.

### STUDY 1

#### Monte Carlo Simulation

##### *Conditions*

Table 3 details the combinations of conditions that were manipulated in this simulation study. The fully-crossed study involved thirty-two combinations. The conditions that were manipulated include the correlation of the residuals at level 2 (0 and .40), the number of middle schools feeding into each high school (2 and 3), the number of levels of each cross-classified factor (here, considered to be the number of middle schools and high schools) (30 and 50), the average middle

school size (20 and 40), and the proportion of the total variance between middle schools and high schools, also termed the intra-unit correlation coefficient (5% and 15%).

#### *Correlation of the Residuals at Level 2*

With real data, it is often (but not always) the case that the factors being crossed are somewhat related. For example, it is more likely that students from low-income middle schools attend low-income high schools. In other words, there is usually some pattern as to which levels of the cross-classified factors can be combined. To mimic reality, two conditions were used: one in which the middle school and high school conditional residuals were correlated moderately ( $\rho = .40$ ) and the other in which there was no correlation between the residuals.

#### *Number of Feeder Schools*

As an indicator of the number of empty cells in the data set, the number of middle schools feeding into each high school was manipulated. Using the NELS:88 data set as a guideline, it was found that typically either two or three middle schools fed into a given high school. The fewer the number of middle school feeders, the more empty cells exist in the data set. So, for example, suppose middle school one feeds only into high schools one and two. In this case, empty cells exist for the combination of middle school one and every other high

school in the data set. Having two feeder schools results in more empty cells than does having three.

#### *Number of Levels of Cross-Classified Factors*

In order to manipulate the overall sample size, the number of middle schools and high schools was varied. The number of middle schools and high schools by which students could be cross-classified was either 30 or 50. The number of middle schools was always equal to the number of high schools.

#### *Average Middle School Size*

The average number of students at each middle school was also manipulated. The overall sample size for a study was a function of not only the number of middle schools and high schools, but also the number of students per middle school. Since it is unusual for the middle schools (and therefore the cells) to contain equal sample sizes, the number of students in each middle school was randomly generated. The middle school sample sizes were drawn from a normal distribution with a known mean (with values of 20 or 40 based on the relevant condition with values rounded to the closest integer, see Table 3) and standard deviation (2). Thus, in the smaller sample size condition, instead of each middle school containing exactly twenty students, the number of students in each middle school randomly varied around twenty, yielding an average per-middle school sample size of twenty.



### *Intra Unit Correlation Coefficients (IUCCs)*

A search of the studies involving the use of CCREM, as well as textbook examples in multilevel modeling textbooks, indicated that actual observed estimates for intra-unit correlation coefficients (IUCCs) range from .009% to 24%, with a mean of 6.6% and standard deviation of .0682. Therefore, in order to mimic real situations, values of 5% and 15% were used to represent small and moderate IUCCs for each cross-classified factor. Conditional IUCCs were generated, such that after including the explanatory variables in the model, the variance between students, middle schools, and high schools had known values.

*Table 3. Simulation Design Conditions*

<b>Condition</b>	<b>Correlation of Residuals</b>	<b>Number of Feeder Schools</b>	<b>Levels of Cross-Classified Factors</b>	<b>Average Middle School Size</b>	<b>IUCC per Cross-Classified Factor</b>
1	.00	2	30	20	.05
2	.00	2	30	20	.15
3	.00	2	30	40	.05
4	.00	2	30	40	.15
5	.00	2	50	20	.05
6	.00	2	50	20	.15
7	.00	2	50	40	.05
8	.00	2	50	40	.15
9	.00	3	30	20	.05
10	.00	3	30	20	.15
11	.00	3	30	40	.05
12	.00	3	30	40	.15
13	.00	3	50	20	.05
14	.00	3	50	20	.15
15	.00	3	50	40	.05
16	.00	3	50	40	.15
17	.40	2	30	20	.05
18	.40	2	30	20	.15
19	.40	2	30	40	.05
20	.40	2	30	40	.15
21	.40	2	50	20	.05
22	.40	2	50	20	.15
23	.40	2	50	40	.05
24	.40	2	50	40	.15
25	.40	3	30	20	.05
26	.40	3	30	20	.15
27	.40	3	30	40	.05
28	.40	3	30	40	.15
29	.40	3	50	20	.05
30	.40	3	50	20	.15
31	.40	3	50	40	.05
32	.40	3	50	40	.15

### ***Data Generation***

SAS (SAS Institute, 2002) was used to generate and analyze the data for each of the 1,000 replications. Data were generated to fit a two-level cross-classified multilevel model, with students at level 1 and middle school and high school crossed at level 2. A total of three predictors, one student-level predictor, and one predictor for each cross-classified factor, were included in the generating model. Specifically, the student-level variable,  $X$ , the middle school predictor,  $Z$ , and the high school predictor,  $W$ , were generated from normal distributions with a mean of 50 and a standard deviation of 10. Each of the three predictors was related to the outcome and the predictors for middle school and high school were modeled as being related to the outcome with the relationship fixed across the other cross-classified factor. Using the notation of Raudenbush and Bryk (2002) and substituting the generating parameter values results in the following generating level-1 equation:

$$Y_{i(jk)} = \beta_{00(jk)} + 0.5X_{i(jk)} + e_{i(jk)} \quad (25)$$

and the level-2 equation is:

$$\beta_{00(jk)} = 100 + 0.5Z_j + 0.5W_k + b_{0j} + c_{0k} \quad (26)$$

where  $j$  indexes middle school and  $k$  indexes high school. Thus, the generating values to be recovered for the fixed effects were: 0.5 for  $X$ , 0.5 for  $W$ , 0.5 for  $Z$

and 100 for the intercept. Equations 25 and 26 can be combined into a single equation such that:

$$Y_{i(jk)} = 100 + 0.5Z_j + 0.5W_k + 0.5X_{i(jk)} + e_{ij} + b_{0j} + c_{0k} \quad (27)$$

The generating values of the variances of each of  $b_{0j}$  and  $c_{0k}$  were either .0556 (representing a conditional IUCC of .05) or .2143 (representing a conditional IUCC of .15) depending on the condition (see Table 3). The values of the variances of  $b_{0j}$  and  $c_{0k}$  were arrived at in the following manner: To yield an IUCC of 0.05 for both middle schools and high schools, the IUCC equations in Equations 16 through 18 were solved where  $\tau_{b0j} = \tau_{c0k}$ ,  $\sigma^2 = 1$ , and  $\rho_b = \rho_c = .05$ . For the conditions in which the generating value for the IUCC is 0.15, the same equations were solved where  $\rho_b = \rho_c = .15$ . The student-level variance was fixed at 1.0.

*Generating the combinations of middle school and high school*

Given a set of middle schools and high schools, it is unlikely that there are students in every possible combination of middle school and high school. Therefore, the data were generated so that only certain cells existed. Specifically, based on the condition in the study, a matrix of middle school and high school residuals was generated so that they were correlated on average either .40 or .00. After this matrix was created, it was sorted by the middle school residual, in ascending order. Then, the cell combinations that were generated to exist were

varied according to the condition in Table 3. When there were two middle school feeders, the middle school sent 60% of its students to the “adjacent” high school (the high school with the same rank associated with the middle school’s residual) and 40% of its students to the high school paired with the middle school with the next highest residual. When there were three middle school feeders, the middle school sent 70% of its students to the adjacent high school, 15% of its students to the high school associated with the middle school with the next lowest residual and 15% of its students to the high school paired with the middle school with the next highest residual. Thus, for example, after the schools were sorted by middle school residual, the second ranked middle school sent (depending on the condition) 60% of its students to its paired high school and 40% to the next highest ranked school for the two middle school feeder condition. In the three middle school condition, the second ranked middle school sent 70% of its students to the next highest ranked school and 15% of its students to each of the first and third ranked high schools. The feeding pattern wrapped around for the first and last schools in the data set. So, for example, in the three feeder school condition with 30 middle schools and high schools, middle school 30 was paired with high schools 29, 30, and one. Similarly, in this condition, high school one received students from middle school one, two, and 30. These “feeding patterns” were based on the NELS:88 data set. The resulting data structure contained

students cross-classified by middle and high school. Note that the above process was also conducted in the conditions where the correlation between residuals was .00, except that no ranking was done. The result was that middle schools were randomly paired with high schools. Figures 1 and 2 on the following pages illustrate these feeding patterns.

Figure 1. Two Feeder School Condition

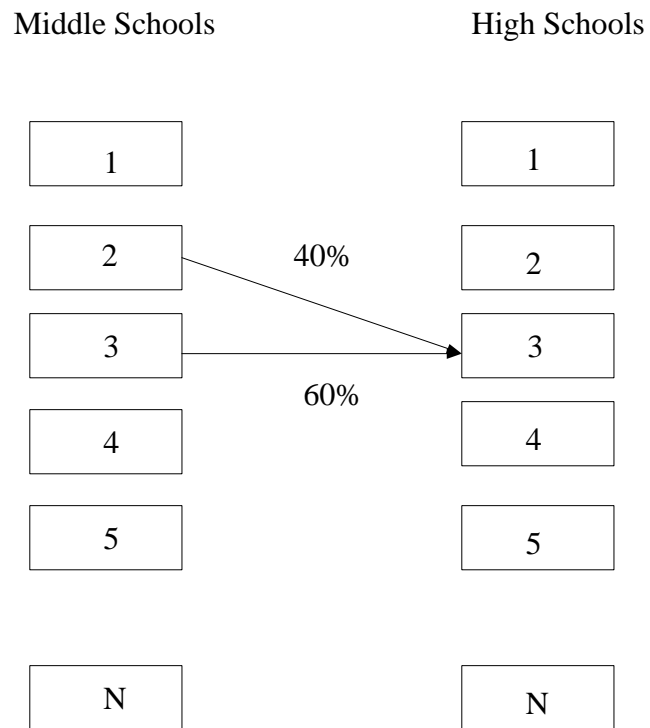
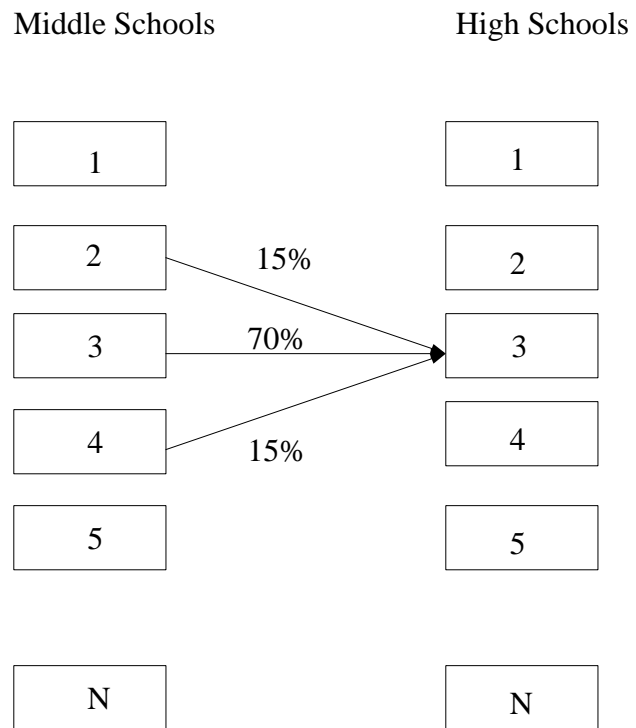


Figure 2. Three Feeder School Condition





## *Analysis*

SAS PROC MIXED (SAS Institute, 2002) was used to fit the appropriate models and summarize the data. For each of the thirty-two conditions, two models were estimated. *Model 1* was a conditional CCREM with one level-1 explanatory variable and an explanatory variable for each cross-classified factor, matching the way the data were generated. *Model 1* was fit assuming that the data structure was as follows. At level 1:

$$Y_{i(jk)} = \beta_{00(jk)} + \beta_{01(jk)}X_{i(jk)} + e_{i(jk)} \quad (28)$$

and at level 2:

$$\begin{cases} \beta_{00(jk)} = \gamma_{00} + \gamma_{01}Z_j + \gamma_{02}W_k + b_{0j} + c_{0k} \\ \beta_{01(jk)} = \gamma_{10} \end{cases} \quad (29)$$

where  $i$  indexes students,  $j$  indexes middle schools, and  $k$  indexes high schools.

Combining Equations 28 and 29 into a single regression equation results in the following:

$$Y_{i(jk)} = \gamma_{00} + \gamma_{10}X_{i(jk)} + \gamma_{01}Z_j + \gamma_{02}W_k + b_{0j} + c_{0k} + e_{i(jk)} \quad (30)$$

The variance of the random interaction effect was not estimated and the effect of the level-1 predictor was assumed fixed as were those of the two level-2 predictors,  $Z_j$  and  $W_k$ .

*Model 2* was a conditional HLM with high school as the level-2 clustering unit, thereby ignoring the cross-classified data structure and including the middle school variable as a level-1 variable. Thus, at level 1 the following data structure was modeled:

$$Y_{ik} = \beta_{0k} + \beta_1 X_{ik} + \beta_{2k} Z_{ik} + r_{ik} \quad (31)$$

and at level 2:

$$\begin{cases} \beta_{0k} = \gamma_{00} + \gamma_{01} W_k + u_{0k} \\ \beta_{1k} = \gamma_{10} \\ \beta_{2k} = \gamma_{20} \end{cases} \quad (32)$$

Combining the two equations into one, *Model 2* is represented by:

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ik} + \gamma_{20} Z_{ik} + \gamma_{01} W_k + r_{ik} + u_{0k} \quad (33)$$

where  $i$  indexes students and  $k$  indexes high schools.

Both fixed and random effect parameter estimates were summarized and compared to known generating values. The values of both the fixed parameter estimates and random effect estimates were summarized across the 1,000 iterations for each combination of conditions. In addition, the estimates of the standard errors associated with each coefficient were compared across models for each condition.

For each of the fixed and random effect estimates, the average value for each estimate across iterations was calculated as was the relative bias of the

parameter estimates (Hoogland & Boomsma, 1998). Relative bias of the parameter estimates is calculated by using the equation

$$B(\hat{\theta}) = \frac{\bar{\hat{\theta}}_i - \theta}{\theta}, i = 1, 2, \dots, t, \quad (34)$$

where  $\theta$  is the value of the  $i$ th parameter,  $\bar{\hat{\theta}}_i$  is the mean of the  $i$ th parameter estimates across the 1,000 replications, and  $t$  is the number of parameters to be estimated.

