Georgia State University
ScholarWorks @ Georgia State University

Educational Policy Studies Dissertations | Department of Educational Policy Studies

8-12-2009

A Monte Carlo Study: The Impact of Missing Data in Cross-Classification Random Effects Models

Meltem Alemdar

Follow this and additional works at: http://scholarworks.gsu.edu/eps_diss

Recommended Citation

Alemdar, Meltem, "A Monte Carlo Study: The Impact of Missing Data in Cross-Classification Random Effects Models." Dissertation, Georgia State University, 2009.
http://scholarworks.gsu.edu/eps_diss/34

This Dissertation is brought to you for free and open access by the Department of Educational Policy Studies at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Educational Policy Studies Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

ACCEPTANCE

This dissertation, A MONTE CARLO STUDY: THE IMPACT OF MISSING DATA IN CROSS-CLASSIFICATION RANDOM EFFECTS MODELS, by MELTEM ALEMDAR, was prepared under the direction of the candidate's Dissertation Advisory Committee. It is accepted by the committee members in partial fulfillment of the requirements for the degree Doctor of Philosophy in the College of Education, Georgia State University.

The Dissertation Advisory Committee and the student's Department Chair, as representatives of the faculty, certify that this dissertation has met all standards of excellence and scholarship as determined by the faculty. The Dean of the College of Education concurs.

____
Carolyn F. Furlow, Ph.D.
Committee Chair

____
Philo A. Hutcheson, Ph.D.
Committee Member

____
Phillip E. Gagne, Ph.D.
Committee Member

____
Sheryl A. Gowen, Ph.D.
Committee Member

____
Date

____
Sheryl A. Gowen, Ph.D.
Chair, Department of Educational Policy Studies

____
R. W. Kamphaus, Ph.D.
Dean and Distinguished Research Professor
College of Education

AUTHOR'S STATEMENT

By presenting this dissertation as a partial fulfillment of the requirements for the advanced degree from Georgia State University, I agree that the library of Georgia State University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote, to copy from, or to publish this dissertation may be granted by the professor under whose direction it was written, by the College of Education's director of graduate studies and research, or by me. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without my written permission.

____

Meltem Alemdar

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University library must be used in accordance with the stipulations prescribed by the author in the preceding statement. The author if this dissertation is:

Meltem Alemdar
1140 North Ave NE Apt: 7
Atlanta, GA 30307

The director of this dissertation is:

Dr. Carolyn F. Furlow
Department of Educational Policy Studies
College of Education
Georgia State University
Atlanta, GA 30303- 3083

VITA

Meltem Alemdar

ADDRESS: 1140 North Ave NE Apt: 7
Atlanta, Georgia 30307

EDUCATION:

Ph.D.	2008	Georgia State University Educational Policy Studies
M.S.	2003	Georgia State University Educational Research
B.S.	1999	University of Hacettepe Testing, Measurement and Evaluation in Education

PROFESSIONAL EXPERIENCE:

2002-Present	Graduate Research Assistant Georgia State University, Atlanta, GA
2003-2008	Program Evaluation Research Associate Georgia State University, Atlanta, GA
2005-2006	Graduate Teaching Assistant Georgia State University, Atlanta, GA
2002-2004	Senior Graduate Student Researcher Georgia Professional Standards Commission, Atlanta, GA

SELECTED PRESENTATIONS

Lingle, J., Alemdar, M., & Gowen, S. (April 2009). Comparing Methods of Propensity Score Estimation with Partially Missing Data. Presentation at the Annual Conference for the American Educational Research Association, San Diego, CA.

Lingle, J., Alemdar, M., Gowen, S., & Skelton, S. (March 2008). School Engagement, After School Activities, and Health-Risk Behaviors: Results from an Evaluation of Community-Based After School Programs. Presentation at the Annual Conference for the American Educational Research Association, New York.

Alemdar, M. & Furlow, C. F. (2007). The impact of missing data in cross-classified models**.** Presentation at the Sixth International Amsterdam Conference on Multilevel Analysis, Amsterdam, Netherlands.

Alemdar, M. & Furlow, C. F. & Gowen, S. A. (2005). Investigation of Missing Data Patterns in Georgia Academic Learn & Serve Program. Presented at 5th Annual International Learn & Serve Conference. East Lansing, MI.

Alemdar, M. & Furlow, C. F. & Gowen, S. A. (2005). Evaluation of the Academic Learn and Serve Program Using HLM. Presentation at the American Psychology Association Annual Convention. Washington DC.

Alemdar, M. (2005). John Dewey's Impact on Today's Turkish Education System. Presentation at the Southeast Philosophy of Education Society Conference. Orlando, FL.

Alemdar, M. & Livingston, C. F. & Gowen, S. A.(2004). Evaluation challenges faced when assessing a service-learning program. Presentation at the 2004 International Service-Learning Research Conference. Greenville, SC.

ABSTRACT

A MONTE CARLO STUDY: THE IMPACT OF MISSING DATA IN CROSS-CLASSIFICATION RANDOM EFFECTS MODELS
by
Meltem Alemdar

Unlike multilevel data with a purely nested structure, data that are cross-classified not only may be clustered into hierarchically ordered units but also may belong to more than one unit at a given level of a hierarchy. In a cross-classified design, students at a given school might be from several different neighborhoods and one neighborhood might have students who attend a number of different schools. In this type of scenario, schools and neighborhoods are considered to be cross-classified factors, and cross-classified random effects modeling (CCREM) should be used to analyze these data appropriately. A common problem in any type of multilevel analysis is the presence of missing data at any given level. There has been little research conducted in the multilevel literature about the impact of missing data, and none in the area of cross-classified models. The purpose of this study was to examine the effect of data that are missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR), on CCREM estimates while exploring multiple imputation to handle the missing data. In addition, this study examined the impact of including an auxiliary variable that is correlated with the variable with missingness (the level-1 predictor) in the imputation model for multiple imputation. This study expanded on the CCREM Monte Carlo simulation work of Meyers (2004) by the inclusion of studying the effect of missing data and method for

handling these missing data with CCREM. The results demonstrated that in general, multiple imputation met Hoogland and Boomsma's (1998) relative bias estimation criteria (less than 5% in magnitude) for parameter estimates under different types of missing data patterns. For the standard error estimates, substantial relative bias (defined by Hoogland and Boomsma as greater than 10%) was found in some conditions. When multiple imputation was used to handle the missing data then substantial bias was found in the standard errors in most cells where data were MNAR. This bias increased as a function of the percentage of missing data.

A MONTE CARLO STUDY: THE IMPACT OF MISSING DATA IN CROSS-CLASSIFICATION RANDOM EFFECTS MODELS

by
Meltem Alemdar

A Dissertation

Presented in Partial Fulfillment of Requirements of the
Degree of
Doctor of Philosophy
in
Educational Policy Studies
in
the Department of Educational Policy Studies
in
the College of Education
Georgia State University

Atlanta, GA
2008

Copyright by
Meltem Alemdar
2008

DEDICATION

This dissertation is dedicated to my brother, Bulent, who is my mentor and inspiration, and also to my family: Annem Gulizar, Babam Fuat, Kizkardeslerim Meryem ve Biriz

I hope I make you proud.

ACKNOWLEDGMENTS

I am extremely grateful to so many people for supporting me, day in and day out, during this long journey. I have learned so much, not just through the preparation of this dissertation but also through the many thoughtful discussions and talks with friends and faculty that I have become a completely different person. This dissertation is therefore not just a simulation study, but it is also a record of the journey I have made in becoming this new person.

I must first express my gratitude towards my advisor, Dr. Carolyn Furlow, who has been an extraordinary advisor by providing intellectual challenge and friendship throughout the preparation of this dissertation. I will be always in debt to her for believing in me and holding my hand through the rough times of this process. Very special thanks must go to Dr. Phillip Gagne, a passionate teacher and scholar who assisted me during the analysis of my dissertation, who spent long hours reviewing my dissertation and making sure I did not give up. I extend my gratitude and admiration also to Dr. Philo Hutcheson for his words of wisdom and support as I learned from his expertise and breadth of knowledge. I would also like to give special thanks to Dr. Sheryl Gowen for providing me financial support, giving me the opportunity to work on many interesting research projects, and being my mentor and trusting me throughout this journey.

This dissertation is dedicated to my brother and other immediate family members for many reasons. Without the support and encouragement of my brother, Bulent, I would not be at this point at all. I want to thank him not only for being my mentor and friend, but for taking the responsibility to ensure that I have always stayed on the right track and for reminding me to stay happy. I also want to thank my sister in-law Biriz. Her strength and kindness has motivated me in times of struggle and I appreciate her more than words could ever express. I am also so fortunate to have the undying love and support of my family in Turkey: my mom Gulizar, my dad Fuat, and my sister Meryem. They have always supported me in my efforts to succeed, even when my success meant that they had to make sacrifices and live apart from me for many years. Although my parents did not really understand what I was doing in attending school for so long, sometimes asking, ”Now tell me again why you’re still going to school?,” they always showed me unconditional love, and for that I am grateful to them forever. *Hersey icin sonsuz tesekkurler*.

I have been incredibly lucky with friends who have shown me so much love and understanding. Special thanks also go to my dearest friends Sinem, Nevbahar, and Funda. They are all constant reminders of what life is all about - good people, great laughs, honesty, and love. Other special thanks also go to my best friend Burcu, with whom I have been friends since middle school. She supported me anyway she could during my time here. I also express my appreciation to Muge who has been a great source of strength and happiness for me. And to Hakan who helped me dearly to start this journey and supported in many ways that I never had a chance to show my appreciation to him.

I thank to my roommate and best friend Elcin who became my family in the United States. There are no words to express my gratitude to her. Thank you for just being you and making me your sister.

I would also like to thank my colleague and best friend, Jeremy. I can’t imagine what this journey would have been like had it not been for his friendship. I can’t thank him enough for making me laugh and preparing me for yet another day. I will miss our long coffee hours, lunch breaks, and dissertation discussions, but I know there are many other good experiences yet to come.

I also would like to express my gratitude to my dear friend Jennie for keeping my mind away from my dissertation and providing me her unconditional friendship; to Evren for being there for me whenever I needed; to Gonul and Sinasi for keeping me company and being my cheerleaders; to Atay, Bulent and Eray for being a long time supporters and great friends; to Rebecca, Dawson, Amy and Mike for helping me to keep my mind and soul happy, and providing me good food; to Jeanne and Sue for nurturing me to inconceivable levels; to Syreeta for making sure I was hanging in there; and to the EPS staff for making my life easy during this journey.

And to those who witnessed part of my journey in Atlanta then moved thousand miles away: Even when I did not return phone calls, emails, or text messages, they still loved me and cared about me. I am grateful to Pinar, Volkan, Elif, Deniz, Timucin and Orkun for changing my life in some ways that I would never imagine.