In order to summarize the standard error estimates, relative bias was calculated for those estimates as well. This required the calculation of two types of standard error estimates. First, for each coefficient estimate, the standard deviation of the estimates across the 1,000 iterations was calculated yielding the empirical standard error,  $\hat{se}_\theta$ . In addition, the mean of the standard error estimates associated with each coefficient was calculated, yielding the mean standard error estimate,  $\bar{\hat{se}}_\theta$ . The relative bias of the standard error (Hoogland & Boomsma, 1998) is defined as:

$$B(\hat{se}_\theta) = \frac{\bar{\hat{se}}_\theta - \hat{se}_\theta}{\hat{se}_\theta}, i = 1, 2, \dots, t, \quad (35)$$

Finally, Akaike's Information Criterion (AIC), and Schwarz's Bayesian Criterion (SBC) were used to identify the best fitting model. These fit indices are used when the models being compared are not nested, as is the case with this

study. Smaller values for these fit indices indicate better fit. Because the data were generated to fit a CCREM, a tally was kept describing the proportion of the 1,000 iterations that the CCREM was chosen over the HLM to represent the proportion of correct model identifications.

Analyses of variance (ANOVAs) were conducted when either the relative bias or standard error bias fell outside of the acceptable range for multiple conditions. Hoogland and Boomsma (1998) recommend a cutoff for acceptability of .05 for the relative bias of the parameter estimates and .10 for the relative bias of the standard error estimates. To determine which of the study's factors contributed to the observed bias, ANOVAs were conducted on the aggregate values (across the 1,000 iterations) with the bias (either relative bias or standard error bias) as the dependent variable and each manipulated variable (number of feeder schools, etc.) as a factor. Only main effects were investigated. The .05 alpha-level was used to determine statistical significance. In addition, for factors that resulted in statistical significance, eta squared effect sizes were computed as a measure of practical significance. Eta squared is calculated by dividing the sum of squares for the effect by the total sum of squares. Conventionally, values of .01, .06, and .14 represent small, moderate, and large effects, respectively (Green, Salkind, & Akey, 2000).

ANOVAs were also conducted to determine which factors affected the proportion of correct model identifications. Here, the percentage of correct model identifications was the dependent variable and each manipulated variable was a factor. Eta squared was calculated for any factor found to be statistically significant.

## **STUDY 2**

### **Data**

The National Educational Longitudinal Study (NELS:88) provides an ideal data set for an investigation into the potential effects of ignoring cross-classified random effects. The study began in 1988 when the 25,000 student participants were in eighth grade and then followed the students for three follow-ups at two-year intervals. Because most American students transition from junior high (or middle school) to high school between eighth and ninth grade, there is an inherent cross-classification in the data. All students who attended a given middle school in 1988 did not subsequently attend the same high school. Therefore, students are nested within a cross-classification of middle and high schools.

### **Variables**

As an investigation into the functioning of CCREM, a simple example of cross-classified random effects modeling was conducted. The dependent measure was modeled as being nested within the cross-classification of middle school and

high school (both at level 2). This example, using the NELS:88 data set, mimics the cross-classified example provided in the MLwiN user's guide (Goldstein, Rasbash, et al. 1998).

### ***Dependent Variable***

The dependent variable was a standardized Reading/Math composite test score for the second wave of data, when the students were in tenth grade. This test score was modeled as a function of individual student-level characteristics, as well as characteristics of the middle school the student attended and the high school he or she currently attends.

### ***Predictor Variables***

For the sake of simplicity, one student-level variable was used as well as one school-level variable to describe each of the types of school (middle school versus high school). The student-level variable used was gender. Although not purposely based upon any specific theory, it is not far-fetched to think that there might be gender differences in student achievement. At the high school-level, the percentage of the students in the free or reduced lunch program was used. Again, although not intentionally based on a specific theory, the percentage of these students at a given school may provide insight into the quality of educational resources available. It is reasonable to postulate that the percentage of

students in this program is related to average student achievement. Finally, as it is reasonable to think that the population of a school might impact average test performance, the total school enrollment was used as a predictor at the middle school-level.

### **Analysis**

SPSS (SPSS, Inc., 2002) was used to import the NELS:88 data set for analysis. Since hierarchical linear modeling (and cross-classified random effects modeling) cannot be conducted when there is missing data for the level-2 identification variable, any observation that did not have both a middle school and a high school identification code (29.3% of the observations) was deleted from the data set.

SAS (SAS Institute, 2002) was then used to analyze the data. It should be noted that the NELS:88 data employs a complex weighting design. The weighting system was ignored for the purposes of this analysis. Two hierarchical linear models were fit that ignored the data's cross-classification. The models are differentiated by which cases were included in the analysis. The first of the models, referred to as *Model A*, involved deleting cases. For a given high school, the largest middle school feeder into the school was used. Only students at said high school who had attended the primary feeder middle school were included in the analysis. Any students at a high school that came from a middle school other

than the primary feeder school were deleted from the data set. This process resulted in dropping 5.1% of the cases. By deleting students not from the primary feeder schools, the cross-classification was removed from the data set. Test scores were nested within the middle school-high school “sequence.” It should be noted that another way to remove the cross-classification is to, given a middle school, remove those students who did not subsequently attend the “main” high school. However, with the NELS:88 data set, the sample sizes are too small to obtain stable results after deleting cases in this manner. The high school (level-2) explanatory variable was the percentage of students at the high school in the free or reduced lunch program and the middle school variable (also at level 2) was the total enrollment of the middle school. The HLM equations (using the notation of Raudenbush & Bryk (2002)) for *Model A* were as follows:

For level 1:

$$Y_{ik} = \beta_{0k} + \beta_{1k}(Gender)_{ik} + r_{ik} \quad (36)$$

and for level 2:

$$\begin{cases} \beta_{0k} = \gamma_{00} + \gamma_{01}(per\_free)_k + \gamma_{02}(school\_size)_k + u_{0k} \\ \beta_{1k} = \gamma_{10} \end{cases} \quad (37)$$

Note that the effect of gender was assumed to be constant across schools. It should also be noted that  $k$  is used to index high school in the HLM models to match its use an index in the corresponding CCREM.



Combining these two levels becomes:

$$Y_{ik} = \gamma_{00} + \gamma_{10}(Gender)_{ik} + \gamma_{01}(per\_free)_k + \gamma_{02}(school\_size)_k + u_{0k} + r_{ik} \quad (38)$$

*Model B* also involved ignoring the cross-classified nature of the data.

However, this model included all students in the data set and involved disaggregating the data. Observations were modeled as being nested within *high school*. The high school descriptor variable, percent of students in the lunch program, was included at level 2, the high school-level. However, the middle school descriptor variable, *school size*, was treated as a student-level (level-1) variable. Specifically, each student in a given middle school was given the same value for this variable. The HLM equations for *Model B* were as follows:

At level 1:

$$Y_{ik} = \beta_{0k} + \beta_{1k}(Gender)_{ik} + \beta_{2k}(school\_size)_{ik} + r_{ik} \quad (39)$$

And at level 2:

$$\begin{cases} \beta_{0k} = \gamma_{00} + \gamma_{01}(per\_free)_k + u_{0k} \\ \beta_{1k} = \gamma_{10} \\ \beta_{2k} = \gamma_{20} \end{cases} \quad (40)$$

Here, the effects of *gender* and *school size* were assumed fixed across high schools. Combining these two levels, *Model B* becomes:

$$Y_{ik} = \gamma_{00} + \gamma_{10}(Gender)_{ik} + \gamma_{01}(per\_free)_k + \gamma_{20}(school\_size)_{ik} + r_{ik} + u_{0k} \quad (41)$$

Finally, *Model C* involved modeling the cross-classification appropriately. Test scores were nested within the cross-classification of middle school and high school. Each middle school had a value for *school size* and each high school had a value for the percentage of students in the free or reduced lunch program. This model can be represented by the following equations. For level 1:

$$Y_{i(jk)} = \beta_{0(jk)} + \beta_{1(jk)}Gender_{i(jk)} + e_{i(jk)} \quad (42)$$

And, at level 2:

$$\begin{cases} \beta_{0(jk)} = \gamma_{00} + \gamma_{01}(school\_size)_j + \gamma_{02}(per\_free)_k + b_{0j} + c_{0k} \\ \beta_{1(jk)} = \gamma_{10} \end{cases} \quad (43)$$

Again, the effect of gender was assumed to be constant across both middle schools and high schools. Also, the effect of *school size* was modeled as being constant across high schools, and the effect of *percent free lunch* was modeled as being constant across middle schools. Combining Equations 42 and 43, the single equation describing *Model C* becomes:

$$Y_{i(jk)} = \gamma_{00} + \gamma_{10}(gender)_{i(jk)} + \gamma_{01}(school\_size)_j + \gamma_{02}(per\_free)_k + b_{0j} + c_{0k} + e_{i(jk)} \quad (44)$$

The variance due to the random interaction effect,  $\tau_{d00}$ , was not estimated since it is often hard to estimate because it is difficult to distinguish it from the within-cell variance,  $\sigma^2$  (Raudenbush & Bryk, 2002).

The results of *Models A* and *B* were compared with *Model C* to assess differences in the results. Specifically, the estimates and statistical significance of each fixed effect were compared across models. In addition, the estimates of the variance components between students and between high schools were compared.

Finally, the Akaike's Information Criterion (AIC) and Schwarz's Bayesian Criterion (SBC) fit statistics were compared across models to identify the best fitting model. The model associated with the lowest values of the AIC and SBC was identified as the model that the data best fit.

When comparing parameter estimates using real data, the researcher does not know "truth." Therefore, the above study was conducted to demonstrate use of CCREM as well as discover potential differences in estimates based on the model assumed.

## Chapter 4: Results

### STUDY 1

#### Fixed Effect Estimates

##### *X, The Student-Level Predictor*

The statistics for describing the estimation of the coefficient ( $\gamma_{10}$ ) associated with the level-1 predictor variable,  $X$ , are provided in Table 4 and the bias statistics follow in Table 5. The conditions are identified by number. Refer to Table 3 for the detailed description of the conditions.

##### Relative Bias

Hoogland and Boomsma's (1998) criteria of acceptability for relative bias is that the observed bias magnitude be less than .05. Using that criterion, all estimates of the fixed effect of  $X$  ( $\gamma_{10}$ ) were acceptable in terms of the relative bias. Neither the correlation of the residuals at level 2, the number of feeder schools, the number of levels of each cross-classified factor, the average middle school size, or the IUCCs had any noticeable impact on the relative bias of  $\gamma_{10}$ , the effect of  $X$ . Both the CCREM and HLM performed well.

##### Standard Error Bias

Hoogland and Boomsma (1998) recommend a cutoff for acceptability of .10 for the magnitude of the bias of the standard errors. Using that criterion, all standard errors met the acceptable standard. None of the factors being manipulated seemed to have any impact on the bias of the standard error of the

estimate of  $\gamma_{10}$ , the coefficient of  $X$ . Again, both CREM and HLM performed well.

Table 4. Descriptive Statistics for the fixed effect coefficient,  $\gamma_{10}$ , associated with the student-level variable

Condition	Estimating Model			
	CCREM		HLM	
	Mean	SD	Mean	SD
1	0.50002	0.0043	0.50003	0.0043
2	0.49995	0.0044	0.49993	0.0046
3	0.50004	0.0030	0.50007	0.0030
4	0.50007	0.0029	0.50009	0.0029
5	0.49986	0.0033	0.49985	0.0034
6	0.50012	0.0034	0.50012	0.0035
7	0.49995	0.0023	0.49996	0.0023
8	0.49997	0.0024	0.49999	0.0024
9	0.50002	0.0043	0.50003	0.0075
10	0.49989	0.0044	0.49986	0.0046
11	0.50007	0.0030	0.50006	0.0031
12	0.50005	0.0031	0.50010	0.0033
13	0.50010	0.0033	0.50011	0.0033
14	0.50009	0.0034	0.50007	0.0035
15	0.50009	0.0024	0.50009	0.0024
16	0.50005	0.0023	0.50005	0.0024
17	0.50002	0.0044	0.50001	0.0044
18	0.49995	0.0043	0.49990	0.0043
19	0.50010	0.0031	0.50011	0.0030
20	0.49999	0.0029	0.50001	0.0029
21	0.50007	0.0033	0.50006	0.0033
22	0.49994	0.0033	0.49994	0.0033
23	0.50005	0.0024	0.50005	0.0024
24	0.49999	0.0022	0.50000	0.0022
25	0.50000	0.0043	0.50000	0.0044
26	0.49999	0.0043	0.50000	0.0044
27	0.50018	0.0030	0.50018	0.0029
28	0.50004	0.0031	0.50003	0.0031
29	0.50011	0.0034	0.50012	0.0034
30	0.50016	0.0035	0.50016	0.0035
31	0.49993	0.0023	0.49993	0.0023
32	0.49993	0.0024	0.49994	0.0024

Note. The value to be recovered is .50 for the fixed effect.

Table 5. Bias Statistics for the fixed effect coefficient,  $\gamma_{10}$ , associated with the student-level variable.

Condition	Estimating Model			
	CCREM		HLM	
	Param. Est. Relative Bias	SE Relative Bias	Param. Est. Relative Bias	SE Relative Bias
1	0.00004	0.0079	0.00006	0.0109
2	-0.00010	-0.0091	-0.00015	-0.0110
3	0.00008	0.0020	0.00013	0.0041
4	0.00042	0.0431	0.00019	0.0552
5	-0.00027	0.0031	-0.00030	0.0012
6	0.00024	0.0044	0.00024	0.0084
7	-0.00010	0.0073	-0.00008	0.0064
8	-0.00007	-0.0146	-0.00002	-0.0185
9	0.00004	0.0163	0.00005	0.0166
10	-0.00022	0.0054	-0.00029	0.0060
11	0.00014	-0.0144	0.00012	-0.0169
12	0.00010	-0.0383	0.00020	-0.0450
13	0.00020	0.0261	0.00023	0.0237
14	0.00018	0.0239	0.00015	0.0079
15	0.00018	-0.0156	0.00018	-0.0248
16	0.00011	0.0229	0.00009	0.0252
17	0.00004	-0.0149	0.00002	-0.0113
18	-0.00010	0.0283	-0.00021	0.0334
19	0.00020	-0.0266	0.00023	-0.0217
20	-0.00002	0.0390	0.00001	0.0450
21	0.00014	-0.0009	0.00012	0.0020
22	-0.00012	0.0206	-0.00013	0.0220
23	0.00011	-0.0215	0.00009	-0.0192
24	-0.00003	0.0427	-0.00001	0.0477
25	-0.00001	0.0122	0.00000	0.0067
26	-0.00002	0.0153	0.00000	0.0224
27	0.00036	0.0163	0.00036	0.0190
28	0.00009	-0.0140	0.00006	-0.0192
29	0.00023	0.0103	0.00023	0.0113
30	0.00032	-0.0248	0.00031	-0.0235
31	-0.00013	0.0170	-0.00013	0.0175
32	-0.00013	-0.0104	-0.00012	-0.0087

Note. Values below .05 are considered acceptable for parameter bias and values below .10 are considered acceptable for the standard error bias.

### ***Z, The Middle School-Level Predictor***

The statistics describing the estimation of the fixed effect coefficient ( $\gamma_{01}$  for *Model 1* or  $\gamma_{20}$  for *Model 2*) associated with the predictor variable, *Z*, for one of the cross-classified factors (here, considered middle school) are provided in Table 6 and the bias statistics follow in Table 7. This effect coefficient will be referred to as  $\gamma_z$  here.

#### *Relative Bias*

All thirty-two conditions met the Hoogland and Boomsma (1998) criteria for acceptability of the relative bias estimates. None of the factors had a noticeable impact on the relative bias of the effect of *Z* ( $\gamma_z$ ), and both CCREM and HLM performed well.

#### *Standard Error Bias*

Unlike the relative bias, the standard error bias of the effect of *Z* ( $\gamma_z$ ) was unacceptable for all conditions when assuming the HLM model. Assuming HLM and ignoring the variance at the middle school level appears to substantially underestimate the standard errors of the fixed effect for the middle school-level predictor.

The correlation of the residuals at level 2 had a moderate effect on the negative standard error bias,  $F(1,26) = 209.688$ ,  $p < .001$ ,  $\eta^2 = .0777$ . In conditions where the correlation was 0, the average standard error bias was -.400



for the HLM compared to only -.211 when the correlation was .40. Thus, it appears that when the residuals are uncorrelated the standard errors for the estimates of the fixed effect of the middle school predictor variable are more negatively biased than when the residuals are moderately correlated.

The IUCC also had a moderate effect on the observed negative standard error bias,  $F(1,26) = 189.504$ ,  $p < .001$ ,  $\eta^2 = .0701$ . Assuming HLM and ignoring the variance at the middle school level, the average standard error bias when the IUCC was .05 was only -.216 compared to -.395 when the IUCC was .15. The more variance existing between middle schools and high schools, the larger the magnitude of the standard error bias.

The average number of students per middle school had a small effect on the observed negative standard error bias,  $F(1,26) = 62.258$ ,  $p < .001$ ,  $\eta^2 = .0231$ . Assuming HLM, the average standard error bias when there were twenty students per middle school was -.264 compared to -.357 when forty were used. Increasing the number of students at the middle school level was associated with a slight increase in the magnitude of the negative standard error bias, although the difference was small.

Despite its statistical significance, the effect of the number of schools that were crossed on the observed standard error bias was not very large, as evidenced by its small eta squared value,  $F(1,26) = 6.834$ ,  $p < .05$ ,  $\eta^2 = .0025$ . Assuming

HLM, the average standard error bias when thirty schools was used was  $-.32$  compared to  $-.29$  when fifty schools were used. The difference between these means is negligible.

Table 6. Descriptive Statistics for the fixed effect coefficient,  $\gamma$ , associated with the middle school- level variable.

Condition	Estimating Model			
	CCREM		HLM	
	Mean	SD	Mean	SD
1	0.50028	0.0073	0.50026	0.0073
2	0.49979	0.0109	0.49967	0.0115
3	0.49997	0.0060	0.49998	0.0061
4	0.50029	0.0101	0.50034	0.0112
5	0.50024	0.0052	0.50027	0.0053
6	0.50041	0.0081	0.50025	0.0086
7	0.50009	0.0046	0.50008	0.0048
8	0.49995	0.0077	0.49987	0.0086
9	0.50026	0.0074	0.50003	0.0075
10	0.50041	0.0114	0.50033	0.0118
11	0.49994	0.0061	0.49992	0.0062
12	0.49987	0.0102	0.49986	0.0109
13	0.49990	0.0054	0.49991	0.0055
14	0.50005	0.0085	0.50007	0.0090
15	0.50000	0.0046	0.49997	0.0047
16	0.49992	0.0079	0.49994	0.0084
17	0.49984	0.0075	0.49981	0.0070
18	0.50060	0.0099	0.50033	0.0087
19	0.50023	0.0061	0.50024	0.0055
20	0.50004	0.0085	0.49998	0.0070
21	0.49958	0.0051	0.49966	0.0048
22	0.49984	0.0007	0.49981	0.0061
23	0.49998	0.0041	0.50000	0.0038
24	0.49999	0.0056	0.50011	0.0049
25	0.49981	0.0069	0.49979	0.0065
26	0.49985	0.0100	0.49967	0.0086
27	0.49999	0.0060	0.50001	0.0053
28	0.50002	0.0087	0.50013	0.0072
29	0.49996	0.0053	0.49992	0.0050
30	0.49969	0.0072	0.49976	0.0063
31	0.49983	0.0041	0.49985	0.0037
32	0.49986	0.0057	0.49977	0.0050

Note. The value to be recovered is .50 for the fixed effect.

Table 7. Bias Statistics for the fixed effect coefficient,  $\gamma_z$ , associated with the middle school- level variable.

Condition	Estimating Model			
	CCREM		HLM	
	Param. Est. Relative Bias	SE Relative Bias	Param. Est. Relative Bias	SE Relative Bias
1	0.00056	-0.0568	0.00051	-0.2456
2	-0.00043	-0.0152	-0.00066	-0.4402
3	-0.00006	-0.0014	-0.00005	-0.3263
4	0.00057	-0.0095	0.00067	-0.5838
5	0.00048	0.0103	0.00054	-0.2006
6	0.00083	0.0255	0.00049	-0.4247
7	0.00018	0.0043	0.00016	-0.3416
8	-0.00011	0.0034	-0.00027	-0.5822
9	0.00052	-0.0548	0.00042	-0.2563
10	0.00083	-0.0620	0.00065	-0.4575
11	-0.00011	-0.0278	-0.00016	-0.3482
12	-0.00026	-0.0184	-0.00028	-0.5825
13	-0.00019	-0.0349	-0.00018	-0.2337
14	0.00010	-0.0318	0.00014	-0.4564
15	0.00000	-0.0014	-0.00006	-0.3378
16	-0.00017	-0.0297	-0.00012	-0.5850
17	-0.00031	-0.0569	-0.00037	-0.1646
18	0.00119	-0.0449	0.00066	-0.2358
19	0.00046	-0.0806	0.00047	-0.2203
20	0.00009	-0.0349	-0.00005	-0.3400
21	-0.00084	0.0147	-0.00068	-0.0627
22	-0.00033	-0.0279	-0.00039	-0.1802
23	-0.00004	-0.0181	0.00000	-0.1313
24	-0.00002	-0.0479	0.00022	-0.2937
25	-0.00038	0.0022	-0.00043	-0.1105
26	-0.00031	-0.0550	-0.00065	-0.2628
27	-0.00002	-0.0771	0.00002	-0.2115
28	0.00004	-0.0542	0.00025	-0.3793
29	-0.00008	-0.0275	-0.00015	-0.1182
30	-0.00062	-0.0799	-0.00047	-0.2202
31	-0.00330	-0.0117	-0.00030	-0.1482
32	-0.00027	-0.0424	-0.00045	-0.3036

Note. Values below .05 are considered acceptable for parameter bias, and values below .10 are considered acceptable for the standard error bias.

### ***W, The High School-Level Predictor***

The descriptive statistics for the fixed effects of  $W$  ( $\gamma_{02}$  in *Model 1* and  $\gamma_{01}$  in *Model 2*) are provided in Table 8 and the bias statistics follow in Table 9. This fixed effect coefficient will be referred to as  $\gamma_w$  here. Recovering  $\gamma_w$  appeared to present no problems.

#### *Relative Bias*

The relative bias in estimation of  $\gamma_w$  met the criteria for all thirty-two conditions and both estimating models. None of the factors had a noticeable impact on the relative bias of the effect of  $W$ , and estimation assuming both CCREM and HLM performed well.

#### *Standard Error Bias*

All thirty-two conditions and both models had standard error bias values under the .10 cutoff.

Table 8. Descriptive Statistics for the fixed effect coefficient,  $\gamma_w$ , associated with the high school- level variable.

Condition	Estimating Model			
	CCREM		HLM	
	Mean	SD	Mean	SD
1	0.50001	0.0071	0.49995	0.0073
2	0.49973	0.0114	0.49971	0.0122
3	0.49993	0.0062	0.49990	0.0066
4	0.49971	0.0103	0.49970	0.0113
5	0.50006	0.0053	0.50001	0.0054
6	0.50024	0.0081	0.50016	0.0088
7	0.49994	0.0049	0.49995	0.0051
8	0.49948	0.0078	0.49949	0.0087
9	0.49975	0.0073	0.49968	0.0075
10	0.49997	0.0110	0.50004	0.0115
11	0.49983	0.0061	0.49978	0.0064
12	0.50053	0.0105	0.50051	0.0117
13	0.50010	0.0055	0.50011	0.0056
14	0.50002	0.0086	0.49991	0.0092
15	0.49994	0.0046	0.49995	0.0049
16	0.50008	0.0078	0.49983	0.0087
17	0.49980	0.0081	0.49982	0.0084
18	0.49967	0.0140	0.49953	0.0154
19	0.50011	0.0072	0.50002	0.0077
20	0.49976	0.0125	0.49977	0.0143
21	0.50011	0.0064	0.50010	0.0067
22	0.49996	0.0111	0.49997	0.0118
23	0.49994	0.0059	0.49994	0.0063
24	0.49996	0.0100	0.49991	0.0107
25	0.49991	0.0085	0.49998	0.0091
26	0.49944	0.0138	0.49944	0.0152
27	0.50005	0.0076	0.50005	0.0082
28	0.49947	0.0131	0.49952	0.0151
29	0.49979	0.0062	0.49981	0.0064
30	0.49974	0.0111	0.49980	0.0119
31	0.49995	0.0059	0.49993	0.0064
32	0.49889	0.0097	0.49881	0.0105

Note. The value to be recovered is .50 for the fixed effect.

Table 9. Bias Statistics for the fixed effect coefficient,  $\gamma_w$ , associated with the high school- level variable.

Condition	Estimating Model			
	CCREM		HLM	
	Param. Est. Relative Bias	SE Relative Bias	Param. Est. Relative Bias	SE Relative Bias
1	0.00003	-0.0221	-0.00009	-0.0160
2	-0.00055	-0.0544	-0.00058	-0.0282
3	-0.00015	-0.0297	-0.00021	-0.0387
4	-0.00058	-0.0257	-0.00060	0.0040
5	0.00013	-0.0127	0.00001	0.0027
6	0.00024	0.0205	0.00024	0.0230
7	-0.00011	-0.0556	-0.00011	-0.0326
8	-0.00103	-0.0162	-0.00102	-0.0009
9	-0.00049	-0.0547	-0.00064	-0.0469
10	-0.00005	-0.0259	0.00007	0.0099
11	-0.00034	-0.0293	-0.00044	-0.0043
12	0.00107	-0.0393	0.00101	-0.0231
13	0.00020	-0.0369	0.00021	-0.0157
14	0.00003	-0.0333	-0.00018	-0.0200
15	-0.00012	-0.0098	-0.00010	-0.0283
16	0.00015	-0.0220	-0.00033	-0.0125
17	-0.00040	-0.0220	-0.00036	0.0136
18	-0.00066	-0.0594	-0.00094	-0.0473
19	0.00022	-0.0109	0.00004	0.0101
20	-0.00048	-0.0007	-0.00047	0.0071
21	0.00021	-0.0386	0.00020	-0.0288
22	-0.00009	-0.0334	-0.00006	-0.0282
23	-0.00012	-0.0443	-0.00013	-0.0349
24	-0.00008	0.0259	-0.00018	0.0293
25	-0.00018	-0.0732	-0.00004	-0.0680
26	-0.00112	-0.0528	-0.00112	-0.0329
27	0.00010	-0.0716	0.00010	-0.0556
28	-0.00107	-0.0692	-0.00095	-0.0524
29	-0.00043	0.0030	-0.00038	0.0245
30	-0.00052	-0.0437	-0.00041	-0.0350
31	-0.00010	-0.0556	-0.00013	-0.0575
32	-0.00222	0.0337	-0.00238	0.0504

Note. Values below .05 are considered acceptable for parameter bias, and values below .10 are considered acceptable for the standard error bias.

## **Random Effect Estimates**

Bias was only reported for conditions in which the correlation between middle school and high school residual was zero. In the conditions in which the level-two residuals were correlated .40, the data were generated such that only certain combinations of middle schools and high schools could be paired, and thus the true value of the variance components, or IUCCs, was not known.

Recall that the generating values for the variance between students was 1.0 and the variance between both middle schools and high schools was either .0556 for the .05 IUCC conditions or .2143 for the .15 IUCC conditions. The next sections describe the bias associated with the estimates of the variance between students and the variance between high schools for conditions in which the correlation between residuals was zero.

### ***Variance Between Students***

The descriptive statistics for the random effect estimates of the variance between students under the zero correlation between residuals conditions are given in Table 10 and the bias statistics follow in Table 11.

### ***Relative Bias***

There were some problems with the estimates of the variance between students when assuming HLM, with excessive bias noted for several conditions (see Table 11). Of the factors investigated, only the number of feeder schools and



the IUCC seemed to impact the relative bias of the estimate of the variance between students. There were no problems with bias when assuming CCREM.