Last, but not least, to my significant other, Brian Kammer, the surprise and still incomplete part of this journey: Thank you for your love, support and faith in me, without you I would never have gotten through the last couple months of this dissertation process. I am thankful for our past, present and future together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

CCREM	Cross-classified Random Effects Modeling
GPA	Grade Point Average
HLM	Hierarchical Linear Modeling
ICC	Intraclass Correlation Coefficient
IUCC	Intra Unit Correlation Coefficient
LD	Listwise Deletion
MAR	Missing at Random
MCAR	Missing Completely at Random
MCMC	Bayesian Markov Chain Monte Carlo
MNAR	Missing Not at Random
MI	Multiple Imputation
RLR	Reduced Lunch Rate
SES	Socio-economic Status

CHAPTER 1

INTRODUCTION

The present study investigated the impact of missing data with cross-classification random effects modeling (CCREM), where individuals belong to more than one unit at a given level. Unlike multilevel data with a purely nested structure, data that are cross-classified may not only be clustered into hierarchically ordered units; they may also belong to more than one unit at a given level of a hierarchy. For example, in a purely nested design, students who belong to one neighborhood would all attend the same group of schools and these schools would have only one neighborhood where students belong. In a cross-classified design, however, students at a given school might be from several different neighborhoods and one neighborhood might have students who attend a number of different schools. In this type of scenario, schools and neighborhoods are considered to be cross-classified factors and CCREM should be used to appropriately analyze these data.

Meyers and Beretvas (2006) investigated the impact of inappropriate modeling of cross-classified data by analyzing both real and simulated data. In the simulation study, their major finding was that when the cross-classified structure was ignored, the fixed effect estimates were unbiased; however, the standard errors associated with the incorrectly modeled variables were underestimated. Such underestimation would result in an inflation of the Type I error rate.

A common problem in any type of multilevel analysis is the presence of missing data at any given level. There has been little research conducted in the multilevel literature about the impact of missing data, and no research to date has investigated the impact of missing data in the area of cross-classified models. Using a search of applied articles from PsycInfo database, researchers conducting CCREM primarily used the typical default for handling missing data, listwise deletion (LD). None of these studies used more advanced techniques, such as multiple imputation (MI), for handling missing data. While LD can be correctly implemented when data are missing completely at random (MCAR), it is not robust to data that are missing at random (MAR; Shafer & Graham, 2002). Using LD can result in biased results if the data are MAR or missing not at random (MNAR). MI, on the other hand, can be accurately used if the data are at least MAR and if data are MNAR then use of an auxiliary variable can improve its performance (Collins, Schafer, & Kam, 2001).

The purpose of this study was to examine the impact of MCAR, MAR, and MNAR data on CCREM fixed effect estimates while exploring MI to handle the missing data in a Monte Carlo simulation study. This research expanded on the CCREM work of Meyers and Beretvas (2006), where they compared both real and simulated data by examining the impact of inappropriate modeling of cross-classified data, by the inclusion of studying the impact of missing data and using MI for handling these missing data with CCREM.

CHAPTER 2

LITERATURE REVIEW

This chapter begins with an introduction to hierarchical linear modeling (HLM), a brief history of its use, and its application in the area of educational research. Next, an explanation of CCREM models is given along with an example of cross-classified data in an educational setting. Finally, a discussion of missing data in CCREM is given, including missing data patterns and methods for handling missing data.

Hierarchical Linear Modeling

Hierarchical linear models (HLM) investigates the impact of both individual and grouping factors on some individual level outcome through incorporating data from multiple levels. For example, student achievement may be a function of student level characteristics (e.g., IQ and study habits), classroom level factors (e.g., instruction style and textbook), school level factors (e.g., Adequate Yearly Progress), and so on. In educational settings, students are nested within a classroom and classrooms nested within a school. When nesting occurs, the relationship between predictors and the dependent variable can be extended to more than one level. For example, student achievement can be predicted not only by student-level variables but also by school-level variables.

Most traditional statistical techniques are not capable of taking into account data with a hierarchical structure (Raudenbush & Bryk, 2002). Ordinary least squares (OLS) regression, which uses single-level procedures for modeling, disaggregates all higher order variables to the individual level. For example, in a scenario where schools are sampled and then students are sampled from those schools, the analysis is usually done where school characteristics are assigned to the individual level and variance in the dependent variable attributable to school characteristics is ignored. The problem with this approach is that OLS assumes that observations are independent of one another. Students in the same school, however, have the same value on each of the school variables. Shared experiences within schools can create dependent observations, which violate this independence of errors assumption. Thus, ignoring a hierarchical structure might violate this assumption, which results in biased standard errors and high Type I error rates (Raudenbush & Bryk, 2002).

Another traditional analysis approach is to aggregate the individual-level variables up to the next level. Thus, student characteristics are aggregated across schools. When the student characteristics are aggregated, however, within-group information is not taken into account. Therefore, a good amount of important information, which could explain the variation within groups, is lost. Because of the accompanying reduction in sample size, power to detect an effect also diminishes. As a result, both aggregating and disaggregating are not reasonable when the data structure is hierarchical (Raudenbush & Bryk, 2002).

Instead of ignoring variation in the dependent variable attributable to group characteristics, HLM can accurately incorporate group-level and individual-level predictors because it takes into account error structures at each level. HLM investigates the relationships within a single level and between or across hierarchical levels. For example, students within a particular school are likely to be more similar with each other than with students in other schools. In this case, HLM recognizes that students may not provide independent observations. HLM allows the partial interdependence of students within a school while it combines both student-level and school-level responses. In addition, while keeping constant the correct level of analysis for the independent variable, HLM estimates both lower and higher level variance in the dependent variable. This allows researchers to use individual independent variables at the individual level and group independent variables at the group level (Raudenbush & Bryk, 2002).

Bryk and Raudenbush (1988) pioneered much of what is currently known about HLM in their textbook. The authors described within-group and between-group equations along with demonstrations of HLM in educational settings. Because much of the research in education involves nested data structures, Raudenbush and Bryk (1988) demonstrated that HLM measures these data better than single-level methods.

Numerous authors have contributed to the development of HLM to address issues related to the hierarchical structure of data. Tate and Wongbundhit (1983) showed that single-level analysis is problematic in the presence of a hierarchical structure when they compared single and multilevel models. Browne, Goldstein, and Rasbash (2001) applied HLM to educational data, and Goldstein (1987) demonstrated the application of multilevel modeling in educational and social science research. In addition, Goldstein (1995) applied multilevel linear and nonlinear modeling approaches involving school effectiveness and cross-classified data. Kreft and de Leeuw (1998) provided examples using multilevel modeling in social and educational research, where they addressed practical issues and potential problems of doing multilevel analyses. Lastly, Hox (2002) discussed the extension of HLM models and special application areas. He emphasized understanding the methodological and statistical issues involved in using multilevel models.

In the literature, various terms have been used to describe HLM models. These include covariance component models (Goldstein, 1987; Longford, 1987), random-effects and mixed effects models (Laird & Ware, 1982; Singer, 1998), hierarchical linear modeling (Bryk & Raudenbush, 2002) and multilevel regression models (Hox, 2002). Researchers have also employed a variety of statistical programs, such as HLM 6 (Raudenbush, Bryk, & Congdon, 2005), SAS PROC MIXED (Institute, 2006), ML*wiN* (Rasbash, Browne, & Goldstein, 2000), MPLUS (Hagenaars & Van de Pol, 2002) and BMDP (Bentler, 1989) for data analysis and modeling purposes.

Two-level HLM Models

HLM can be described as the extension of a linear regression model, which adds random effects at multiple levels. Suppose that a researcher is interested in using grade point average (GPA) to predict scholastic aptitude test (SAT) scores. In this case, both variables are student level variables and simple regression could be used. However, in an education setting usually students (individual level) are nested within classes (group level), and classes are nested within schools. Therefore, using HLM allows the researcher to divide the variance in the outcome variable at the group and individual levels.

A typical two-level HLM model in an educational setting assigns students to level-1 and schools to level-2. The first estimated HLM is typically a fully unconditional model; no explanatory variables are included at either level. With this model, the amount of variance in the outcome variable that is attributable to within-group characteristics (here, students) and to between-group differences (schools) can be identified.

The researcher would start the analysis by running a fully unconditional model at level-1:

$$Y_{ij} = \beta_{0j} + r_{ij} \qquad (1)$$

where Y_{ij} is the dependent variable (SAT) for student *i* within school *j*, and β_{0j} (the intercept) is the mean SAT test score for school *j,* and r_{ij} is the level-1 residual by which student *i*'s score differs from school *j*'s mean test score. At level-2, the coefficients from level-1 are treated as outcome variables. The level-2 unconditional model is:

$$\beta_{oj} = \gamma_{00} + u_{0j} \qquad (2)$$

where β_{oj} is the outcome variable, and γ_{00} (intercept) is the grand mean of the school SAT score averages.

The residual term associated with level-1, r_{ij}, is the unique effect on the outcome variable corresponding to the *i*th student in the *j*th school; it is assumed to be normally and independently distributed with a mean of zero and constant variance, σ^2. The residual term associated with level-2, u_{0j}, is the unique effect of school *j* on the outcome variable; it is assumed to be normally and independently distributed with a mean of zero and constant variance, τ_{00}. The level-2 residual variance captures the variation across school means.

The level-1 and level-2 models can be combined into a single equation using substitution:

$$Y_{ij} = \gamma_{00} + r_{ij} + u_{oj} \qquad (3)$$

One important component in HLM is the intraclass correlation coefficient (ICC), which is a measure of the extent to which individuals are not independent of a grouping variable (here, schools). The ICC represents the proportion of the variance in the outcome variable that is between level-2 units (Raudenbush & Bryk, 2002).

Once the fully unconditional model is run and the proportion of between-group variation determined, then a conditional model can be examined where predictors are included at either one level or both levels. In this example, level-1 takes into account the student's GPA to explain the variability between individuals on the dependent variable (SAT). The level-1 equation follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + r_{ij} \tag{4}$$

where Y_{ij} is the outcome for the *i*th student in the *j*th school, β_{0j} is the predicted test score (the intercept) for the *j*th school when GPA equals to 0. When a predictor does not have a true zero value, the intercept can be problematic for interpretation; therefore; centering is sometimes necessary. β_{1j} is the expected change in test score associated with one point increase in GPA, and X_{1ij} is the GPA for individual *i* in classroom *j*, and r_{ij} is the level-1 residual.