The IUCC had a strong effect on the accuracy of the student random effect estimates,  $F(1,11)= 5210.534$ ,  $p < .001$ ,  $\eta^2= .2500$ . Under the assumption of the HLM (Model 2), the average student random effect bias was .024 when the IUCC was .05 and .091 when the IUCC was .15. The larger the value of the IUCC, the more positively biased were the estimates of the variance between students.

Despite its statistical significance, the number of feeder schools effect on the positive bias associated with the estimates of the variance between students did not appear to be very large, as evidenced by its small eta squared value,  $F(1,11) = 20.713$ ,  $p < .01$ ,  $\eta^2= .0009$ . When assuming HLM and ignoring the clustering at the middle school level (Model 2), the average student random effect bias was .056 when there were two feeder schools and .060 when there were three. The difference between these means is negligible.

#### Standard Error Bias

There were also several problems with the standard errors associated with the estimate of the variance between students when HLM was assumed. There were no problems when assuming CCREM.

The IUCC had a strong effect on the standard errors associated with the estimates of the variance between students,  $F(1,11)= 85.685$ ,  $p < .001$ ,  $\eta^2= .3491$ .

When assuming HLM and under the condition of an IUCC of .05, the average student random effect standard error bias was only -.013 compared to -.114 when the IUCC was .15. It appears that when assuming HLM for data fitting the CCREM, the greater the IUCC, the more negatively biased are the standard errors associated with the student random effect.

The number of students per middle school had a small effect on the negative standard error bias associated with the estimates of the variance between students,  $F(1,11) = 9.977$ ,  $p < .01$ ,  $\eta^2 = .0407$ . When assuming HLM and there were twenty students per middle school, the average student random effect standard error bias was -.047 compared to -.080 when there were forty students per middle school. The larger the sample size at the middle school level, the worse was the standard error bias of the student random effect, although the differences were small.

*Table 10. Descriptive Statistics for the Estimates of the Variance Between Students in Conditions where the Correlation = 0*

Condition	Estimating Model			
	CCREM		HLM	
	Mean	SD	Mean	SD
1	1.00118	0.0609	1.02442	0.0623
2	1.00186	0.0661	1.09029	0.0744
3	0.99948	0.0416	1.02195	0.0427
4	0.99921	0.0422	1.08534	0.0515
5	0.99710	0.0473	1.02079	0.0488
6	0.99819	0.0476	1.09045	0.0554
7	1.00020	0.0335	1.02308	0.0340
8	0.99846	0.0332	1.08821	0.0423
9	1.00191	0.0640	1.02736	0.0656
10	0.99834	0.0637	1.09261	0.0752
11	1.00135	0.0431	1.02601	0.0441
12	0.99926	0.0429	1.09453	0.0547
13	1.00102	0.0494	1.02561	0.0021
14	0.99618	0.0486	1.09219	0.0575
15	0.99969	0.0340	1.02513	0.0351
16	0.99896	0.0323	1.09533	0.0422

*Note.* The value to be recovered was 1.0 for the variance in all conditions.

Table 11. Bias Statistics for the Estimates of the Variance Between Students in Conditions where Correlation = 0

Condition	Estimating Model			
	CCREM		HLM	
	Param. Est. Relative Bias	SE Relative Bias	Param. Est. Relative Bias	SE Relative Bias
1	0.00118	0.0228	0.02499	0.0103
2	0.00186	-0.0499	0.09029	-0.0986
3	-0.00052	0.0179	0.02195	0.0065
4	-0.00079	0.0066	0.08534	-0.1125
5	-0.00290	0.0155	0.02079	-0.0057
6	-0.00181	0.0174	0.09045	-0.0623
7	0.00020	-0.0187	0.02308	-0.0205
8	-0.00154	-0.0095	0.08821	-0.1612
9	0.00191	-0.0110	0.02736	-0.0235
10	-0.00166	-0.0036	0.09261	-0.0947
11	0.00135	-0.0079	0.02601	-0.0149
12	-0.00074	-0.0037	0.09453	-0.1534
13	0.00102	-0.0094	0.02561	-0.0151
14	-0.00382	0.0083	0.09219	-0.0830
15	-0.00031	-0.0272	0.02513	-0.0434
16	-0.00104	0.0252	0.09533	-0.1491

Note. Values below .05 are considered acceptable for parameter bias, and values below .10 are considered acceptable for the standard error bias.

### ***Variance Between High Schools***

The descriptive statistics for the random effect estimates of the variance between high schools are given in Table 12 and the bias statistics follow in Table 13.

#### **Relative Bias**

When assuming HLM and ignoring the clustering at the middle school level, the relative bias of the estimate of the high school variance was substantial. The positive relative bias far exceeded the acceptable level in every condition. The primary source of the bias appeared to be the number of feeder schools. The observed bias was acceptable when assuming CCREM.

Despite its statistical significance, the number of feeder schools effect on the accuracy of the high school random effect estimates did not appear very large, as evidenced by its small eta squared value,  $F(1,11)= 52.007$ ,  $p < .001$ ,  $\eta^2 = .0018$ . When assuming HLM, the average high school random effect bias was .566 when there were two feeder schools and .521 when there were three. The difference between these means is negligible.

#### **Standard Error Bias**

There were no problems with the standard errors associated with the estimates of the variance between high schools, as all estimates for both the

CCREM and HLM fell within Hoogsland and Boomsma's (1998) acceptable range.

*Table 12. Descriptive Statistics for the Estimates of the Variance Between High Schools in Conditions where the Correlation = 0*

Condition	Estimating Model			
	CCREM		HLM	
	Mean	SD	Mean	SD
1	0.05697	0.0382	0.08722	0.0398
2	0.21627	0.0892	0.33570	0.1088
3	0.05689	0.0282	0.08779	0.0315
4	0.21385	0.0766	0.33198	0.0962
5	0.05656	0.0294	0.08722	0.0305
6	0.21545	0.0673	0.33407	0.0830
7	0.05582	0.0216	0.08771	0.0240
8	0.21241	0.0566	0.33394	0.0748
9	0.05513	0.0367	0.08398	0.0386
10	0.21322	0.0903	0.32091	0.1078
11	0.05667	0.0276	0.08438	0.0306
12	0.21734	0.0761	0.32923	0.0950
13	0.05593	0.0288	0.08431	0.0299
14	0.21694	0.0667	0.32776	0.0812
15	0.05574	0.0200	0.08440	0.0229
16	0.21676	0.0579	0.33002	0.0728

*Note.* In the odd numbered conditions, the value to be recovered for the variance was .0556 and in the even numbered conditions, it was .2143.

Table 13. Bias Statistics for the Estimates of the Variance Between High Schools in Conditions where Correlation = 0

Condition	CCREM		HLM	
	Param. Est. Relative Bias	SE Relative Bias	Param. Est. Relative Bias	SE Relative Bias
1	0.02464	0.0170	0.56867	-0.0279
2	0.00919	-0.0072	0.56648	-0.0245
3	0.02317	-0.0096	0.57895	-0.0221
4	-0.00212	-0.0163	0.54914	0.0046
5	0.01734	-0.0160	0.56876	-0.0386
6	0.00538	0.0037	0.55890	-0.0302
7	0.00402	-0.0237	0.57754	-0.0250
8	-0.00880	0.0137	0.55830	-0.0095
9	-0.00838	0.0351	0.51048	-0.0096
10	-0.00504	-0.0447	0.49746	-0.0487
11	0.01933	-0.0180	0.51767	-0.0206
12	0.01418	-0.0147	0.53628	0.0105
13	0.00596	-0.0047	0.51638	-0.0293
14	0.01232	0.0061	0.52943	-0.0197
15	0.00250	0.0258	0.51806	-0.0035
16	0.01150	-0.0090	0.54001	0.0076

Note. Values below .05 are considered acceptable for parameter bias, and values below .10 are considered acceptable for the standard error bias.



***Random Effects when the correlation is .40.***

Descriptive statistics for the student-level random effects in the conditions in which the correlation between residuals was .40 are listed in Table 14 and the descriptive statistics for the high school-level random effects are listed in Table 15. Because the data were generated so that only similar middle schools and high schools were paired, the true IUCC was not known. Therefore, bias was not calculated for these conditions. Based on the following tables, however, it appears that the estimates of the variance between students were not very biased, but that the estimate of the variance between high schools was overestimated.

Table 14. Descriptive Statistics for the Estimates of the Variance Between Students in Conditions where the Correlation = .40

Condition	Estimating Model			
	CCREM		HLM	
	Mean	SD	Mean	SD
1	0.99539	0.0620	1.00858	0.0623
2	1.00530	0.0601	1.03096	0.0618
3	0.99781	0.0422	1.00711	0.0425
4	0.99763	0.0420	1.02650	0.0445
5	0.99332	0.0482	1.00339	0.0483
6	0.99730	0.0474	1.02042	0.0485
7	0.99646	0.0335	1.00360	0.0335
8	0.99946	0.0344	1.01894	0.0349
9	0.99326	0.0635	1.00687	0.0641
10	0.99593	0.0629	1.03073	0.0647
11	0.99847	0.0426	1.00861	0.0427
12	0.99781	0.0431	1.03002	0.0456
13	0.99430	0.0480	1.00515	0.0482
14	0.99557	0.0498	1.02097	0.0504
15	0.99930	0.0341	1.00758	0.0345
16	0.99965	0.0332	1.02212	0.0342

Note. The value to be recovered for the variance was 1.0 in all conditions.

Table 15. Descriptive Statistics for the Estimates of the Variance Between High Schools in Conditions where the Correlation = .40

Condition	Generating Model			
	CCREM		HLM	
	Mean	SD	Mean	SD
17	0.09878	0.0554	0.14165	0.0509
18	0.39750	0.1668	0.54171	0.1604
19	0.10224	0.0497	0.14162	0.0442
20	0.37971	0.1437	0.54951	0.1468
21	0.11158	0.0486	0.14787	0.0406
22	0.46748	0.1343	0.57403	0.1249
23	0.11803	0.0408	0.14836	0.0360
24	0.46060	0.1190	0.56668	0.1230
25	0.09868	0.0573	0.14038	0.0541
26	0.38635	0.1716	0.54952	0.1616
27	0.10154	0.0485	0.14125	0.0437
28	0.36498	0.1387	0.54330	0.1533
29	0.11209	0.0481	0.14705	0.0418
30	0.45363	0.1445	0.56498	0.1304
31	0.11579	0.0386	0.14765	0.0346
32	0.44204	0.1244	0.56219	0.1228

*Note.* In the odd numbered conditions, the value to be recovered for the variance was .0556 and in the even numbered conditions, it was .2143.

## Fit Indices

Table 16 presents the means and standard deviations of the Akaike's Information Criterion (AIC) and Schwarz's Bayesian Criterion (SBC) fit statistics for estimation both under the CCREM and HLM for each condition. Also provided is the percentage of iterations in which the correct model (CCREM) was identified using the AIC. The SBC correctly identified the CCREM as the better fitting model in 100% of the replications across all conditions. This is consistent with previous research showing that Bayesian Model Averaging, such as Schwarz's Bayesian Criterion, has outperformed competing methods for model selection with both theoretical results and simulation studies (see Raftery & Yeng, 2003 for a more detailed account).

None of the factors investigated had moderate to strong effects on the proportion of correct model identifications using the AIC, but several factors had small effects.

The correlation of the residuals at level 2 had a small effect on the percentage of correct model identifications using the AIC,  $F(1,26) = 109.443$ ,  $p < .001$ ,  $\eta^2 = .0203$ . When there is no correlation, the correct model was chosen an average of 93.18% of the time compared to only 69.46% of the time when the correlation is .40.

The IUCC also had a small effect on the fit indices,  $F(1,26)= 81.306$ ,  $p < .001$ ,  $\eta^2= .0151$ . When the IUCC was .05, the CCREM was only chosen 71.10% of the time compared to 91.54% of the time when it was .15.

Despite its statistical significance, the average middle school size effect on the proportion of correct model identifications using the AIC did not appear very large, as evidenced by its small eta squared value,  $F(1,26)= 30.245$ ,  $p < .001$ ,  $\eta^2= .0056$ . When twenty students per middle school were used the correct model was only chosen 75.09% of the time compared to 87.56% when forty students per middle school were used. The difference between these means is negligible.

Table 16. Means (and Standard Deviations) for Goodness of Fit Indices

Condition	CCREM		HLM		AIC% Correct
	AIC	SBC	AIC	SBC	
	M(SD)	M(SD)	M(SD)	M(SD)	
1	1657.3 (47)	1651.6 (47)	1660.4 (48)	1663.2 (47)	69.7
2	1697.0 (48)	1691.0 (48)	1720.6 (50)	1723.4 (50)	99.4
3	3381.3 (57)	3375.4 (57)	3391.8 (58)	3394.6 (58)	91.4
4	3429.9 (58)	3423.9 (58)	3491.3 (64)	3494.1 (64)	100.0
5	2745.2 (58)	2739.2 (58)	2751.2 (59)	2755.0 (59)	83.5
6	2813.2 (61)	2807.2 (61)	2855.4 (65)	2859.2 (65)	100.0
7	5621.4 (75)	5615.4 (75)	5639.9 (75)	5643.8 (75)	98.8
8	5701.4 (75)	5695.4 (75)	5809.7 (86)	5813.5 (86)	100.0
9	1618.0 (49)	1612.3 (49)	1621.4 (50)	1624.2 (50)	71.6
10	1657.8 (49)	1651.8 (49)	1682.1 (52)	1684.9 (52)	99.1
11	3345.8 (62)	3339.9 (62)	3357.1 (63)	3359.9 (62)	94.1
12	3394.9 (63)	3388.9 (63)	3462.2 (69)	3465.0 (69)	100.0
13	2681.4 (67)	2675.5 (67)	2687.2 (67)	2691.0 (67)	84.0
14	2745.5 (65)	2739.5 (65)	2787.6 (69)	2791.4 (69)	100.0
15	5555.4 (81)	5549.4 (81)	5575.8 (82)	5579.6 (82)	99.3
16	5640.2 (77)	5634.2 (77)	5754.4 (87)	5758.2 (87)	100.0
17	1659.3 (48)	1653.7 (48)	1660.4 (48)	1663.2 (48)	51.2
18	1697.1 (46)	1691.3 (46)	1701.4 (47)	1704.2 (47)	73.5
19	3384.5 (57)	3378.7 (57)	3386.7 (57)	3389.5 (57)	59.5
20	3428.7 (57)	3422.7 (57)	3441.8 (59)	3444.6 (59)	91.2
21	2747.7 (60)	2742.0 (60)	2748.8 (60)	2752.6 (60)	50.3
22	2816.3 (60)	2810.4 (60)	2820.7 (60)	2824.5 (60)	73.2
23	5619.2 (77)	5613.4 (77)	5621.3 (77)	5625.1 (77)	60.2
24	5698.2 (79)	5692.3 (79)	5709.5 (79)	5713.3 (79)	91.9
25	1618.7 (50)	1613.1 (50)	1619.7 (50)	1622.5 (50)	49.9
26	1659.5 (50)	1653.6 (50)	1664.8 (50)	1667.6 (50)	75.5
27	3345.1 (61)	3339.4 (61)	3347.7 (61)	3350.5 (61)	62.1
28	3392.8 (63)	3386.9 (63)	3407.8 (65)	3410.6 (65)	93.4
29	2684.9 (64)	2679.2 (64)	2686.0 (64)	2689.8 (64)	48.0
30	2747.3 (67)	2741.4 (67)	2752.4 (67)	2756.2 (67)	72.5
31	5563.0 (77)	5557.1 (77)	5565.7 (77)	5569.5 (77)	64.0
32	5633.4 (80)	5627.5 (80)	5647.2 (81)	5651.0 (81)	95.0

Note. The proportion of correct model identifications is reported for the AIC. The SBC correctly identified the CCREM as the better fitting model in 100% of the replications across all conditions.

## STUDY 2

### Descriptive Statistics

The NELS:88 data set contains data for 19,392 students nested within 1,014 middle schools and 1,698 high schools. Descriptive statistics of the variables included in the analyses are provided in Table 17 below. The percentage of students in the free or reduced lunch program and the size of the middle school were measured on ordinal scales.

*Table 17. Descriptive Statistics of variables included in analysis*

	<b>N</b>	<b>Mean</b>	<b>SD</b>	<b>Skewness</b>	<b>Kurtosis</b>
<b>Gender</b>	17,322	1.503	0.500	-.013	-2.00
<b>Per_free</b>	16,454	1.486	0.840	-.148	-.598
<b>Size</b>	17,465	3.752	1.563	.470	-.517
<b>Test score</b>	17,322	50.976	10.117	-.006	-1.083

In the NELS:88 data set, there are an average of 19.12 students per middle school. In addition, partitioning the total variance into its components yielded the following unconditional IUCCs: .7522 at the student level, .1750 at the middle school level, and .0728 at the high school level. Finally, typically two or three middle schools feed into a given high school. These values are similar to the conditions generated in the simulation study.

The coding system used by NELS:88 is as follows. For the variable *gender*, males were coded with a one and females with a two. The variable *per\_free* refers to the percentage of students at a given high school that were in the free and reduced lunch program. *Per\_free* was measured on a 0 to 3 scale with values as follows: 0 = 0% of the students in the program, 1 = between 1% and 10% of the students in the program, 2 = between 11% and 50% of the students in the program, and 3 = between 51% and 100% of the students in the program. The variable *size* refers to the total enrollment of the middle school and had the following scale values : 1 = 1-199 students, 2 = 200-399 students, 3 = 400-599 students, 4 = 600-799 students, 5 = 800-999 students, 6 = 1000-1199 students, and 7 = 1200 or more students. The variable *test score* refers to the dependent variable, a standardized Reading/Math test score. The test was administered when the students were in tenth grade and was measured on an interval scale. The ordinal variables were treated as though they were interval-scaled. The skew and kurtosis statistics are included in Table 17 above.

### **Fixed Effect Estimates**

The three models resulted in similar results for the fixed effect estimates. The fixed effect estimates for the three predictor variables and the associated statistical significance for each of the three models are listed in Table 18 below. The largest difference between models was for the variable *size*, the middle



school-level variable. Despite the observed difference, the coefficient associated with the variable *size* was not statistically significant at the .05 alpha-level for any of the models.

Table 18. Fixed Effect Estimates

	Model A	Model B	Model C
<b>Intercept</b>	56.0069 **	55.9049 **	55.6351 **
<b>Per_free</b>	-3.4799 **	-3.4541 **	-3.2131 **
<b>Size</b>	-0.1778	-0.1654	-0.1863
<b>Gender</b>	0.7493 **	0.7512 **	0.7332 **

Note. Model A involved deleting cases, Model B involved disaggregating the variable *size*, and Model C involved correctly modeling via CCREM.

\*\* p < .0001

\* p < .05

### Fixed Effect Standard Error Estimates

Estimation of the three models provided similar results for the standard errors of the fixed effects as well. The standard errors for the fixed effect estimates for each of the three models are listed in Table 19 below. The standard errors associated with the fixed effect estimates for the middle school variable *size* were smaller for Models *A* and *B*, but the difference appeared small. The fixed effect for the middle school variable is the one that was modeled incorrectly in Models *A* and *B*.

Table 19. Fixed Effect Standard Errors

	Model A	Model B	Model C
<b>Intercept</b>	0.5010	0.4795	0.5013
<b>Per_free</b>	0.1760	0.1728	0.1722
<b>Size</b>	0.0958	0.0892	0.0959
<b>Gender</b>	0.1557	0.1525	0.1521

*Note.* Model A involved deleting cases, Model B involved disaggregating the variable size, and Model C involved correctly modeling via CCREM.

### Random Effect Estimates

The random effect estimates and associated statistical significance for each of the three models are listed in Table 20 below. Very minor differences were found for the estimates of the variance between students across the three models. However, when estimating the variance between high schools, *Models A* and *B* had similar results, but *Model C*'s (the CCREM) estimate was roughly half that of the other two models. In *Model C*, the remaining variance was estimated to be between middle schools.

Table 20. Random Effect Estimates

Variance between	Model A	Model B	Model C
<b>Middle schools</b>	N/A	N/A	8.109**
<b>High schools</b>	16.138**	15.976**	8.374**
<b>Students</b>	76.503**	76.544**	76.052**

*Note.* Model A involved deleting cases, Model B involved disaggregating the variable size, and Model C involved correctly modeling via CCREM.

\*\* p < .0001

\* p < .05

### Random Effect Standard Error Estimates

The standard errors of the random effect estimates for each of the three models are listed in Table 21 below. There are minor differences for the standard errors associated with the variance between students. This is not true of the standard errors associated with the variance estimates between high schools. Both *Models A* and *B* have substantially smaller standard errors for this variance component.

Table 21. Random Effect Standard Errors

Variance Between	Model A	Model B	Model C
Middle schools	N/A	N/A	1.436
High schools	1.051	1.024	1.431
Students	.962	0.940	0.936

*Note.* Model A involved deleting cases, Model B involved disaggregating the variable size, and Model C involved correctly modeling via CCREM.

### Fit Indices

Finally, the AIC and SBC fit indices for each of the three models are listed in Table 22 below. Using both the AIC and SBC fit indices, *Model A* exhibits the best fit. In addition, data was found to fit *Model C* better than *Model B*. It is not surprising that *Model A* was identified as having the best fit. This model deleted cases and forced a hierarchy. *Model A* was based on different sample of cases than the other two models.

Table 22. *Fit Indices*

	<b>Model A</b>	<b>Model B</b>	<b>Model C</b>
<b>AIC</b>	99792.1	104335.8	104294.0
<b>SBC</b>	99806.2	104345.9	104288.0

*Note.* Model A involved deleting cases, Model B involved disaggregating the variable size, and Model C involved correctly modeling via CCREM.

## Chapter 5: Discussion

### SUMMARY

Study One was a Monte Carlo Simulation in which data were generated such that students were cross-classified by middle school and high school. Both a CCREM and a HLM (where the middle school predictor variable was included as a student-level predictor) were estimated for each replication.

The fixed effect estimates were virtually unaffected by the model being fit. Neither the model assumed nor the condition simulated seemed to affect the precision of the fixed effect estimates. However, the standard errors associated with the fixed effect estimates presented some evidence of bias. The standard error estimates associated with the middle school predictor variable were negatively biased across all conditions when HLM was assumed instead of CCREM. Both the correlation of the residuals and the IUCC had moderate effects on the negative standard error bias of the fixed effect coefficient of  $Z$  ( $\gamma$ ). The average middle school size had a small effect.

The middle school predictor,  $Z$ , was modeled inappropriately when assuming HLM. Instead of being used at the middle school-level, it was treated as a student-level variable. Previous research with traditional HLM has shown that incorrect modeling can lead to biased standard errors, so this finding is consistent. Raudenbush and Bryk note that “misestimated standard errors occur

with multilevel data when we fail to take into account the dependence among individual responses within the same organization”( Raudenbush & Bryk, 2002, p. 100). The resulting impact of underestimated standard errors is an inflation in the Type I error rate. Researchers are more likely to conclude that a variable is statistically significant when it actually is not.

Here, by assuming HLM and modeling middle school as part of level 1 instead of assuming CCREM, the variance between middle schools was being modeled incorrectly. The clustering at the middle school-level was ignored. This likely explains the observed negative standard error bias of the middle school-level variable. The correlation of the middle school and high school residuals appears to disrupt the estimating procedure substantially. In addition, the more variance there was at the middle school-level, the more negatively biased were the standard errors for the variables at that level. Logically, incorrectly modeling the variance between middle schools resulted in more bias when there was more of this variance. Finally, larger overall sample sizes measured by the number of students at the middle school level, were associated with slightly more negatively biased standard errors for the fixed effect  $\gamma$ , although the effect was small. This result is likely due to what is known as the *design effect* in cluster sampling (Kalton, 1983). The magnitude of the design effect is influenced by the per-cluster sample size. The larger the per-cluster sample size, the larger is the design

effect, and as a result, the more biased are estimates when the within-cluster homogeneity is ignored. *Model 2* ignored the within-cluster homogeneity, so this could potentially explain why the bias under the assumption of HLM increased as the sample size increased. The sample size effect was very small, though, so the observed difference might be inconsequential.

The random effect estimates were also biased when HLM was assumed instead of CCREM. Substantial differences in the relative bias of the random effects for the between-student variance were found when the IUCC was manipulated. As the variance between high schools (and middle schools) increased, the accuracy of the estimation of the variance between students decreased when assuming HLM. The larger the variance, the more positively biased were the estimates of the variance between students. Again, it is more important to model the cross-classification appropriately when there is more variance between each cross-classified factor.

The standard errors associated with the random effect estimates of the variance between students were negatively biased when HLM was used instead of CCREM. The standard error bias varied as a function of the number of students and the IUCC. Although it was a statistically significant factor, the number of students had only a small effect. Again, this is potentially due to the design effect. The differences were small, though, so there was little practical

significance. In addition, the more variance there was between high schools, the more biased were the student-level estimates and standard errors. Again, underestimated standard errors can lead to an inflation of the Type I error rate.

The random effect estimates of the variance between high schools also exhibited some bias. When HLM was assumed instead of CCREM, the relative bias was in the unacceptable range in all conditions. The HLM estimates were positively biased. However, none of the factors examined had a substantial impact on the observed bias. The standard errors associated with the high school random effect were robust. For applied researchers assuming HLM although analyzing cross-classified data, having an inflated estimate of the variance between high schools could lead to adding high school-level variables in an attempt to explain more of this variance. However, as was the case in this study, because the variance might really be at the middle school level, adding explanatory variables at the high school level will not help.

Several factors seemed to influence the proportion of iterations in which CCREM was chosen as the correct model over HLM using the AIC, but none had a substantial effect. The correlation of residuals and the IUCC between each cross-classified factor each had a small effect. When there was no correlation between the residuals at level 2 (meaning middle schools and high schools were randomly paired) and when there was a larger proportion of variance between



middle schools and high schools, the correct model was chosen more often, although the differences were small.

The presence of correlated residuals seems to adversely impact all estimates except for the fixed effect estimates. When the residuals were not correlated, CCREM was chosen more consistently. Finally, CCREM was chosen more often when there was more variance between middle schools. Logically, the more variance there is between each cross-classified factor, the more necessary it is to model the structure appropriately. It was expected that CCREM would exhibit better fit, since it was the more parameterized of the two models being investigated.

Study Two found similar results to the simulation study. The main finding was that modeling the cross-classified situation incorrectly led to differences in the estimates of the variance between high schools. The two HLM models appeared to transfer the variance that should have been identified between middle schools to the high school-level. Both HLM models had estimates of the variance between high schools that were roughly twice as large as the CCREM estimate. In addition, the standard errors associated with the variance between high schools were much lower in the HLM models than in the CCREM. This is consistent with the findings of the simulation study, which showed that standard errors were underestimated when variance within a cluster was ignored.

## LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

This dissertation provided only an introductory investigation into how the results of assuming the CCREM differ from those when assuming different possible HLM models. Consequently, there are several limitations to this study.

In the simulation study, only moderate effect sizes were used. The conclusions provided here are applicable only to situations in which the effects of the predictor variables are moderate in nature. Furthermore, the true effects of each of the variables were the same at each level. Future research should vary the effect sizes at the different levels.

Another major limitation of the simulation study involves the generation of the data. Half of the thirty-two conditions involved a moderate correlation of the residuals between level-2 factors. This presented several problems. Two things were actually occurring when the data were generated this way. First of all, the characteristics of the middle school and high school were intentionally made similar ( $\rho=.40$ ) in all cells. Second, sorting the data and pairing the middle schools and high schools created an inherent additional dependency in the data. Only *similar middle schools* could be paired with a given high school. It is possible to have a correlation between middle school and high school characteristics without creating the dependency *among* the middle schools paired with a high school via the sorting that was done. Therefore, the similarity of the

middle schools that could be paired with a given high school and the similarity of each paired middle school and high school were confounded in this study. It is impossible to tell which one led to the observed results. Future research should distinguish between these two factors.