At level-2, explanatory variables can be added to explain any variability between schools in the level-1 intercepts or slopes. The average number of years teaching experience for teachers, W_{ij}, within the schools can be used as a predictor at level-2. Intercepts β_{0j} and slopes β_{1j} can be modeled in the level-2 between-school model as follows:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} W_{1j} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} W_{1j} + u_{1j} \end{aligned} \tag{5}$$

where γ_{00} is the mean SAT score for a school where the average teacher's number of years teaching is equal to zero, W_{1j} represents the average teacher's years of experience for school *j*, γ_{01} is the amount of change in the dependent variable (SAT) with a one year increase in average number of years teaching, γ_{10} is the mean GPA-SAT slope where average teacher's number of years teaching equals zero, γ_{11} represents the amount of change in the GPA-SAT slope with one point increase on average years of teaching experience. u_{oj} is the random effect of school *j,* representing random variation in the average SAT among schools and u_{1j} is also random effect of school *j,* representing random variation on the SAT slope within schools. u_{oj} and u_{1j} are unique effects with means of zero and variances τ_{00} and τ_{11}, respectively. They represent the variability in β_{0j} and β_{1j} remaining after controlling for W_{1j}. The combined model can be shown as:

$$Y_{ij} = \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{10}X_{1ij} + \gamma_{11}W_{1j}X_{1ij} \left[+ u_{0j} + u_{1j}X_{1ij} + r_{ij} \right] \quad (6)$$

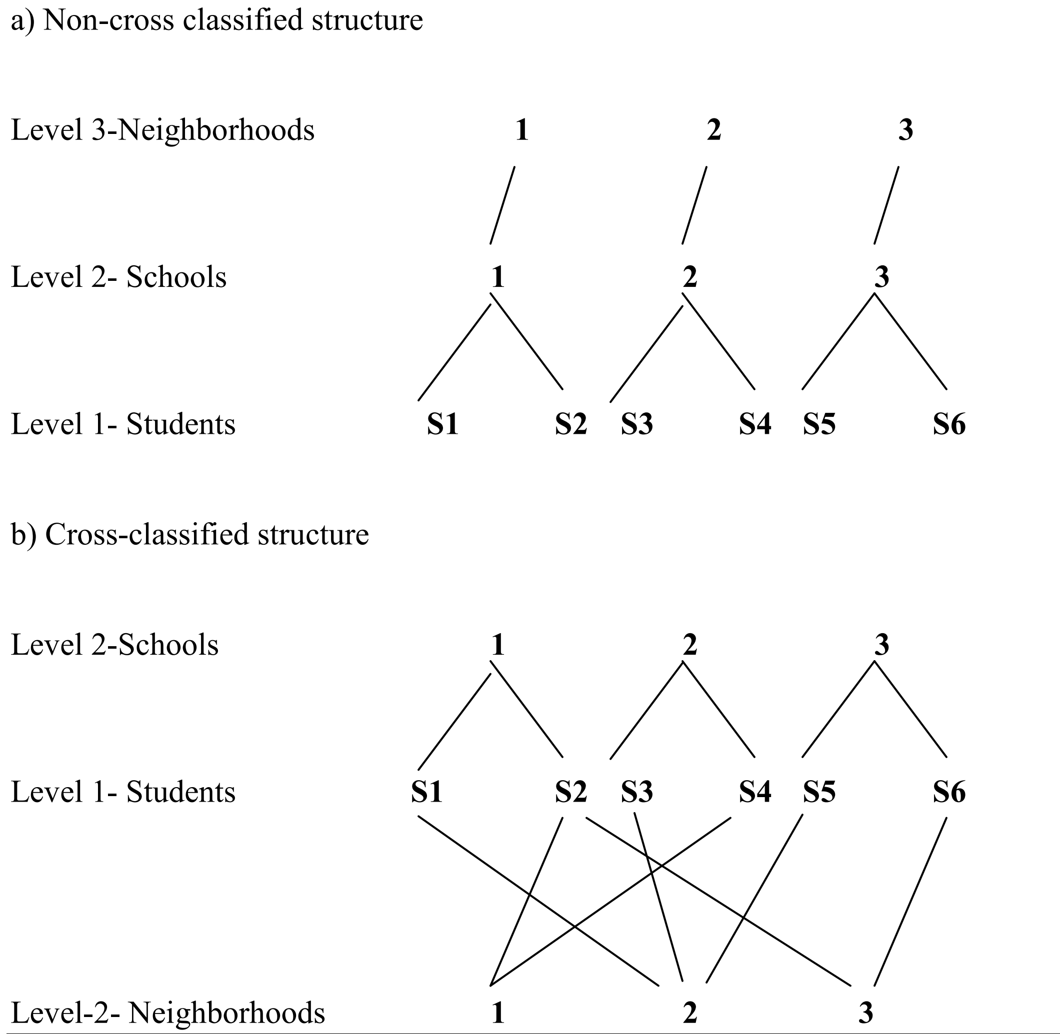
where the random effects are contained within the brackets.

Cross-Classified Data

In the previous section, the discussion pointed out a scenario where students are nested in schools in a two-level hierarchical structure. To extend the modeling further, the researcher could model the schools as nested within neighborhoods (a three level structure: students at level-1, schools at level-2, and neighborhoods at level-3). In a purely nested design, students who belong to one neighborhood would all attend the same group of schools and these schools would have only one neighborhood where students belong. In a cross-classified design, however, students at a given school might be from several different neighborhoods and one neighborhood might have students who attend a number of different schools. In this type of scenario, schools and neighborhoods are considered to be cross-classified factors. To appropriately analyze these data, Cross-classified random effects modeling (CCREM) should be used (Raudenbush and Bryk, 2002).

Cross-classification takes into account the influences from two different contexts that are not purely nested: in this example they are schools and neighborhoods. Compared with traditional HLM, CCREM can increase the precision of standard error estimates of random effects by considering clustered data at both of those levels (Meyers & Beretvas, 2006). Furthermore, inappropriate modeling of cross-classified data could be a problem if a researcher is interested in evaluating the effects of variables at the ignored level (i.e., if one of the cross-classified factors is not included in the analysis and only a standard two-level model is examined). In Figure 1 below, the first diagram illustrates a non-cross-classified structure. In this structure there are three levels: students, schools and neighborhoods. A typical hierarchical model has students nested in schools, and schools nested within neighborhoods. In a cross-classified structure, however, students in a given neighborhood might attend different schools, and one school might draw students from different neighborhoods. Subsequently the cross-classified data structure occurs at level-2.

Figure 1. Cross-Classified Structures.



Note: Six students at level-1 are nested within a neighborhood and school cross-classification at level-2. This figure was adapted from Goldstein and Fielding (2006).

In Figure 1, the difference between HLM and CCREM lies in the structure of the data. In the cross-classified structure, for example, students 1 and 2 attend the same school but come from different neighborhoods, whereas students 3 and 5 come from the same neighborhood but attend different schools.

In this example, researchers can take into account influences coming from two different factors: neighborhoods and schools. This may improve the estimation of explanatory variable effects ensuring the structure is properly specified (Goldstein & Fielding, 2006). In a cross-classified model, before introducing explanatory variables, the variation in outcomes could be examined by looking at the differences between schools, between neighborhoods, and between individual students after controlling for neighborhood and school effects. This will give a basis for then extending the model to identify which neighborhood, school, and student characteristics might explain some part of these component variances. After adding explanatory variables, the residual components of variance can be estimated which will provide more information about the variation in outcomes (Rasbash & Browne, 2001)

Cross-Classification Random Effects Modeling

The first and most well known study using CCREM was carried out by Raudenbush (1993), who reanalyzed data from a previous study he had conducted (Garner & Raudenbush, 1991). In the first study, the authors had investigated the impact of school effects for neighborhoods (living environment) on student achievement in one school district in Scotland. In the first analysis the data were considered as purely hierarchical: students were nested in neighborhoods. Raudenbush (1993) later reanalyzed the data taking into account the cross-classified data structure; schools and neighborhoods. In the study, there were 2310 children from 524 neighborhoods who attended 17 schools. The data were cross-classified because students from one neighborhood attended a host of schools and students from a particular school came from multiple neighborhoods. Some neighborhoods contributed as many as eight students, some contributed only one student, and some neighborhoods contributed no students. In the original study, Garner and Raudenbush (1991) ignored the variance between schools, instead focusing only on neighborhoods. In reanalyzing the data, Raudenbush (1993) estimated both the school and neighborhood variance using cross-classified modeling. With the cross-classified model, Raudenbush was able to examine the variation in student achievement resulting from neighborhoods, schools, and students after taking into account student, neighborhood, and school characteristics. In reanalyzing the data, the variance attributable to both schools and neighborhoods were estimated. The comparison of both studies' results indicated that neighborhood poverty had an impact on student achievement. However, part of the variability that had been incorrectly attributed to individual difference was now attributed to school effects such as teacher and classroom size effects.

Goldstein and Sammons (1997) utilized cross-classified modeling to investigate the effects of different primary school characteristics on student exam performance. The study was designed to see what carryover effects the primary school attended might have on students' progress in middle school. The research sample consisted of 758 students in 48 primary schools that went on to 116 different middle schools. In a purely nested data structure, students at a particular middle school come from the same primary school; however, in this study, not everyone at a particular primary school attended the same middle school. Therefore, students were cross-classified by primary and middle schools. The essential feature is that level-2 was a cross-classification of the different combinations of primary school and middle school that students (level-1) attended. They concluded that students with high average test scores at a given primary school also tended to have high scores in a given middle school.

CCREM can also be utilized with longitudinal designs where data might be collected from students during their middle school and high school years (a growth curve design). A researcher might be interested in looking at test score growth over multiple years. The test could be given first at seventh grade and then every year through eleventh grade. Although students are nested in both middle schools and high schools, students in the same middle school may not attend the same high school and vice versa. In this case, students are cross-classified by both the middle school and high school they attended (Goldstein, 1995). It is very likely that both middle and high school characteristics contribute to the variance in test scores; thus, CCREM should be used.

CCREM Example: Fully Unconditional Model

Raudenbush and Bryk's (2002) example of cross-classified neighborhoods and schools will be used to illustrate a simple CCREM model. The fully unconditional model estimates the variation between neighborhoods, between schools, and within schools and neighborhoods. Each potential combination of school and neighborhood is referred to as a cell in the CCREM literature. The level-1 or within-cell model represents the relationships among the student-level variables; the level-2 or between-cell model captures the influence of school- and neighborhood level factors. In the model, there are $i = 1,2,...,n_{jk}$ level-1 units (e.g., students) nested within cells cross-classified by $j = 1,..., J$ first level-2 units (e.g., neighborhoods), designated as rows, and $k = 1,..., K$ second level-2 units (e.g., schools), designated as columns. The notation of (jk) indicates that the cells are conceptually at the same level: the $(jk)^{th}$ neighborhood/school (Raudenbush & Byrk, 2002). At level-1, students (within-cell) are nested within each cell of the cross-classification. The level-1 unconditional model is

$$Y_{i(jk)} = \beta_{0(jk)} + e_{i(jk)} \tag{7}$$

where $Y_{i(jk)}$ is the test score for student *i* within the cross-classification of neighborhood *j* and school *k;* $\beta_{0(jk)}$ is the mean achievement score for students who attend school k and live in neighborhood *j*; and $e_{i(jk)}$ is the random student effect (a residual error term). The error is assumed to be normally and independently distributed with a mean of 0 and a constant variance, σ^2.

At level-2 (between cells), variation between cells represents the components of school effects, neighborhood effects, and a school by neighborhood interaction effect. The level-2 unconditional model is

$$\beta_{0(jk)} = \gamma_{000} + b_{0j0} + c_{00k} + d_{0(jk)} \tag{8}$$

where $\beta_{0(jk)}$ is the intercept from level-1, γ_{000} is the average achievement score for all students; b_{0j0} is the random effect of neighborhood *j*, which is assumed to be normally distributed with a mean of 0 and a constant variance, τ_{b000}; c_{00k} is the random effect of school *k*, which is also assumed to be normally distributed with a mean of 0 and a constant variance τ_{c000}; and $d_{0(jk)}$ is the random interaction effect of school and neighborhood, which is assumed to be normally distributed with a mean of 0 and a constant variance, τ_{d00}. The random interaction effect is the deviation of the cell mean from the two main effects (neighborhoods and schools). Usually, the within-cell sample sizes are not big enough to distinguish the variance associated with the interaction effect, τ_{d000} from the within-cell variance, σ^2. Therefore, Raudenbush and Bryk (2002) typically recommend dropping the interaction effect from the model.