The simulation study is further limited in that it only fit one alternative HLM. There are other ways to manipulate the HLM to describe a cross-classified data structure other than the way presented here. Disaggregating the middle school predictor variable is only one possible alternative to assuming the CCREM. Additional as yet unexplored models may be more or less accurate. Future research should test other manipulations of HLM.

It is also possible that the findings presented here are due to the specific levels of each factor chosen. Specifically, the values for the conditions did not vary substantially and do not cover the full range of values. For instance, more insight might have been gained from investigating additional values for the correlation between residuals at level 2. In addition, other values for the average number of students per middle school could be investigated. This is especially true of the IUCCs chosen. While the values were chosen because they reflect the actual values reported in the literature, additional values could be investigated.

In addition to the actual values of the IUCCs, this study was limited in that the same value was generated for both the middle school and high school IUCC.

These values are usually not the same in reality. Varying the levels of the IUCCs per cross-classified factor might provide more insight into conditions impacting the need to assume CCREM.

Similarly, the number of middle schools and high schools were generated to be equal. Empirical data sets, such as NELS:88 usually have differing numbers of each type of cross-classified factor. Future research might vary the levels of each cross-classified factor to see how the results differ from those presented here.

The number of feeder schools factor also presented a problem. The way the data were generated, the cell size was not kept constant. In the two feeder school condition, not only are there more empty cells than in the three feeder school condition, but also the average cell size is larger because the same number of students is being distributed to fewer cells. Thus, the number of empty cells and the average cell size is confounded. Future research should attempt to find ways to generate cross-classified data that avoid this problem.

Finally, the CCREM modeling presented here is simplistic. It is possible to have interaction effects as well as to model random effects in which the effect of a variable of one cross-classified factor (say, middle schools) varies across levels of the other cross-classified factor (say, high schools). Future research

might involve assessment of the performance of CCREM in the presence of interactions and random effects.

The primary limitation of Study Two is the extremely large sample size. Most data sets do not contain 19,000 records. The results could have been similar across the models investigated due to the large sample size. The results might be more discrepant if there were fewer subjects. Future research should replicate the study on smaller data sets.

A secondary limitation to the study with empirical data is that the analysis presented here did not take into account the complex sampling design employed in the NELS:88 data set. The data set involves a complicated weighting design. Future research should accommodate the sampling design to see how the results might differ from those presented here.

## **EDUCATIONAL IMPORTANCE AND CONCLUSION**

In educational research, students are often nested within cross-classifications. The two most common examples are students within middle schools and high schools and students within neighborhoods and schools. With the recent passage of the No Child Left Behind Act (NCLB; Public Law 107-110, 2001), cross-classified situations are likely to become even more common, so

understanding the functioning of the techniques available to model these situations has become even more important.

A cornerstone of NCLB is an emphasis on increased accountability. NCLB strengthens accountability for Title I schools by requiring states to implement statewide accountability systems. The accountability systems include annual testing for all students in grades 3-8 and annual statewide progress objectives ensuring that all groups of students reach proficiency by the end of the 2013-2014 academic year. Test results and state progress objectives must be broken down by poverty, race, ethnicity, disability, and limited English proficiency to ensure that no social group of students is “left behind.” Schools and school districts that do not make adequate yearly progress (AYP) towards statewide proficiencies become, over time, subject to improvement, corrective action and other measures aimed towards getting them back on course towards reaching state-defined standards. Because of the increased importance placed on tests and assessments, it is essential that tools exist to properly measure concepts such as “achievement” and “progress.” With NCLB, it has become increasingly important to be able to adequately model change over time.

In addition to the focus on accountability, a cornerstone of the NCLB Act is the availability of “more choices for parents and students.” The NCLB Act significantly increases the choices available to the parents of students attending

Title I schools that fail to meet state standards. Under NCLB, local education agencies are required to give students who attend schools which have been identified as having deficiencies of any level the opportunity to attend a “better” school within the school district. As more and more students begin to transfer between schools, the evaluation of institutional effectiveness needs to be enhanced. It is not sufficient to simply control for students’ performance before entering a school: “‘value-added’ approaches to comparing Secondary schools adjusting for intake achievement but not for previous schooling, may be seriously deficient” (Goldstein & Sammons, 1997, p. 228-229). A result of NCLB is that there will likely be more students attending schools outside of their neighborhood and more students changing schools over time – the two most common cross-classified situations in educational research. Use of CCREM provides a flexible modeling technique that is well matched to these cross-classified data structures, so it is imperative that the performance of these methods is well understood.

Because measuring students’ and schools’ educational achievement and progress has taken on added emphasis under the NCLB Act, and because there are likely to be more mobile students under NCLB, understanding the functioning of the statistical methods available to evaluate performance is of utmost importance.

Both the simulated and empirical data analyses conducted here demonstrate that assuming HLM instead of CCREM leads to different results. As

seen in the simulation study, while the fixed effect estimates are similar, the standard errors associated with the variables that are modeled incorrectly (here, the middle school–level variable) are negatively biased. In addition, the random effect estimates and standard error estimates for HLM are biased. Biased standard error estimates can lead to incorrect conclusions regarding the statistical significance of the parameter estimates, including an inflation in the Type I. error rate. The suggestion here is that CCREM be assumed instead of HLM whenever the data are cross-classified. CCREM is becoming more widely available in statistical software packages, and the computing time for these complicated models is steadily decreasing. While the results of CCREM can be more difficult to interpret than HLM, failing to model the data correctly can lead to incorrect conclusions.



## Appendix A

### SAS Program for Two Feeder Schools

```
/*Generating middle school and high school residuals based on a known
correlation*/

%macro generate;

proc iml;
  resid=normal(j(50,2,0));/*change this to 30 in other condition*/
  corr={1.00 .40,
        .40 1.00}; /*change to .00 to remove correlation*/
  corr_root=root(corr);
  fin_resid=resid*corr_root;
  fin_resid2=.2358#fin_resid; /*change to .0556 to change IUCC to .05*/

  create sasresid from fin_resid2;
  append from fin_resid2;

quit;

data sasresid;set sasresid;
rename COL1= ms_res;
rename COL2= hs_res;
run;

proc sort data=sasresid;by ms_res;
run;

/*main feeder*/

data one;set sasresid;
retain counter 0;
counter=counter+1;
ms_id=counter;
hs_id=counter;
ms_size= int(20+2*rannor(0)); /*change to 40 in other condition*/
cell_size = int(.60*ms_size); /*number of students*/
size= 50+10*rannor(0);
middleschool_effect= .50*size + ms_res;
free=50+10*rannor(0);
highschool_effect=.50*free + hs_res;
run;

/*secondary feeder*/

data onemore;set one;
keep ms_id hs_id cell_size;
ms_id=counter;
hs_id=counter+1;
cell_size=int (.40*ms_size);
```

```

if cell_size = 0 then cell_size = 1;
if hs_id gt 50 then hs_id=1; /* change to 30 in other condition*/
ms_res=ms_res;
run;

proc sort data=onemore;by ms_id;run;
data merged;
merge one(in=A) onemore (in=B);by ms_id;
if A and B;
keep ms_id hs_id cell_size ms_res size middleschool_effect;
run;

proc sort data=onemore;by hs_id;run;
data merged1;
merge one(in=A) onemore(in=B);by hs_id;
if A and B;
keep ms_id hs_id hs_res free highschool_effect;
run;

proc sort data=merged;by hs_id;run;
data merged2;
merge merged(in=A) merged1(in=B);by hs_id;
if A and B;
keep ms_id hs_id cell_size ms_res hs_res size free middleschool_effect
highschool_effect;
run;

data grand;
keep ms_res hs_res ms_id hs_id cell_size size free middleschool_effect
highschool_effect;
set one merged2;
run;

/*creating student observations*/

data student;set grand;
do student= 1 to cell_size;
SAT=50+10* rannor(0);
student_effect=.50*SAT + 1*rannor(0);
test_score=100+student_effect+middleschool_effect+highschool_effect;
iter=&i;
output;
end;
run;
%mend;

%macro runall;
%do i =1 %to 1000; /*number of iterations*/
%generate;
proc datasets;
append base=fulldata data=student;
%end;
%mend;

```

```

proc iml;
%runall;
quit;

ods output covparms=random_effects;
ods output solutionf = fixed_effects;
ods output FitStatistics = fit_statistics;

/* Fitting the cross-classified model */

proc sort data=fulldata;by iter;run;
proc mixed data= fulldata noclprint covtest noinfo noitprint;
class ms_id hs_id;
model test_score= free size SAT/s;
random int/sub= ms_id;
random int/sub=hs_id;
by iter;
run;

/*transposing random effects*/

data random_effects;set random_effects;by iter;
if Subject = ' ' then Subject = 'resid';
run;

proc transpose data=random_effects out=random_trans;
var estimate stderr zvalue probz;
id subject;
by iter;
run;

data randeffs;set random_trans;by iter;
retain msest hsest resest msstd hsstd resstd msZ hsZ resZ msProbZ hsProbZ
resProbZ iter;

if _NAME_ = 'Estimate' then do;
msest=ms_id;
hsest=hs_id;
resest=resid;
end;
if _NAME_='StdErr' then do;
msstd=ms_id;
hsstd=hs_id;
resstd=resid;
end;
if _NAME_='ZValue' then do;
msZ=ms_id;
hsZ=hs_id;
resZ=resid;
end;
if _NAME_='ProbZ' then do;
msProbZ=ms_id;

```

```

hsProbZ=hs_id;
resProbZ=resid;
end;
keep msest hsest resest msstd hsstd resstd msZ hsZ resZ msProbZ hsProbZ
resProbZ iter;
if last.iter;
run;

proc transpose data = fixed_effects out = fixed_trans;
var estimate StdErr tvalue Probt;
id effect;
by iter;
run;

/*renaming fixed effects*/

data fixedeff;set fixed_trans; by iter;
retain intercept_est free_est SAT_est size_est intercept_std free_std
SAT_std size_std intercept_t free_t SAT_t size_t intercept_probt
free_probt SAT_probt size_probt iter;

if _NAME_ = 'Estimate' then do;
intercept_est = intercept;
free_est = free;
SAT_est = SAT;
size_est=size;
end;
if _NAME_='StdErr' then do;
intercept_std = intercept;
free_std = free;
SAT_std = SAT;
size_std=size;
end;
if _NAME_='tValue' then do;
intercept_t = intercept;
free_t = free;
SAT_t = SAT;
size_t=size;
end;
if _NAME_='Probt' then do;
intercept_probt = intercept;
free_probt = free;
SAT_probt = SAT;
size_probt=size;
end;
keep intercept_est free_est SAT_est size_est intercept_std free_std
SAT_std size_std intercept_t free_t SAT_t size_t intercept_probt
free_probt SAT_probt size_probt iter;
if last.iter;
run;

```

```

/*Getting Fit indices in tabular form*/

proc transpose data = fit_statistics out = fit_trans;
var value;
id Descr;
by iter;
run;

/*Renaming Fit Statistics*/

data fit_indices;set fit_trans;by iter;
_2RLL = _2_Res_Log_Likelihood;
AIC = AIC__smaller_is_better_;
AICC= AICC__smaller_is_better_;
BIC = BIC__smaller_is_better_;
keep iter AIC BIC;
run;

data results;
merge randeffs
      fixedeff
      fit_indices;
by iter;
run;

/*Now, repeating the above for the standard HLM */

ods output covparms = random_effects2;
ods output solutionf = fixed_effects2;
ods output FitStatistics = fit_statistics2;

/*fitting the standard HLM model with HS as the clustering unit*/

proc mixed data = fulldata noclprint covtest noinfo noitprint;
class hs_id;
model test_score = size free SAT/s;
random int/sub = hs_id;
by iter;
run;

/*Getting random effects in tabular form */

data random_effects2;set random_effects2;by iter;
if Subject = ' ' then Subject = 'resid';
run;
proc transpose data = random_effects2 out = random_trans2;
var estimate StdErr ZValue ProbZ;
id subject;
by iter;
run;

```

```

/*renaming random effects*/

data randeff2;set random_trans2; by iter;
retain hsest2 resest2  hsstd2 resstd2  hsZ2 resZ2  hsProbZ2 resProbZ2
iter;

if _NAME_ = 'Estimate' then do;
hsest2=hs_id;
resest2=resid;
end;
if _NAME_='StdErr' then do;
hsstd2=hs_id;
resstd2=resid;
end;
if _NAME_='ZValue' then do;
hsZ2=hs_id;
resZ2=resid;
end;
if _NAME_='ProbZ' then do;
hsProbZ2=hs_id;
resProbZ2=resid;
end;
keep hsest2 resest2  hsstd2 resstd2  hsZ2 resZ2  hsProbZ2 resProbZ2 iter;
if last.iter;
run;

/* Getting fixed effects in tabular form*/

proc transpose data = fixed_effects2 out = fixed_trans2;
var estimate StdErr tvalue Probt;
id effect;
by iter;
run;

data fixedeff2;set fixed_trans2; by iter;
retain intercept_est2 free_est2 SAT_est2 size_est2 intercept_std2
free_std2 SAT_std2 size_std2 intercept_t2 free_t2 SAT_t2 size_t2
intercept_probt2 free_probt2 SAT_probt2 size_probt2 iter;

if _NAME_ = 'Estimate' then do;
intercept_est2 = intercept;
free_est2 = free;
SAT_est2 = SAT;
size_est2=size;
end;
if _NAME_='StdErr' then do;
intercept_std2 = intercept;
free_std2 = free;
SAT_std2 = SAT;
size_std2=size;
end;
if _NAME_='tValue' then do;
intercept_t2 = intercept;
free_t2 = free;
SAT_t2 = SAT;

```

```

size_t2=size;
end;
if _NAME_='Probt' then do;
intercept_probt2 = intercept;
free_probt2 = free;
SAT_probt2 = SAT;
size_probt2=size;
end;
keep intercept_est2 free_est2 SAT_est2 size_est2 intercept_std2 free_std2
SAT_std2 size_std2 intercept_t2 free_t2 SAT_t2 size_t2 intercept_probt2
free_probt2 SAT_probt2 size_probt2 iter;
if last.iter;
run;

/*Getting Fit indices in tabular form*/

proc transpose data = fit_statistics2 out = fit_trans2;
var value;
id Descr;
by iter;
run;

/*Renaming Fit Statistics*/

data fit_indices2;set fit_trans2;
_2RLL2 = _2_Res_Log_Likelihood;
AIC2 = AIC__smaller_is_better_;
AICC2= AICC__smaller_is_better_;
BIC2 = BIC__smaller_is_better_;
keep iter AIC2 BIC2;
run;

/*Merges fixed effects, random effects, and statistical tests */

data results2;
merge randeff2
      fixedeff2
      fit_indices2;
by iter;
run;

/*Summarizes data across iterations*/

data final;
merge results
      results2;
by iter;
run;

/*Descriptive Statistics for all of the estimates*/

proc means data=final;run;