Similar to the ICC, the intra-unit correlation coefficient (IUCC) is used with CCREM to determine the proportion of variability that is between schools and that which is between neighborhoods. The following formula calculates the IUCC for students attending the same school and living in the same neighborhood for models where the interaction has not been estimated (Raudenbush & Bryk, 2002):

$$P_{bcd} = \frac{\tau_{b000} + \tau_{c000}}{\tau_{b000} + \tau_{c000} + \sigma^2} \quad . \tag{9}$$

This formula gives the correlation between the outcomes of students who attend the same school and live in the same neighborhood. This can also be considered as the proportion of variance in the outcome that lies between each of the cross-classified factors (here, between schools, between neighborhoods, and between cells – the interaction between neighborhood and schools).

The following formula provides the IUCC for students attending different schools but living in the same neighborhood (Raudenbush & Bryk, 2002).

$$P_b = \frac{\tau_{b000}}{\tau_{b000} + \tau_{c000} + \sigma^2} \quad . \tag{10}$$

This provides the proportion of variance that is attributable to neighborhoods. Finally, the following formula calculates the IUCC for students attending the same schools and but living in different neighborhoods:

$$P_c = \frac{\tau_{c000}}{\tau_{b000} + \tau_{c000} + \sigma^2} \quad . \tag{11}$$

The formula gives the correlation between outcomes of students who attend the same school but live in the different neighborhoods. This provides the proportion of variance that is attributable to school effects.

CCREM Example: Fully Conditional Model

After estimating the unconditional model, predictors can be included to explain variation. In this example, two level-2 variables can be included: the proportion of students receiving a reduced lunch rate *(RLR)*, which is a school characteristic, and socio-economic status (*SES)*, which is a neighborhood characteristic. Additionally, one student-level variable (*gender*: female=1, male=0) will be included in the level-1 model. The level-1 equation is as follows:

$$Y_{i(jk)} = \beta_{0(jk)} + \beta_{1(jk)} X_{i(jk)} + e_{i(jk)} \tag{12}$$

At level-1, $Y_{i(jk)}$ is the math test score, and $X_{i(jk)}$ is the gender of student *i* in neighborhood *j* and school *k*, $\beta_{0(jk)}$ is the intercept, or predicted achievement score for males in cell (*jk*), and $\beta_{1(jk)}$ is the regression coefficient, or the predicted change in the test score for females in cell (*jk*). Furthermore, $e_{i(jk)}$ is the within-cell random effect, or the deviation of student *i* in neighborhood *j* and school *k*'s achievement score from the predicted score after taking gender into account. It is assumed normally and independently distributed with a mean of 0 and a constant variance, σ^2.

At level-2, each of the level-1 coefficients becomes an outcome that represents the variation between cells created by the crossing of the two factors. This variation can be modeled as a function of level-2 predictors, such as characteristics of the schools or the neighborhoods, and the interactions between neighborhood and school predictors.

To explain variability in test scores, one school (k) characteristic (RLR) and one neighborhood (j) characteristic (SES) will be included at level-2. The level-2 equation follows:

$$\beta_{0(jk)} = \gamma_{000} + \gamma_{010}W_j + \gamma_{020}Z_k + b_{0j0} + c_{00k} \qquad (13)$$

$$\beta_{1(jk)} = \gamma_{100}$$

W_j is the SES of the neighborhood, and γ_{010} is its effect on the outcome; Z_k is the RLR of the school, and γ_{020} is its effect on the outcome; and b_{0j0} and c_{00k} are residual random effects of neighborhood and schools, respectively. The effect of neighborhood SES, γ_{010}, could also be modeled as randomly varying across schools and similarly the coefficient of Z_k (the RLR of the school), γ_{020}, could be modeled to vary across neighborhoods. The effect of level-1 predictor and each cross-classified factor's characteristics (Z_k and W_j) are fixed across the cross-classified factors.

Then, combining level-1 and level-2 and creating a single equation becomes:

$$\begin{aligned} Y_{i(jk)} &= \gamma_{000} + \gamma_{010}W_j + \gamma_{020}Z_k + \gamma_{100}X_{i(jk)} \\ &+ \left[b_{0j0} + c_{00k} + e_{i(jk)}\right] \end{aligned} \qquad (14)$$

Cross-Classified Random Effect Models in the Literature

To date, the use of CCREM to analyze cross-classified data is not very common in educational research (Meyers, 2005). There have also been few methodological CCREM studies that have been conducted (see Browne, Goldstein, & Rasbash, 2001; Clayton & Rasbash, 1999; Meyers & Beretvas, 2006).

The purpose of Meyers and Beretvas' (2006) study was to investigate the impact of inappropriate modeling of cross-classified data by analyzing both real and simulated data. Their study was comprised of two parts. First, a Monte Carlo simulation was designed to investigate possible factors influencing the use of CCREM. Additionally, the impact of ignoring the cross-classification of data was investigated. In order to determine which study conditions would be employed, a review was first conducted of the dataset from the National Education Longitudinal Study of 1988 (NELS:88). Students in this dataset were nested within a cross-classification of middle and high schools. In addition to creating simulation study conditions that were reflective of this dataset, they also used this example to illustrate the simulation model that was utilized. A variety of conditions were manipulated to explore what differences in variances would occur as compared to properly modeled cross-classified data. In order to illustrate the CCREM simulation model, they used an example where students were cross-classified by middle and high school. In this example, they considered students who were measured on a standardized test during high school but the effects of high school and middle school characteristics were of interest for their impact on the outcome variable. They included one student predictor (gender) at level-1 and one predictor for each of the two cross-classified factors (school size for middle schools and reduced lunch rate for high schools) were included in the model at level-2.

The first study factor examined was the correlation between the residuals for the two cross-classified factors. Meyers and Beretvas (2006) created a logical pattern for combining the levels of the cross-classified factors based on the assumption that real data situations there likely show correlations among the cross-classified factors (Bryk & Raudenbush, 2002). For example, students from a low-income middle school most likely will also attend a low-income high school. Therefore, the authors used two levels of this factor: one in which the middle school and high school conditional residuals were correlated (ρ =0.40), and the other in which no correlation was present between the residuals (ρ =0.00).

The second manipulated study factor was the number of middle schools feeding into each high school. Using the NELS:88 dataset as a guideline, Meyers and Beretvas (2006) found that typically either two or three middle schools fed into a given high school; hence, these two levels were included in their study.

Next, in order to manipulate sample size, the number of middle schools and high schools were manipulated. The number of schools, 30, was considered to represent a small sample size within each cross-classified factor (middle schools and high schools) and 50, represented a large number of schools within each cross-classified factor. The average number of students sampled from each high school was also manipulated in the study. The number of students in each high school was randomly generated as the sample size across high schools would tend to be different. The number of students for each high school was drawn from a normal distribution with either a mean of 20 or a mean of 40 and a standard deviation of 2. Again, the sample size was chosen to mirror the real data situation from the NELS:88 dataset.

Lastly, based on a search of applied studies utilizing CCREM as well as examples in multilevel modeling textbooks, Meyers and Beretvas (2006) found that the conditional IUCC estimates ranged from 0.009% to 24.0%. In order to represent small and moderate IUCCs, Meyers and Beretvas chose 5% and 15%, respectively as levels to include in their simulation study. The random interaction effect of middle school and high school was not modeled as “it is typically not well estimated” (Meyers & Beretvas, 2006, p. 483).

The estimated fixed and random effects as well as their standard errors were summarized and compared across the 1,000 replications in each condition. The percent relative bias for the parameter and standard error estimates were calculated. Conditions where there was no correlation (ρ =0.00) between the two cross-classified factors' residuals showed more biased standard errors when the cross-classified structure was ignored. The authors explained this result as when the middle school and high school were related, this relationship might explain some of the variation attributable to the other factors. Additionally, the standard error bias magnitude increased when the sample size of the cross-classified factors was large when the cross-classified structure was ignored. This result might be related to "*design effect*" in cluster sampling. The sample size in per-cluster usually influences the degree of design effect (Kalton, 1983). Estimates of the level-1 variance and associated standard error estimates were also affected by model misspecification. When the cross-classified structure was ignored, more positive bias emerged in the level-1 parameter estimates.

In the second part of their study, Meyers and Beretvas (2006) conducted an applied analysis where they used the NELS:88 dataset to compare parameter estimates when the cross-classified data were modeled and then in two scenarios where the data were modeled without taking into account the cross-classification. The NELS:88 data set consisted of 25,000 eighth grade students who were tested over eight years. It should be noted that although the data collected in NELS:88 were longitudinal, the proposed models examined in both the simulation and the applied portion of their study were not growth models but instead cross-sectional. Instead, the impact of middle and high school characteristics were taken into account for a standardized measure assessed in high school. Students were nested within a cross-classification of middle school and high school. The two HLM analyses were separated as HLM-Delete and HLM-Complete, where HLM-Delete removed all cases that did not meet the conditions of the primary case. Specifically, in order to make a more purely nested scenario, they only included students in their analysis if in attending a particular high school they also attended the primary feeder middle school. Any student at a high school that came from a middle school other than the primary feeder middle school was deleted from the data set. The purpose of this method was to remove the cross-classification from the data set by deleting students not from the same primary feeder schools. A two-level HLM model was then used where one explanatory variable (gender) was included at level-1 and two explanatory variables (reduced lunch rate and school size) were included at level-2.

The second traditional HLM method employed, HLM-Complete, was also designed to ignore the cross-classification of the data. In this analysis the variability due to middle schools was ignored and only the nestedness from high school membership was included. The two level HLM model with two explanatory variables (gender and middle school size) at level-1 and one explanatory variable (reduced lunch rate) at level-2 were used.

The results of HLM-Complete and HLM-Delete were then compared with the results from when the cross-classification of the data was correctly taken into account (when CCREM was utilized). In particular, the authors were interested in comparing the estimates and statistical significance of each fixed effect across the models. The comparison of the results of the three models showed that the parameter estimates differed for data analyzed using CCREM as opposed to the other two models, which means that the analysis for a misspecified model would be misleading. Additionally, they concluded that when a researcher models only high school (ignoring middle school clustering), he or she may find a more significant high school variance component than would be found if the cross-classification was taken into account. This may lead the generalizability of the study in a wrong way, explaining all variability with only high school characteristics.

Meyers and Beretvas (2006) found from both studies that if a cross-classified factor is ignored, its related variables will show that they have more of an influence than they actually do. The major finding from the simulation study was that when the cross-classification was ignored, the fixed effect parameter estimates were not biased; however, the standard errors were underestimated where it might result in Type I error rate for test of statistical significance when compared with traditional HLM.

Missing Data

A common problem in any type of multilevel analysis is the presence of missing data at any given level. Missing data have become an important issue in many empirical studies (Lepkowski, Landis & Stehouwer, 1987; Little & Rubin, 1987) but to date little attention has been paid to their effects on HLM analyses. Lack of response can occur for a variety of reasons, especially in studies with large sample sizes and in longitudinal studies where the probability of subject attrition increases. Most statistical procedures were designed for complete data sets; thus, dealing with missing data is problematic for data analyses. Moreover, the most important concern regarding missing data is the extent to which the missing information can affect study results (McKnight, McKnight, Sidani, & Figueredo, 2007). Missing data may bias parameter and standard error estimates, inflate Type I and Type II error rates, and result in a reduction of statistical power (Allison, 2001).

The framework of missing data inference was developed by Little & Rubin (1987). He discussed three different missing data mechanisms:

1. *Missing completely at random (MCAR).* MCAR occurs when missing data on variable *Y* is not related to the value of *Y* itself or to the values of other variables in the data set. For example, one student might accidentally skip an item on a given test. In other words, the missingness occurs purely by chance.
2. *Missing at random (MAR).* MAR occurs when the probability of missing data on *Y* is not related to the value of *Y*, but is related to the value of one or more completely observed variables in the data set. For example, suppose a drug use questionnaire was administered and non-response occurred mostly among high socioeconomic status (SES) students. However, the missing values were not related to the frequency of drug use for students with the same SES status. In other words, the missingness on drug use is not related after controlling for SES. Most methods for handling missing data are designed under this assumption.
3. *Missing Not At Random (MNAR).* If the missingness depends on the value of *Y*, then MNAR occurs. For example, when asking students for their test scores in a survey, it might be that missing values are more likely to occur when students have low test scores and are embarrassed to report them.

Several authors have stated that the MCAR condition may rarely occur in practice (e.g., Muthen, Kaplan, & Hollis, 1987; Enders, 2003). MCAR is also considered to be a special case of MAR (Enders, 2003). MNAR creates a situation that is difficult to handle in a study. Because the missingness of the values is related to the variable itself, there is usually no information to access the missing observations. To distinguish the MNAR and MAR, the relationship between the missing information and observed data needs to be known (McKnight et al., 2007). This relationship is difficult to distinguish unless a researcher asks the respondent the reason of non-response to the particular question. There are a number of widely employed approaches to handle missing data. In the next section, several techniques for handling missing data will be discussed.

Methods for Dealing with Missing Data

There are many approaches for handling missing data including listwise deletion, pairwise deletion, hotdecking, mean substitution, maximum likelihood estimation, and multiple imputation (Allison, 2001). This paper will focus on the sophisticated and rapidly increasing in use multiple imputation.

Listwise Deletion (LD). LD deals with missing data in an intuitive way by dropping all cases with missing values. It is the default in most multilevel modeling software packages such as HLM 6.0 (Raudenbush, Bryk, Cheong & Congdon, 2005), SPSS (version 15.0, SPSS, 2007) and SAS PROC MIXED (SAS Institute, 2006). According to Allison (2001), a researcher employing LD might delete a large proportion of cases with a resulting loss of statistical power, data variability, and generalizability. This is particularly a problem when researchers have a small sample size and high rates of missing data, because they might end up losing a large portion of the data. LD is appropriate when data are MCAR because reduced sample size will be a random subset sample of the original sample; therefore, parameter estimates should be unbiased (Little & Rubin, 1987). However, when the data are MAR or MNAR, LD can yield biased estimates (Allison, 2001). When the probability of missingness on any of the independent variables depends on the values of the dependent variable, the estimates of an analysis (e.g. regression) using LD could be biased (Allison, 2001) However, if data are MCAR and small amounts of data are missing then using LD can be appropriate.

Multiple Imputation (MI). MI is rapidly increasing as a preferred method and is one of the most sophisticated methods for handling missing data. Rubin (1987) developed the basic idea of MI, which was to treat missing data as random variables and replace the missing values with more than one simulated value. The goal was to create simulated values for the missing values so that the complete data set would replicate the original variance-covariance matrix. MI makes the assumption that the data are at least MAR. Thus, MI will also work well with MCAR (Schafer & Olsen, 1998).

MI has three phases: imputation, analysis, and pooling of parameter estimates (Peugh & Enders, 2004). In the imputation phase, each missing value is replaced by a set of $m > 1$ plausible values, where m is typically a small value (3-10) (Schafer & Olsen, 1998). The imputation phase is an iterative procedure. A series of predicted values are derived from a covariance matrix of existing data, and missing values are replaced by the predicted scores from those equations (Schafer, 1997). This process will be elaborated on below. In the second phase, each of the complete datasets is analyzed by the desired statistical methods (here, CCREM). At the pooling stage, the results (parameter and standard error estimates) are combined using rules provided by Rubin (1987) to produce overall parameter estimates and standard errors that reflect the missing data uncertainty (Schafer & Olsen, 1998). Furthermore, Schafer (1997) suggested that no more than 10 imputations are required if the fraction of missing data is not large. Gold, Bentler, and Kim (2003) suggested that with 30% or less missing data, five imputed data sets should be sufficient.

Allison (2001) summarized MI in six steps. In the first step, the Bayesian Markov Chain Monte Carlo (MCMC) method uses the expectation maximization (EM) algorithm where the means and standard deviations from available cases are calculated as the initial estimates. The resulting estimates are then used to begin the MCMC process. Because the model parameters are estimated from the observed and filled-in data, the parameters can be considered to have a posterior probability distribution. In the second step, for each missing data pattern, a series of regression equations are created using the initial estimates from the first step (the means and covariance matrix of the variables). In the third step, missing values are imputed from the predicted values from the series of regression equations and a random draw is made from the distribution of residuals where it is added to the predicted values. After imputing all missing data, the fourth step starts where all the parameters are re-estimated using the "completed" data set. In the fifth step, based on the newly calculated means and covariances, a random draw takes place from the posterior distribution of the means and covariances (the posterior step). In the last step, steps 2 through 5 are repeated until the distribution of covariance matrices stops changing in a substantial way and every *k*th iteration is extracted to be utilized for data analysis.

Each of the *m* datasets is then analyzed with HLM and then the results are pooled together using Rubin's rules (1987) for combining parameter and standard error estimates to result in a single set of estimates. A single estimate for each parameter is obtained by averaging across the *m* imputed data sets:

$$\overline{Q} = \frac{1}{m}\sum_{i=1}^{m}\hat{Q}_i \tag{15}$$

where $\hat{Q}_i$ is the parameter estimate from the *i*th imputed data set.

The variance estimate for each parameter takes into account the variability within each imputed dataset

$$\overline{U} = \frac{1}{m}\sum_{i=1}^{m}\hat{U}_i \tag{16}$$

where $\hat{U}_i$ is the variance estimate from the *i*th imputed data set, as well as the variance between imputations

$$B = \frac{1}{(m-1)}\sum_{i=1}^{m}(\hat{Q}_i - \overline{Q})^2 . \tag{17}$$

Finally, the total variance will be calculated, which is the variance estimate associated with $\hat{Q}$

$$T = \overline{U} + (1 + \frac{1}{m})B . \tag{18}$$

MI can provide more information in the presence of small sample sizes or high rates of missing data with multivariate statistical approaches. MI can be used almost in any setting, especially if the data are MAR (Allison, 2001). As Allison (2001) notes,

> All the common methods for salvaging information from cases with missing data typically make things worse: They introduce substantial bias, make the analysis more sensitive to departures from MAR and MCAR, or yield standard error estimates that are incorrect (usually too low) (p.12).

Multiple imputation is also an increasingly popular method in applied multilevel modeling. Sample size is a crucial factor in multilevel analysis due to its statistical power. Because variances at all levels of the data are analyzed simultaneously, the multilevel analyses require larger sample sizes than other multivariate procedures (Zhang, 2005). Therefore, handling missing data in multilevel models is important since it requires a large sample size. MI also can be easily implemented with SAS, using PROC MI and PROC MIANALYZE (SAS Institute, 2005). In addition; MI can be implemented through other programs: NORM (Schafer, 1999), CAT (Schafer, 1997a), PAN (Schafer, 1997b) SPSS (version 15.0, SPSS, 2007) and S-PLUS (version 6.0, Insightful Corporation, 2001).

Missing Data and Multilevel Modeling in the Literature

The topic of missing data with multilevel modeling has not been frequently explored. Gibson and Olejnik (2003) conducted a simulation study comparing eight missing data techniques: LD, pairwise deletion (where all available data are used and no method is used to replace the missing values), single imputation (utilizing methods such as regression for imputing missing values), mean substitution (replacing all missing data in a variable by the mean of that variable), weighting imputation (weighting parameters based on the observed data), group mean substitution (replacing missing data in a variable by the mean of sub groups' value of that variable) the EM algorithm (the beginning idea of imputation), and MI with a traditional HLM. In their simulation study, the missing data were at the second level of a two-level HLM. The manipulated conditions included the number of level-2 variables (2 and 4 explanatory variables), level-2 sample size (sample size of 30 and 160), level-1 intercept-slope correlation (correlated either at .2 or .8), and percentage of missing data (10% and 40% missingness). They found that LD, the EM algorithm and group mean substitution performed equally well, whereas overall mean substitution and MI performed poorly when compared to complete data condition (as a baseline condition). They attributed the poor performance of MI (which was hypothesized to perform well) to the fact that only level-2 data were missing and when the sample size at level-2 was small (30), results based on the MI may have become unstable. Furthermore, MI performed better (not biased parameter estimates) with a level-2 sample size of 160, which was the large sample condition when a missing data proportion was 10%. When the missing data proportion was 40% and the sample size was 160, MI yielded overestimates of the variability of the school means (intercepts) and also overestimates of the variability of slopes when comparing to complete data condition.

Another simulation study examining multilevel modeling and multilevel structural equation modeling (MSEM) with missing data was conducted by Zhang (2005). The influence of non-normality and performance of multiple imputation utilizing the expectation maximization (EM) algorithm was investigated. The statistical power, bias for the parameter and standard error estimates for the main effects and cross-level interaction in a two-level model were compared across four manipulated conditions: analysis methods (HLM and MSEM), percent of missing data at level-1 and level-2 (15%

and 30% missingness), total sample size (300, 500, 1500, and 2500) and normality condition (three configurations: normality, moderate non-normality, severe non-normality). These configurations were achieved by different combinations of skewness and kurtosis following Fleishma's power transformation method. All methods for handling missing data were compared to baseline conditions with complete data. Because of the model difference between HLM and MSEM only the main and interaction effects were compared and presented. The power based on imputed data was compared with complete data condition, and both HLM and MSEM were similar in terms of their estimation of the main effects and cross-level interaction. The bias of the parameter and standard error estimates was evaluated with the root mean square of the difference RMSD across the 100 iterations. The means and standard deviations of the RMSD for the main and cross-level interaction effects and their standard errors were compared using factorial ANOVAs. Using MI procedure, all the interactions among missingness proportion, analysis method and sample size were significant. When comparing parameter estimates, a higher portion of missing data (30%) generated larger bias magnitude with MI. Furthermore, their results showed that a higher proportion of missing data (30%) tended to produce more bias in parameter and standard error estimates.