```

```

/*Proportion of time CCREM is chosen over HLM*/

data fit;set final;
by iter;
AIC_fit = AIC2-AIC;
BIC_fit=BIC2-BIC;
if AIC_fit ge 0 then Afit = 1;
else Afit = 0;
if BIC_fit ge 0 then Bfit = 1;
else Bfit = 0;
run;

proc freq;
tables Afit Bfit;
run;

data fixed_bias;set final;
by iter;
keep msest hsest resest intercept_est free_est SAT_est size_est hsest2
resest2 intercept_est2 free_est2 SAT_est2 size_est2;
run;

/*Calculating Relative Bias of the Parameter Estimates*/

proc means data=fixed_bias;
output out=fixed_means mean(msest hsest resest intercept_est free_est
SAT_est size_est hsest2 resest2 intercept_est2 free_est2 SAT_est2
size_est2)=mean_msest mean_hsest mean_resest mean_intercept_est
mean_free_est mean_SAT_est mean_size_est mean_hsest2 mean_resest2
mean_intercept_est2 mean_free_est2 mean_SAT_est2 mean_size_est2;
run;

data parameter_bias;set fixed_means;
msest_bias=(mean_msest -.0556)/.0556; /*changes when IUCC changes*/
hsest_bias=(mean_hsest -.0556)/.0556; /*changes when IUCC changes*/
resest_bias=(mean_resest - 1)/1;
int_bias=(mean_intercept_est-100)/100;
free_bias=(mean_free_est-.5)/.5;
SAT_bias=(mean_SAT_est-.5)/.5;
size_bias=(mean_size_est-.5)/.5;
hsest_bias2=(mean_hsest2 -.0556)/.0556; /*changes when IUCC changes*/
resest_bias2=(mean_resest2 - 1)/1;
int_bias2=(mean_intercept_est2-100)/100;
free_bias2=(mean_free_est2-.5)/.5;
SAT_bias2=(mean_SAT_est2-.5)/.5;
size_bias2=(mean_size_est2-.5)/.5;
run;

data parameter_bias;set parameter_bias;
keep msest_bias hsest_bias resest_bias int_bias free_bias SAT_bias
size_bias hsest_bias2 resest_bias2 int_bias2 free_bias2 SAT_bias2
size_bias2;
run;

```



```

proc print data=parameter_bias;run;

/*Calculating Relative Bias of the Standard Errors*/

data se_bias;set final;
by iter;
keep msest hsest resest intercept_est free_est SAT_est size_est hsest2
resest2 intercept_est2 free_est2 SAT_est2 size_est2 msstd hsstd hsstd2
resstd resstd2 intercept_std intercept_std2 free_std free_std2 SAT_std
SAT_std2 size_std size_std2;
run;

proc means data=se_bias;
output out=se_means mean (msest hsest resest intercept_est free_est
SAT_est size_est hsest2 resest2 intercept_est2 free_est2 SAT_est2
size_est2 msstd hsstd hsstd2 resstd resstd2 intercept_std intercept_std2
free_std free_std2 SAT_std SAT_std2 size_std size_std2)=mean_msest
mean_hsest mean_resest mean_intercept_est mean_free_est mean_SAT_est
mean_size_est mean_hsest2 mean_resest2 mean_intercept_est2 mean_free_est2
mean_SAT_est2 mean_size_est2 mean_msstd mean_hsstd mean_hsstd2
mean_resstd mean_resstd2 mean_intercept_std mean_intercept_std2
mean_free_std mean_free_std2 mean_SAT_std mean_SAT_std2 mean_size_std
mean_size_std2stddev (msest hsest resest intercept_est free_est SAT_est
size_est hsest2 resest2 intercept_est2 free_est2 SAT_est2 size_est2 msstd
hsstd hsstd2 resstd resstd2 intercept_std intercept_std2 free_std
free_std2 SAT_std SAT_std2 size_std size_std2)=std_msest std_hsest
std_resest std_intercept_est std_free_est std_SAT_est std_size_est
std_hsest2 std_resest2 std_intercept_est2 std_free_est2 std_SAT_est2
std_size_est2 std_msstd std_hsstd std_hsstd2 std_resstd std_resstd2
std_intercept_std std_intercept_std2 std_free_std std_free_std2
std_SAT_std std_SAT_std2 std_size_std std_size_std2;
run;

data standard_error_bias;set se_means;
ms_se_bias=(mean_msstd- std_msest)/std_msest;
hs_se_bias=(mean_hsstd - std_hsest)/std_hsest;
residual_se_bias=(mean_resstd-std_resest)/std_resest;
intercept_se_bias=(mean_intercept_std -
std_intercept_est)/std_intercept_est;
free_se_bias=(mean_free_std-std_free_est)/std_free_est;
SAT_se_bias=(mean_SAT_std- std_SAT_est)/std_SAT_est;
size_se_bias=(mean_size_std-std_size_est)/std_size_est;
hs_se_bias2=(mean_hsstd2 - std_hsest2)/std_hsest2;
residual_se_bias2=(mean_resstd2-std_resest2)/std_resest2;
intercept_se_bias2=(mean_intercept_std2 -
std_intercept_est2)/std_intercept_est2;
free_se_bias2=(mean_free_std2-std_free_est2)/std_free_est2;
SAT_se_bias2=(mean_SAT_std2- std_SAT_est2)/std_SAT_est2;
size_se_bias2=(mean_size_std2-std_size_est2)/std_size_est2;
run;

data standard_error_bias;set standard_error_bias;

```

```
keep ms_se_bias hs_se_bias residual_se_bias intercept_se_bias
free_se_bias SAT_se_bias size_se_bias hs_se_bias2 residual_se_bias2
intercept_se_bias2 free_se_bias2 SAT_se_bias2 size_se_bias2;
run;
proc print data=standard_error_bias;
run;

/*more descriptive statistics*/

data fulldata;set fulldata;
proc means;var ms_res hs_res middleschool_effect highschool_effect
cell_size;
run;

/*correlation of residuals*/

proc corr;var ms_res hs_res;run;
```

## Appendix B

### SAS Program for 3 Feeder Schools

```
/*Generating middle school and high school residuals based on a known
correlation*/

%macro generate;

proc iml;
  resid=normal(j(50,2,0));/*change this to 30 in other condition*/
  corr={1.00 .40,
        .40 1.00}; /*change to .00 to remove correlation*/
  corr_root=root(corr);
  fin_resid=resid*corr_root;
  fin_resid2=.4629#fin_resid; /*change to .0556 to change IUCC to .05*/

  create sasresid from fin_resid2;
  append from fin_resid2;

quit;

data sasresid;set sasresid;
  rename COL1= ms_res;
  rename COL2= hs_res;
run;

proc sort data=sasresid;by ms_res;
run;

/*main feeder school*/

data one;set sasresid;
  retain counter 0;
  counter=counter+1;
  ms_id=counter;
  hs_id=counter;
  ms_size= int(20+2*rannor(0)); /*change to 40 in other condition*/
  cell_size = int(.70*ms_size);
  size= 50+10*rannor(0);
  middleschool_effect= .50*size + ms_res;
  free=50+10*rannor(0);
  highschool_effect=.50*free + hs_res;
run;

/*secondary feeder school*/

data onemore;set one;
  keep ms_id hs_id cell_size;
  ms_id=counter;
  hs_id=counter+1;
```

```

cell_size=int (.15*ms_size);
if cell_size = 0 then cell_size = 1;
if hs_id gt 50 then hs_id=1; /* change to 30 in other condition*/
ms_res=ms_res;
run;

/*other secondary feeder school*/

data oneless;set one;
keep ms_id hs_id cell_size;
hs_id=counter;
ms_id=counter+1;
cell_size=int (.15*ms_size);
if cell_size=0 then cell_size=1;
if ms_id gt 50 then ms_id=1; /*change to 30 in other condition*/
run;

proc sort data=onemore;by ms_id;run;
data merged;
merge one(in=A) onemore (in=B);by ms_id;
if A and B;
keep ms_id hs_id cell_size ms_res size middleschool_effect;
run;

proc sort data=onemore;by hs_id;run;
data merged1;
merge one(in=A) onemore(in=B);by hs_id;
if A and B;
keep ms_id hs_id hs_res free highschool_effect;
run;

proc sort data=merged;by hs_id;run;
data merged2;
merge merged(in=A) merged1(in=B);by hs_id;
if A and B;
keep ms_id hs_id cell_size ms_res hs_res size free middleschool_effect
highschool_effect;
run;

data merged3;
merge one(in=A) oneless(in=B);by hs_id;
if A and B;
keep ms_id hs_id hs_res free highschool_effect;
run;

proc sort data=oneless;by ms_id;run;
data merged4;
merge one(in=A) oneless (in=B);by ms_id;
if A and B;
keep ms_id hs_id cell_size ms_res size middleschool_effect;
run;

proc sort data =merged4;by hs_id;run;
data merged5;
merge merged3 (in=A) merged4 (in=B); by hs_id;

```

```

if A and B;
keep ms_id hs_id cell_size ms_res hs_res size free middleschool_effect
highschool_effect;
run;

data grand;
keep ms_res hs_res ms_id hs_id cell_size size free middleschool_effect
highschool_effect;
set one merged2 merged5;
run;

/*creating student observations*/

data student;set grand;
do student= 1 to cell_size;
SAT=50+10* rannor(0);
student_effect=.50*SAT + 1*rannor(0);
test_score=100+student_effect+middleschool_effect+highschool_effect;
iter=&i;
output;
end;
run;
%mend;

%macro runall;
%do i =1 %to 50;
%generate;
proc datasets;
append base=fulldata data=student;
%end;
%mend;

proc iml;
%runall;
quit;

ods output covparms=random_effects;
ods output solutionf = fixed_effects;
ods output FitStatistics = fit_statistics;

/* Fitting the cross-classified model */

proc sort data=fulldata;by iter;run;
proc mixed data= fulldata noclprint covtest noinfo noitprint;
class ms_id hs_id;
model test_score= free size SAT/s;
random int/sub= ms_id;
random int/sub=hs_id;
by iter;
run;

/*transposing random effects*/

data random_effects;set random_effects;by iter;

```

```

if Subject = ' ' then Subject = 'resid';
run;

proc transpose data=random_effects out=random_trans;
var estimate stderr zvalue probz;
id subject;
by iter;
run;

data randeffs;set random_trans;by iter;
retain msest hsest resest msstd hsstd resstd msZ hsZ resZ msProbZ hsProbZ
resProbZ iter;

if _NAME_ = 'Estimate' then do;
msest=ms_id;
hsest=hs_id;
resest=resid;
end;
if _NAME_='StdErr' then do;
msstd=ms_id;
hsstd=hs_id;
resstd=resid;
end;
if _NAME_='ZValue' then do;
msZ=ms_id;
hsZ=hs_id;
resZ=resid;
end;
if _NAME_='ProbZ' then do;
msProbZ=ms_id;
hsProbZ=hs_id;
resProbZ=resid;
end;
keep msest hsest resest msstd hsstd resstd msZ hsZ resZ msProbZ hsProbZ
resProbZ iter;
if last.iter;
run;

proc transpose data = fixed_effects out = fixed_trans;
var estimate StdErr tvalue Probt;
id effect;
by iter;
run;

/*renaming fixed effects*/

data fixedeff;set fixed_trans; by iter;
retain intercept_est free_est SAT_est size_est intercept_std free_std
SAT_std size_std intercept_t free_t SAT_t size_t intercept_probt
free_probt SAT_probt size_probt iter;

if _NAME_ = 'Estimate' then do;

```

```

intercept_est = intercept;
free_est = free;
SAT_est = SAT;
size_est=size;
end;
if _NAME_='StdErr' then do;
intercept_std = intercept;
free_std = free;
SAT_std = SAT;
size_std=size;
end;
if _NAME_='tValue' then do;
intercept_t = intercept;
free_t = free;
SAT_t = SAT;
size_t=size;
end;
if _NAME_='Probt' then do;
intercept_probt = intercept;
free_probt = free;
SAT_probt = SAT;
size_probt=size;
end;
keep intercept_est free_est SAT_est size_est intercept_std free_std
SAT_std size_std intercept_t free_t SAT_t size_t intercept_probt
free_probt SAT_probt size_probt iter;
if last.iter;
run;

/*Getting Fit indices in tabular form*/

proc transpose data = fit_statistics out = fit_trans;
var value;
id Descr;
by iter;
run;

/*Renaming Fit Statistics*/

data fit_indices;set fit_trans;by iter;
_2RLL = _2_Res_Log_Likelihood;
AIC = AIC__smaller_is_better_;
AICC= AICC__smaller_is_better_;
BIC = BIC__smaller_is_better_;
keep iter AIC BIC;
run;

data results;
merge randeffs
      fixedeff
      fit_indices;
by iter;
run;

```

```

/*Now, repeating the above for the standard HLM */

ods output covparms = random_effects2;
ods output solutionf = fixed_effects2;
ods output FitStatistics = fit_statistics2;

/*fitting the standard HLM model with HS as the clustering unit*/

proc mixed data = fulldata noclprint covtest noinfo noitprint;
class hs_id;
model test_score = size free SAT/s;
random int/sub = hs_id;
by iter;
run;

/*Getting random effects in tabular form */

data random_effects2;set random_effects2;by iter;
if Subject = ' ' then Subject = 'resid';
run;
proc transpose data = random_effects2 out = random_trans2;
var estimate StdErr ZValue ProbZ;
id subject;

by iter;
run;

/*renaming random effects*/

data randeff2;set random_trans2; by iter;
retain hsest2 resest2  hsstd2 resstd2  hsZ2 resZ2  hsProbZ2 resProbZ2
iter;

if _NAME_ = 'Estimate' then do;
hsest2=hs_id;
resest2=resid;
end;
if _NAME_='StdErr' then do;
hsstd2=hs_id;
resstd2=resid;
end;
if _NAME_='ZValue' then do;
hsZ2=hs_id;
resZ2=resid;
end;
if _NAME_='ProbZ' then do;
hsProbZ2=hs_id;
resProbZ2=resid;
end;
keep hsest2 resest2  hsstd2 resstd2  hsZ2 resZ2  hsProbZ2 resProbZ2 iter;
if last.iter;
run;