One of the benefits of MI is the use of other related variables (termed auxiliary variables) to improve the quality of the imputed data. Rubin (1996) explored the possibility of including auxiliary variables in the context of MI. He argued that when auxiliary variables are added to an imputation procedure, the bias and efficiency may improve even though the auxiliary variables are not included in the final statistical analyses (e.g., the estimation of the HLM). Furthermore, they are not related to the interest of the study, but possibly related to the inclination of the missing data. Several previous methodological studies (Collins, Schafer & Kam,2001; Peugh &Enders, 2004) suggested that the use of auxiliary variables can improve the performance of MI. Using auxiliary variables might reduce possible bias (Collins, 2006). For example, a researcher might be interested in looking at student achievement on a number of educational predictors. In the study, it is suspected that socioeconomic status (SES) may be related to the reason why data are missing, but SES is not a variable that is of interest in the HLM. SES can be used as a predictor in the imputation model and this should enhance the results of MI since achievement and SES are correlated.

An important methodological study regarding missing data was conducted by Peugh and Enders (2004). In their study, the authors provided an overview of missing-data theory, maximum likelihood estimation, and multiple imputation. They also reviewed the methods of 23 applied research studies with missing-data and provided a demonstration of MI and maximum likelihood estimation (ML) using the Longitudinal Study of American Youth data. They also explored the use of an auxiliary variable with MI in a simulation study. The results indicated that exploring the impact of missing data in applied research increased significantly between 1999 and 2003, but the use of ML or MI was rare; the applied studies relied almost completely on listwise and pairwise deletion. The results of their simulation study indicated that the HLM parameter estimates were comparable when either ML or MI was utilized. However, the study results showed that the use of an auxiliary variable with MI improved the parameter estimates; hence, reducing bias.

Collins, Schafer, and Kam (2001) also explored the use of an auxiliary variable with MI. A simulation study was conducted to compare the minimal and maximum (the dosage) use of auxiliary variables with ML and MI. The minimal use of auxiliary variables was described as including few or no auxiliary variables whereas the maximum use of auxiliary variables was described as including numerous auxiliary variables. When missingness was MAR, including one or more auxiliary variable resulted in a decrease in standard error bias, and thus an increase in statistical power. Furthermore, they advised that missing data software programs should be designed to encourage researchers to use auxiliary variables.

A search of recent educational research (1995-2008) using the terms "cross-classified data," "multiple imputation," "CCREM," "missing data," and "incomplete data" in the PsycInfo database indicated that there were 14 studies using CCREM models, but none of these studies discussed whether there were missing data or how missing data were handled; hence, techniques for handling missing data were not discussed.

Statement of the Problem

Because cross-classified modeling is relatively new, there has been no simulation research on the impact of missing data on its parameter and standard error estimates. Using a search of applied articles from PsycInfo database, researchers conducting CCREM were typically seen to use the default for handling missing data, LD. Additional research is needed to explore more advanced techniques such as MI. A Monte Carlo simulation study was conducted to investigate the impact of missing data using an educational setting illustration where students belong to more than one unit at a given level.

The purpose of this simulation study was to examine the impact of MCAR, MAR, and MNAR data on CCREM estimates when MI is utilized. The performance of this method was compared under a number of manipulated conditions including the percentage of missing data and the use of an auxiliary variable. This research expanded on the CCREM simulation work of Meyers (2004) by the inclusion of studying the impact of missing data and use of MI with CCREM.

CHAPTER 3

METHOD

The purpose of this Monte Carlo simulation study was to examine the impact of missing data on CCREM estimates when multiple imputation (MI) is employed. This research expanded on the CCREM work of Meyers (2004) by the inclusion of studying the impact of missing data and use of MI with CCREM. In order to illustrate this type of scenario, Meyers and Beretvas's (2006) example of students who are cross-classified within middle schools and high schools was used. Several factors were manipulated to investigate the impact of missing data at level-1 on cross-classified models. These factors included the type of missingness, the percent of missingness, the correlation between the residuals for the two cross-classified factors, the numbers of levels of each cross-classified factor, the number of middle school feeding into each high school, and the correlation between the level-1 predictor and an auxiliary variable. The performance of MI was assessed through the relative bias of the model fixed effects and their standard error estimates.

Study Design

In this simulation study, a simple multilevel scenario where students are cross-classified by middle school and high school was used as the framework to illustrate the multilevel model. Similar to the model used in Meyers and Beretvas (2006), in this simulation study three predictor variables were included in the generating model. At level-1, one predictor, "reading achievement score" (*X*), and at level-2, one predictor for each of the two cross-classified factors were included. As in Meyers and Beretvas' (2006) CCREM model, "school size" (*W*) was included as the middle school predictor and "the percent of students receiving free or reduced lunch" (*Z*) was used as the high school predictor. Furthermore, "SAT scores" (*Y*) was used as the outcome variable. The model estimated is as follows:

Level 1:

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

Level 2:

$$\beta_{0(jk)} = \gamma_{000} + \gamma_{010} W_j + \gamma_{020} Z_k + b_{0j0} + c_{00k} \quad (20)$$

$$\beta_{1(jk)} = \gamma_{100}$$

Combining equations, the final model becomes:

$$Y_{i(jk)} = \gamma_{000} + \gamma_{100} X_{i(jk)} + \gamma_{010} W_j + \gamma_{020} Z_k + \left[b_{0j0} + c_{00k} + e_{i(jk)}\right] \quad (21)$$

where *i* indexes students, *j* indexes high schools, and *k* indexes middle schools. At level-1, $Y_{i(jk)}$ is the "SAT score" for student *i* who attended middle school *k* and high school *j*. The level-1 predictor, $X_{i(jk)}$, is the individual student's reading achievement score. $\beta_{0(jk)}$ is the mean SAT score for students who attended middle school *k* and high school *j*, controlling for the other model predictors. $\beta_{1(jk)}$ is the predicted change in SAT score of a student in cell (*jk*) associated with a one point increase in reading achievement score, and $e_{i(jk)}$ is the random student effect, or the deviation of the student's SAT score from the predicted score based on the reading achievement score. This error term is assumed to be normally and independently distributed with a mean of 0 and a constant variance, σ^2. At level-2, the intercept, $\beta_{0(jk)}$, was modeled in a way that part of its variability could be explained by both level-2 predictors. γ_{000} is the overall average SAT score controlling for W_j and Z_k. The level-2 predictor, Z_k, represents the number of students in middle school *k* and W_j represents the percent of students receiving free or reduced lunch in high school *j*. The conditional residuals, b_{0j0} and c_{00k}, have variances τ_{b000} and τ_{c000}, respectively. Additionally, these residuals are assumed to be normally and independently distributed with a mean of 0 and a constant variance. The effect of reading achievement, $\beta_{1(jk)}$, was modeled to be fixed across middle and high schools at level-2, γ_{100}.

In order to compare the impact of missing data on CCREM models to what has already been examined in the CCREM simulation literature, a number of the same conditions employed in the Meyers and Beretvas (2006) study was included in this study in addition to the missing data conditions. This study was designed to provide preliminary investigation into missing data issues at level-1 with cross-classified data while using MI. In addition, baseline conditions of no missingness were analyzed. The following section briefly describes the conditions that were manipulated in the Meyers and Beretvas (2006) simulation study along with new conditions to investigate the impact of missing data at level-1.

Design Conditions

Correlation of the residuals at level-2. In CCREM models, it is typically assumed that the cross-classified factors (middle schools and high schools) are in some way correlated (Meyers & Beretvas, 2006). For example, it is more likely that students from economically disadvantaged middle schools attend similarly economically disadvantaged high schools, showing a logical explanation of pattern between socioeconomic status of middle school and high school. To mimic data from the NELS:88 dataset, Meyers and Beretvas (2006) used two levels of this factor: one in which the middle school and high school conditional residuals were correlated (ρ =0.40), and the other in which no correlation was present between the residuals, (ρ =0.00). These same two levels were incorporated into this study.

Number of feeder schools. Again using the NELS:88 dataset as a guideline, Meyers and Beretvas (2006) found that typically either two or three middle schools fed into a given high school. Therefore two levels of this factor, two middle schools feeding into each high school and then three middle schools feeding into each high school, were used.

Number of levels of the cross-classified factors. Using the NELS:88 data, Meyers and Beretvas found that 30 represented a small number of schools within a cross-classified factor whereas 50 represented a large number of schools within a cross-classified factor. Thus in this study, the number of middle schools and high schools were simulated to be either 30 or 50. In addition, the number of middle schools was always simulated to be equal to the number of high schools.

Additionally, Meyers and Beretvas (2006) manipulated the number of students at each high school. The number of students in each high school was randomly generated, drawing from a normal distribution with a known mean (with values of 20 or 40) and standard deviation of 2. Thus, the average number of students in each high school randomly varied around 20 or 40. However, to minimize the number of conditions, the number of students sampled from each school was not manipulated and was fixed at 30. The sample size of 30 is selected because this is the average number of students in each high school in Meyers and Beretvas's (2006) study. As a result, when the level of the cross-classified factor was 50, the total sample size consisted of 1500 students, and when it was 30, the total sample size consisted of 900 students.

Percentage of missing data. Missing data were introduced only in the level-1 predictor, reading achievement. Three levels of the percentage of missing data were specified as suggested by Gold, Bentler, and Kim (2003): 15%, 30%, and 45%. The authors considered these three levels to be typical for low, moderate, and high percentages, respectively, of missing data. To illustrate, in conditions with 15% missing data, 15% of simulees was missing a value on their level-1 predictor. The description of how the missing data were introduced can be found below.

Type of missingness. Three different types of missingness were explored in this study. These three types of missing data were the MCAR, MAR, and MNAR missing data mechanisms (Little & Rubin, 1987). With MCAR data, the missingness is not related to values that would have been observed in the dataset. Thus, MCAR missingness was created so that values on the predictor variable at level-1 had an equal probability of being selected as missing. Because the MCAR data were deleted randomly, there was no relationship between the data that were missing and those that were observed. With MAR data, the missingness is related to some values on another variable observed in the dataset. Thus, in the MAR conditions, the level-1 predictor was set to have missingness based on values of another variable. With MNAR data, the missingness is related to the value of the variable itself (Little & Rubin, 1987) and so missingness was created based on the values on the level-1 predictor. The processes for creating these missing data scenarios will be elaborated upon in this chapter.

Correlation of auxiliary variable with the level-1 predictor. In order to test whether the presence of an auxiliary variable is improved the performance of multiple imputation, the last manipulated condition was the level of correlation between an auxiliary variable ("GPA") and the level-1 predictor (reading test score). The correlation between the two variables was set to 0.1 or 0.3, representing small and moderate correlations (Cohen, 1988).

Design Overview

The seven design factors that were examined in this study were fully crossed [2 (correlation of residuals) x 2 (number of feeder schools) x 2 (levels of cross classified factors) x 3 (percentage of level-1 predictor values missing) x 3 (type of missing data) x 2 (auxiliary variable correlations)]. Each of these 144 conditions was analyzed for their performance using multiple imputation. The results from multiple imputation were compared with baseline conditions of no missingness [2 (correlation of residuals) x 2 (number of feeder schools) x 2 (levels of cross classified factors)] for a total of 8 additional conditions. Tables 1 and 2 summarize the conditions of the study design.

Data Generation

Data were generated in SAS (SAS Institute, 2005) to fit a two-level, cross-classified multilevel model with students at level-1 nested within the cross-classified factors of middle school and high school at level-2. The data were generated based on the CCREM model that was described in Equation 21. One thousand replications were conducted for each of the 144 conditions with missing data and 8 conditions with

Table 1

Conditions of the Study Design with Missingness

Correlation of Residuals
1. 0.00
2. 0.40
Number of Feeder Schools
1. 2
2. 3
Levels of Cross Classified Factors
1. 30
2. 50
Percentage of Level-1 Data Missing
1. 15%
2. 30%
3. 45%
Type of Missing Data
1. MCAR
2. MAR
3. MNAR
Correlation of Auxiliary Variable with Level-1 Predictor
1. 0.10
2. 0.30

Table 2

Conditions of Baseline Study Design (No Missingness)

Correlation of Residuals
1. 0.00
2. 0.40
Number of Feeder Schools
1. 2
2. 3
Sample Size for Each Cross Classified Factor
1. 30
2. 50

complete data. The student level variable, *X* (reading test score); the middle school predictor, *Z* (school size); the high school predictor, *W* (percentage of reduced lunch); and the auxiliary variable, *A* (GPA); were generated from a normal distribution with a mean of 0 and a standard deviation of 1. Similar to the Meyers and Beretvas (2006) study, to represent moderate effect sizes, the slope coefficients was fixed at 0.5 (γ_{010}, γ_{020}, and γ_{100}) and the intercept was fixed at 1.0 (γ_{000}).

At level-1, Meyers and Beretvas (2006) included a dichotomous student level predictor (gender); however, even though missingness can occur on dichotomous variables, MI makes the assumption that the data are normally distributed. So while MI has shown to be fairly robust to violations of non-normality (Enders, 2001; Graham & Schafer, 1999), it was of interest in this study to evaluate conditions under which MI should perform more optimally. Therefore, in this study the level-1 predictor was continuous and not dichotomous. After the variable values were generated, then the auxiliary variable was made to correlate with *X* according to the appropriate study condition. In addition, the auxiliary variable was only be used at the imputation stage with MI and was not included for estimation of the final CCREM. Lastly, the generating value for the student-level error variance, σ^2, was fixed at 1.0.

Generating the combinations of middle schools and high schools

Meyers and Beretvas (2006) paired middle schools with high schools by using a "feeding pattern" that depicted a scenario similar to what was found in the NELS:88 data. The feeding pattern refers to middle school- high school paired combinations (here referred to as cells) that were used for data generation. This same pattern was also incorporated in this study. For this feeding pattern there was a cell for each combination of middle school and high school. With real data scenarios, it is not likely that every combination of middle school and high school occur. Therefore, the data were generated in a way that only some cells (combinations of middle schools and high schools) contained students. In order to create this feeding pattern, first a matrix of middle school and high school residuals was generated where the residuals between the two factors was correlated at either $\rho = 0.40$ or $\rho = 0.00$. After the matrix was created, the middle school residuals were sorted in ascending order as were the high school residuals. Each residual was assigned a rank based on its order.

In conditions where two middle schools fed into a high school, the high school received 60% of its students from the middle school with the same rank. Next, the high school received the rest of its students (40%) from the middle school with the closest, highest rank. When there were three middle schools feeding into a high school, the high school received 70% of its students from middle school with the same rank. For the remaining 30%, high school received 15% of its students from each of the two middle schools with the closest ranks (one with the higher rank and the other with the lower rank). These same percentages were used in conditions where the correlation between the residuals was set to zero; however, the ranking of middle schools and high schools was not done. Instead, the pairing of middle school with high school for the feeding pattern was done randomly. Figure 2 and Figure 3 demonstrate these feeding patterns.

In Figure 2 and Figure 3, schools have been ranked. For example, MS3 has the same rank as HS3. In the above example for the two feeder MS conditions, HS3 receives 40% of its students from MS2, and 60% of its students from MS3. In the three feeder MS

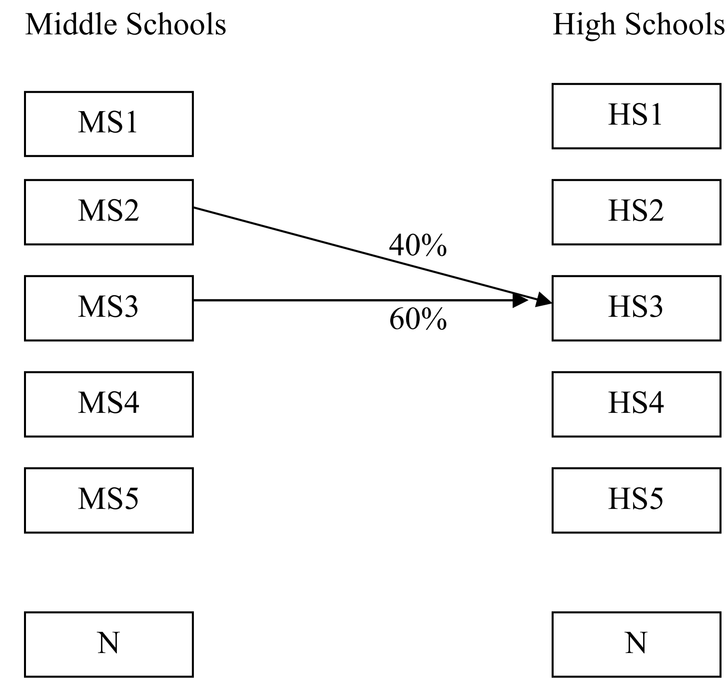


Figure 2. Two Feeder Middle School Condition. Adapted from Meyers & Beretvas, 2006.

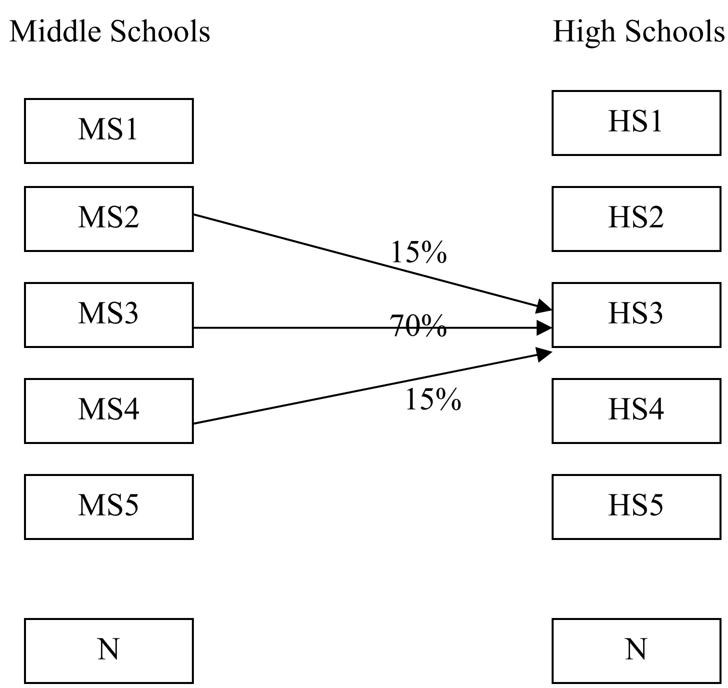


Figure 3. Three Feeder Middle School Condition. Adapted from Meyers & Beretvas, 2006.

Table 3

Patterns of Missingness Simulated

	Missingness	
	Type of Missingness	Percentage of Simulees with Level-1 Predictor Value Missing
Correlation of Residuals		
p=.40	MCAR	15%
	MAR	30%
	MNAR	45%
p=.00	MCAR	15%
	MAR	30%
	MNAR	45%
Number of Feeder Schools		
k= 2	MCAR	15%
	MAR	30%
	MNAR	45%
k=3	MCAR	15%
	MAR	30%
	MNAR	45%
Levels of Cross-Classified Factors		
n= 30	MCAR	15%
	MAR	30%
	MNAR	45%
n=50	MCAR	15%
	MAR	30%
	MNAR	45%

conditions, HS3 receives 15% of its students from MS3, 70% of its students from MS3 and 15% of its students from MS4.

Generating Type of Missingness and Percentage of Missingness

For this study, missing values was introduced in the level-1 predictor. In Table 3, the type and amount of missingness can be seen for each of the manipulated conditions. The simulation of the three missing data mechanisms mirrored the techniques used in Enders's (2003) simulation study where he examined the impact of MCAR, MAR, and MNAR missingness with structural equation modeling data. When data were MCAR, each simulee had an equal probability of having the level-1 predictor value ("reading score") set to missing. The percent of simulees with a missing value on the level-1 predictor (X_i) was generated according to the specified study condition. First, a column vector of uniform random numbers was generated where values were between 0 and 1. The deleting process was accomplished by pairing the column vector of X_i values with this column vector of uniform random numbers. If the uniform random number was smaller than the desired proportion of missing data, the corresponding X_i value was removed.

The simulation of MAR depended on values of the auxiliary variable, A_i, where i refers to simulee i. First, values on A were ranked in ascending order. Then, a percentile rank was assigned for each A_i. Next, a deletion probability, p_i, was generated and it was inversely related to the percentile rank of A_i. For example, when A_i fell at the 20th percentile, p_i was equal to .80, and when A_i fell at the 80th percentile then p_i was equal to .20. Next, a column vector of uniform random numbers was generated to be between 0 and 1 where each random number is indexed by v_i. The vector of X_i values were then paired with this vector of random numbers. Finally, the deletion process started with choosing the lowest A_i, then if $v_i < p_i$ then X_i was set to missing. This deletion process continued from lowest to highest until the required percentage of missing data was reached: 15%, 30% and 45% missingness. In this manner, simulees with low values on A_i had a higher likelihood of having missing values on X_i.

To simulate the MNAR missingness, a process similar to that of the MAR scenario was utilized. The values on X_i were placed in ascending order and then a percentile rank was assigned for each X_i. Next, a deletion probability, p_i, was generated, which was inversely related to the percentile rank of X_i. Next, a column vector of uniform random numbers was generated to be between 0 and 1 where each random number is indexed by v_i. The vector of X_i values was then paired with v_i. Next, the deletion process started by selecting the lowest X_i and X_i was deleted if $v_i < p_i$. The deletion process continued in ascending order, until the desired percentage of missing data was reached: 15%, 30% and 45% missingness.