```



```

/* Getting fixed effects in tabular form*/

proc transpose data = fixed_effects2 out = fixed_trans2;
var estimate StdErr tvalue Probt;
id effect;
by iter;
run;

data fixedeff2;set fixed_trans2; by iter;
retain intercept_est2 free_est2 SAT_est2 size_est2 intercept_std2
free_std2 SAT_std2 size_std2 intercept_t2 free_t2 SAT_t2 size_t2
intercept_probt2 free_probt2 SAT_probt2 size_probt2 iter;

if _NAME_ = 'Estimate' then do;
intercept_est2 = intercept;
free_est2 = free;
SAT_est2 = SAT;
size_est2=size;
end;
if _NAME_='StdErr' then do;
intercept_std2 = intercept;
free_std2 = free;
SAT_std2 = SAT;
size_std2=size;
end;
if _NAME_='tValue' then do;
intercept_t2 = intercept;
free_t2 = free;
SAT_t2 = SAT;
size_t2=size;
end;
if _NAME_='Probt' then do;
intercept_probt2 = intercept;
free_probt2 = free;
SAT_probt2 = SAT;
size_probt2=size;
end;
keep intercept_est2 free_est2 SAT_est2 size_est2 intercept_std2 free_std2
SAT_std2 size_std2 intercept_t2 free_t2 SAT_t2 size_t2 intercept_probt2
free_probt2 SAT_probt2 size_probt2 iter;
if last.iter;
run;

/*Getting Fit indices in tabular form*/

proc transpose data = fit_statistics2 out = fit_trans2;
var value;
id Descr;
by iter;
run;

```

```

/*Renaming Fit Statistics*/

data fit_indices2;set fit_trans2;
  _2RLL2 = _2_Res_Log_Likelihood;
  AIC2 = AIC__smaller_is_better_;
  AICC2= AICC__smaller_is_better_;
  BIC2 = BIC__smaller_is_better_;
keep iter AIC2 BIC2;
run;

/*Merges fixed effects, random effects, and statistical tests */

data results2;
merge randeff2
      fixedeff2
      fit_indices2;
by iter;
run;

/*Summarizes data across iterations*/

data final;
merge results
      results2;
by iter;
run;

/*Descriptive Statistics for all of the estimates*/

proc means data=final;run;

/*Proportion of time CCREM is chosen over HLM*/

data fit;set final;
by iter;
AIC_fit = AIC2-AIC;
BIC_fit=BIC2-BIC;
if AIC_fit ge 0 then Afit = 1;
else Afit = 0;
if BIC_fit ge 0 then Bfit = 1;
else Bfit = 0;
run;

proc freq;
tables Afit Bfit;
run;

data fixed_bias;set final;
by iter;
keep msest hsest resest intercept_est free_est SAT_est size_est hsest2
resest2 intercept_est2 free_est2 SAT_est2 size_est2;
run;

```

```

/*Calculating Relative Bias of the Parameter Estimates*/

proc means data=fixed_bias;
output out=fixed_means mean(msest hsest resest intercept_est free_est
SAT_est size_est hsest2 resest2 intercept_est2 free_est2 SAT_est2
size_est2)=mean_msest mean_hsest mean_resest mean_intercept_est
mean_free_est mean_SAT_est mean_size_est mean_hsest2 mean_resest2
mean_intercept_est2 mean_free_est2 mean_SAT_est2 mean_size_est2;
run;

data parameter_bias;set fixed_means;
msest_bias=(mean_msest -.2143)/.2143; /*changes when IUCC changes*/
hsest_bias=(mean_hsest -.2143)/.2143; /*changes when IUCC changes*/
resest_bias=(mean_resest - 1)/1;
int_bias=(mean_intercept_est-100)/100;
free_bias=(mean_free_est-.5)/.5;
SAT_bias=(mean_SAT_est-.5)/.5;
size_bias=(mean_size_est-.5)/.5;
hsest_bias2=(mean_hsest2 -.2143)/.2143; /*changes when IUCC changes*/
resest_bias2=(mean_resest2 - 1)/1;
int_bias2=(mean_intercept_est2-100)/100;
free_bias2=(mean_free_est2-.5)/.5;
SAT_bias2=(mean_SAT_est2-.5)/.5;
size_bias2=(mean_size_est2-.5)/.5;
run;

data parameter_bias;set parameter_bias;
keep msest_bias hsest_bias resest_bias int_bias free_bias SAT_bias
size_bias hsest_bias2 resest_bias2 int_bias2 free_bias2 SAT_bias2
size_bias2;
run;

proc print data=parameter_bias;run;

/*Calculating Relative Bias of the Standard Errors*/

data se_bias;set final;
by iter;
keep msest hsest resest intercept_est free_est SAT_est size_est hsest2
resest2 intercept_est2 free_est2 SAT_est2 size_est2 msstd hsstd hsstd2
resstd resstd2 intercept_std intercept_std2 free_std free_std2 SAT_std
SAT_std2 size_std size_std2;
run;

proc means data=se_bias;
output out=se_means mean (msest hsest resest intercept_est free_est
SAT_est size_est hsest2 resest2 intercept_est2 free_est2 SAT_est2
size_est2 msstd hsstd hsstd2 resstd resstd2 intercept_std intercept_std2
free_std free_std2 SAT_std SAT_std2 size_std size_std2)=mean_msest
mean_hsest mean_resest mean_intercept_est mean_free_est mean_SAT_est
mean_size_est mean_hsest2 mean_resest2 mean_intercept_est2 mean_free_est2
mean_SAT_est2 mean_size_est2 mean_msstd mean_hsstd mean_hsstd2
mean_resstd mean_resstd2 mean_intercept_std mean_intercept_std2

```

```

mean_free_std mean_free_std2 mean_SAT_std mean_SAT_std2 mean_size_std
mean_size_std2 stddev (msest hsest resest intercept_est free_est SAT_est
size_est hsest2 resest2 intercept_est2 free_est2 SAT_est2 size_est2 msstd
hsstd hsstd2 resstd resstd2 intercept_std intercept_std2 free_std
free_std2 SAT_std SAT_std2 size_std size_std2)=std_msest std_hsest
std_resest std_intercept_est std_free_est std_SAT_est std_size_est
std_hsest2 std_resest2 std_intercept_est2 std_free_est2 std_SAT_est2
std_size_est2 std_msstd std_hsstd std_hsstd2 std_resstd std_resstd2
std_intercept_std std_intercept_std2 std_free_std std_free_std2
std_SAT_std std_SAT_std2 std_size_std std_size_std2;
run;

```

```

data standard_error_bias;set se_means;
ms_se_bias=(mean_msstd- std_msest)/std_msest;
hs_se_bias=(mean_hsstd - std_hsest)/std_hsest;
residual_se_bias=(mean_resstd-std_resest)/std_resest;
intercept_se_bias=(mean_intercept_std -
std_intercept_est)/std_intercept_est;
free_se_bias=(mean_free_std-std_free_est)/std_free_est;
SAT_se_bias=(mean_SAT_std- std_SAT_est)/std_SAT_est;
size_se_bias=(mean_size_std-std_size_est)/std_size_est;
hs_se_bias2=(mean_hsstd2 - std_hsest2)/std_hsest2;
residual_se_bias2=(mean_resstd2-std_resest2)/std_resest2;
intercept_se_bias2=(mean_intercept_std2 -
std_intercept_est2)/std_intercept_est2;
free_se_bias2=(mean_free_std2-std_free_est2)/std_free_est2;
SAT_se_bias2=(mean_SAT_std2- std_SAT_est2)/std_SAT_est2;
size_se_bias2=(mean_size_std2-std_size_est2)/std_size_est2;
run;

```

```

data standard_error_bias;set standard_error_bias;
keep ms_se_bias hs_se_bias residual_se_bias intercept_se_bias
free_se_bias SAT_se_bias size_se_bias hs_se_bias2 residual_se_bias2
intercept_se_bias2 free_se_bias2 SAT_se_bias2 size_se_bias2;
run;
proc print data=standard_error_bias;
run;

```

```

data fullldata;set fullldata;
proc means;var ms_res hs_res middleschool_effect highschool_effect
cell_size;
run;

```

```

/*correlation of the residuals*/
proc corr;var ms_res hs_res;run;

```

## References

- Ainsworth, J.W. (2002). Why Does It Take a Village? The Mediation of Neighborhood Effects on Educational Achievement. *Social Forces*, 81(1), 117-152.
- Browne, W. J., Goldstein, H., & Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modeling*, 1, 103-124.
- Clayton, D. & Rasbash, J. (1999). Estimation in Large Random Effects Models by Data Augmentation. *Journal of the Royal Statistics Society, A*, 162(3): 425-436.
- Fielding, A. (2002). Teaching groups as foci for evaluating performance in cost-effectiveness of GCE Advanced Level provision: Some practical methodological innovations. *School Effectiveness and School Improvement*, 13, 225-246.
- Garner, C.L & Raudenbush, S.W (1991). Neighborhood Effects on Educational Attainment: A Multilevel Analysis. *Sociology of Education*, 64 (4), 251-262.
- Goldstein, H. (1995). *Multilevel Statistical Models* (2<sup>nd</sup> edition). Chapter 8. New York: John Wiley.

- Goldstein, H. & Sammons, P (1997). The Influence of Secondary and Junior Schools on Sixteen Year Examination Performance: A Cross-Classified Multilevel Analysis. *School Effectiveness and School Improvement*, 8(2), 219-230.
- Goldstein, H., Rasbash, J., Plewis, I., Draper, D., et al. (1998). *A user's guide to MLwiN*. London, Institute of Education.
- Green, S.B., Salkind, N.J., & Akey, T.M. (2000). Using SPSS for Windows: Analyzing and Understanding Data.(2<sup>nd</sup> edition). New Jersey: Prentice Hall.
- Hill, P.W. & Goldstein, H.(1998). Multilevel Modelling of educational data with cross classification and missing identification of units. *Journal of Educational and Behavioral Statistics*, 23, 117-128.
- Hoogland, J.J. & Boomsma, A. (1998). Robustness Studies in Covariance Structure Modeling. *Sociological Methods and Research*, 26(3), 329-367.
- Hox, J. J. (2002). Multilevel analysis : techniques and applications. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kalton, G. (1983). Introduction to Survey Sampling. Newbury Park, CA: Sage Publications.
- Kreft, I. & de Leeuw, J. (1998). *Introducing Multilevel Modeling*. Thousand Oaks, CA: Sage Publications, Inc.

- Lillard, L. & Panis, C. (2000). *AML* [Computer Program] Los Angeles:  
EconWare.
- Ma, Xin & Wilkins, J.L.M (2002). The Development of Science Achievement in  
Middle and High School- Individual Differences and School Effects.  
*Evaluation Review*, 26(4), 395-417.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).
- O’Muircheartaigh, C. & Campanelli, P. (1999). A multilevel exploration of the  
role of interviewers in survey non-response. *Journal of the Royal  
Statistics Society*, 162(3), 437-446.
- Raftery, A.E. & Zheng, Y. (2003). *Long Run Performance of Bayesian Model  
Averaging*. Technical Report no. 433, Department of Statistics, University  
of Washington.
- Rasbash, J. & Browne, W. J. (2001). Modelling Non-Hierarchical Structures. In  
Leyland, A. H. and Goldstein, H. (Eds.), *Multilevel Modelling of Health  
Statistics*, 93-105. Chichester: John Wiley & Sons.
- Raudenbush, S. (1993). A crossed random effects model for unbalanced data with  
applications in cross-sectional and longitudinal research. *Journal of  
Educational Statistics*, 18(4), 321-349.

- Raudenbush, S. & Bryk, A.S.(2002). Hierarchical linear models: applications and data analysis methods (2<sup>nd</sup> edition). Chapter 12: Models for Cross-Classified Random Effects. Thousand Oaks.
- Raudenbush, S., Bryk, A., Cheong, Y.F., & Congdon, R. (2000). *HLM 5* [Computer Program]. Chicago: Scientific Software International.
- Roberts, J.Kyle (2000). *The pitfalls of ignoring multilevel design in national datasets*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Bowling Green, KY.
- Simonite, V. & Browne, W.J. (2003). Estimation of a large cross-classified multilevel model to study academic achievement in a modular degree course. *Journal of Royal Statistics Society, 166(1), 119-133*.
- Snijders, T. & Bosker, R. (1999). *Multilevel Analysis*. Thousand Oaks, CA: Sage Publications, Inc.
- Teitler, J.O. & Weiss, C.C. (2000). Effects of Neighborhood and School Environments of Transitions to First Sexual Intercourse. *Sociology of Education, 73(2), 112-132*.
- Yang, M. (2003). A review of random effects modeling in SAS (release 8.2). Retrieved September 5, 2003 from <http://multilevel.ioe.ac.uk/softrev/reviewsas.pdf>.



## **Vita**

Jason Leon Meyers was born in Austin, Texas on December 10, 1975, the son of Charles Lawrence Meyers and Purnie Lefkofsky Meyers. After completing his studies at A.N. McCallum High School in Austin, he entered Emory University in Atlanta, Georgia. In May, 1998, he earned the degree of Bachelor of Arts with a major in Psychology. After working for several years, he entered the graduate program in Educational Psychology at the University of Texas in January of 2001. During his years in graduate school he worked as a Teaching Assistant in the Department of Educational Psychology and as a Graduate Research Assistant at the Measurement and Evaluation Center. He also published several manuscripts. Following completion of his doctoral degree, he plans to pursue employment in the educational testing industry.

Permanent address: 6800 Austin Center Blvd. Apt. 841  
Austin, Texas 78731

This dissertation was typed by the author.