Handling Missing Data

After each replication has been manipulated to meet the appropriate design characteristics, MI was used for each simulated dataset with missing data. Once the missing data were imputed, then PROC MIXED was used to analyze the CCREM model found in Equation 21. The imputation model included the variables from both the student level (level-1) ($Y_{i(jk)}$, $X_,$, and A_i) and level-2 variables (W and Z). PROC MI in SAS, version 9.1, (SAS Institute, 2005) was used to produce five imputed values (thus creating five datasets). According to Rubin (1996), three to five imputations are usually adequate in multiple imputation when the degree of missing information is not large. PROC MI uses the EM algorithm to obtain the initial estimates for the covariances between the variables. Then, the resulting estimates were used to begin the Markov Chain Monte Carlo (MCMC) process. The MCMC imputation method is the default option of SAS (SAS Institute, 2005). Then the values from every 200th iteration (the default in SAS) were extracted to form the five imputed datasets. These five datasets were then each be analyzed in PROC MIXED. Similar to an applied scenario where the auxiliary variable is not of interest for the research question, but may be related to the values of variables with missingness, the auxiliary variable was not included when estimating the CCREM model. After the five complete data sets were analyzed using PROC MIXED, PROC MIANALYZE was used to combine parameter and standard error estimates using Rubin's combination formulas (see Equations 15, 16, 17 and 18).

Data Analysis

The estimated fixed effects as well as their standard errors were summarized and compared across the 1,000 replications in each condition. To evaluate the standard errors, the estimated values were compared with empirical standard errors obtained by computing the standard deviation of the parameter estimates from all the simulated datasets in a condition. Percent relative bias was calculated for both parameter estimates and standard error estimates. The percent relative parameter bias (used with all model coefficients) was estimated by using the following formula (Hoogland & Boomsma, 1998):

$$B(\hat{\theta}_p) = \left(\frac{\bar{\hat{\theta}}_p - \theta_p}{\theta_p} \right) 100 \tag{22}$$

where $\bar{\hat{\theta}}_p$ is the mean of the *p*th parameter for the 1000 parameters estimates and θ_p is the corresponding population parameter.

The relative standard error bias was calculated using the following equation:

$$B(s\hat{e}_{\hat{\theta}_p}) = \left(\frac{s\bar{\hat{e}}_{\hat{\theta}_p} - s\hat{e}_{\theta_p}}{s\hat{e}_{\theta_p}} \right) 100 \qquad (23)$$

where $s\bar{\hat{e}}_{\hat{\theta}_p}$ is the mean of the 1000 estimated standard errors for $\hat{\theta}_p$ and $s\bar{\hat{e}}_{\hat{\theta}_p}$ is the empirical standard error, which will be calculated as the standard deviation of the 1000 estimates of θ_p. According to Hoogland and Boomsma (1998), acceptable cut off values for the relative parameter bias and relative standard error bias are within 5% and 10%, respectively.

CHAPTER 4

RESULTS

The present study was intended to investigate the use of multiple imputation with cross-classified data under different patterns of missing data (MCAR, MAR, and MNAR) while including an auxiliary variable that is correlated with the variable with missingness in the imputation model. A total of 144 conditions was used in this simulation study to explore the use of MI. In each condition, 1000 replications were used, resulting in a total of 144,000 simulated data sets. Hoogland and Boomsma's (1998) criteria of acceptability for relative bias for parameter and standard error estimates were used to evaluate the results of the simulated data sets.

Relative Percentage Bias of Parameter Estimates

The results from the relative bias of the parameter estimates can be seen in Tables 4-11. The relative percent bias for the parameter estimates was considered acceptable if its magnitude was less than 5% (Hoogland & Boomsma, 1998). The parameter relative bias of all coefficients in the baseline conditions (no missingness) was generally very small in magnitude and was always considered acceptable according to Hoogland and Boomsma's criterion. The relative bias magnitude in these baseline cells across all parameters ranged from -0.22% to 0.36%. In addition, the relative bias of γ_{000} ranged from -0.1% to 0.02% (see Tables 4-5). For the effect of the level-1 predictor, γ_{100}, the relative bias ranged from -0.3% to 0.36% (see Tables 6-7). Lastly, the relative bias of the effect of the middle school predictor, γ_{010}, and the effect of the high school predictor,

γ_{020}, ranged between -0.22% and 0.21% , and -0.01% to 0.15%, respectively (see Tables 8-11).

In cells with missing data, regardless of the manipulated condition, the relative bias for the parameter estimates were also always less than five percent in magnitude. In these cells, the relative bias of γ_{000} was small in magnitude and ranged from -1.2% to 1.8% (see Tables 4 & 5). The bias magnitude of the coefficient associated with the level-1 predictor (γ_{100}) was never larger than 4%. In these cells, the bias ranged from -0.24% to 3.8%. When the data were MNAR, the bias magnitude was slightly larger than other patterns of missing data (MCAR and MAR). For instance, with the MNAR data, the bias magnitude ranged from 2.4% to 3.8% whereas with the MCAR and MAR data, the bias ranged from -0.34% to 0.36% and -0.35% to 0.13%, respectively (see Tables 6 & 7). The relative bias of γ_{100} and γ_{020} were also small in magnitude. In these cells, the bias ranged from -0.56% to 0.39% and -0.24% to 0.34%, respectively (see Tables 8, 9, 10, & 11).

Relative Percentage Bias of Standard Error Estimates

Hoogland and Boomsma (1998) recommend a cutoff for acceptability of 10% for the magnitude of the bias of the standard errors. All standard error results can be seen in Tables 12-19. In these tables, any bias results that were greater than 10% in magnitude have been presented in boldface.

Standard error relative bias of γ_{000}. In the baseline condition (no missingness) for γ_{000}, the bias was never above 10 percent in magnitude. For these cells, the bias ranged from -2.7% to 2.6% (see Tables 12-13). In cells with missing data, the bias magnitude ranged between -30.6% and 9.5%. When the data were MCAR and MAR, there was no substantial bias in all cells. For these cells, the bias magnitude ranged from -7.8% to

Table 4

Relative percentage bias of γ_{000} with two feeder schools

Level-1 correlation	Type of missingness	Residual correlation	Missing percentage	Level-2 sample size 30	50
0.1	No Missing	0	0	-0.112	0.011
		0.4	0	-0.047	-0.050
	MCAR	0	15	0.036	-0.033
			30	-0.045	0.029
			45	0.061	0.025
		0.4	15	-0.071	-0.011
			30	0.004	0.132
			45	0.144	0.041
	MAR	0	15	-0.015	-0.057
			30	0.115	-0.020
			45	-0.047	-0.012
		0.4	15	0.079	1.857
			30	-0.016	0.035
			45	0.051	-0.026
	MNAR	0	15	-0.701	-0.705
			30	-1.110	-1.139
			45	-0.862	-0.940
		0.4	15	-0.810	-0.782
			30	-1.211	-1.233
			45	-0.931	-0.974
0.3	No Missing	0	0	-0.052	-0.012
		0.4	0	0.017	0.110
	MCAR	0	15	0.015	0.074
			30	0.112	-0.012
			45	-0.008	0.076
		0.4	15	-0.017	0.060
			30	0.040	0.034
			45	0.103	0.009
	MAR	0	15	0.119	0.111
			30	0.105	-0.034
			45	-0.033	0.013
		0.4	15	-0.093	0.022
			30	0.005	-0.012
			45	0.063	0.110
	MNAR	0	15	-0.770	-0.720
			30	-1.132	-1.136
			45	-0.843	-0.876
		0.4	15	-0.802	-0.795
			30	-1.210	-1.262
			45	-0.945	-0.977

Note. Absolute values below |5%| are considered acceptable for relative bias for parameters.

Table 5

Relative percentage bias of γ_{000} with three feeder schools

Level-1 correlation	Type of missingness	Residual correlation	Missing percentage	Level-2 sample size 30	50
0.1	No Missing	0	0	0.027	-0.008
		0.4	0	0.007	0.009
	MCAR	0	15	0.111	0.062
			30	0.032	-0.084
			45	0.025	0.057
		0.4	15	0.018	0.033
			30	0.157	-0.043
			45	0.128	-0.054
	MAR	0	15	-0.068	0.090
			30	-0.049	0.054
			45	0.106	0.043
		0.4	15	-0.060	0.108
			30	-0.007	0.008
			45	0.091	-0.027
	MNAR	0	15	-0.743	-0.734
			30	-1.125	-1.129
			45	-0.861	-0.897
		0.4	15	-0.796	-0.788
			30	-1.142	-1.186
			45	-0.941	-0.955
0.3	No Missing	0	0	0.013	0.007
		0.4	0	0.014	-0.038
	MCAR	0	15	-0.042	0.029
			30	0.096	0.058
			45	-0.120	0.032
		0.4	15	-0.064	-0.061
			30	0.012	-0.051
			45	-0.059	-0.076
	MAR	0	15	-0.066	0.065
			30	-0.104	0.059
			45	0.012	0.066
		0.4	15	0.070	0.048
			30	0.034	0.035
			45	-0.030	0.069
	MNAR	0	15	-0.726	-0.727
			30	-1.114	-1.067
			45	-0.848	-0.899
		0.4	15	-0.784	-0.762
			30	-1.180	-1.230
			45	-0.908	-0.948

Note. Absolute values below |5%| are considered acceptable for relative bias for parameters.

Table 6

Relative percentage bias of γ_{100} with two feeder schools

Level-1 correlation	Type of missingness	Residual correlation	Missing percentage	Level-2 sample size	
				30	50
0.1	No Missing	0	0	0.164	-0.021
		0.4	0	0.035	0.143
	MCAR	0	15	-0.113	0.073
			30	0.025	0.006
			45	-0.070	0.062
		0.4	15	0.195	-0.005
			30	-0.120	-0.133
			45	-0.140	-0.009
	MAR	0	15	0.046	-0.224
			30	-0.191	-0.019
			45	-0.017	-0.143
		0.4	15	-0.237	-0.213
			30	-0.137	-0.265
			45	-0.280	-0.140
	MNAR	0	15	2.470	2.460
			30	3.521	3.611
			45	2.560	2.609
		0.4	15	2.678	2.664
			30	3.840	3.871
			45	2.746	2.763
0.3	No Missing	0	0	-0.014	0.007
		0.4	0	0.070	-0.039
	MCAR	0	15	-0.006	0.046
			30	-0.214	0.130
			45	0.084	-0.148
		0.4	15	0.064	-0.087
			30	-0.165	-0.220
			45	-0.070	-0.117
	MAR	0	15	-0.357	0.049
			30	-0.160	0.049
			45	-0.115	-0.256
		0.4	15	-0.095	-0.197
			30	-0.041	-0.001
			45	-0.016	-0.172
	MNAR	0	15	2.497	2.480
			30	3.595	3.605
			45	2.477	2.549
		0.4	15	2.563	2.665
			30	3.844	3.899
			45	2.707	2.727

Note. Absolute values |5%| are considered acceptable for relative bias for parameters.

Table 7

Relative percentage bias of γ_{100} with three feeder schools

Level-1 correlation	Type of missingness	Residual correlation	Missing percentage	Level-2 sample size	
				30	50
0.1	No Missing	0	0	0.006	0.369
		0.4	0	0.024	-0.013
	MCAR	0	15	-0.073	-0.130
			30	-0.098	0.020
			45	0.011	-0.094
		0.4	15	-0.072	-0.203
			30	-0.355	0.031
			45	-0.158	-0.059
	MAR	0	15	-0.052	-0.217
			30	0.007	-0.076
			45	-0.235	-0.185
		0.4	15	-0.118	-0.136
			30	-0.003	-0.261
			45	-0.347	-0.161
	MNAR	0	15	2.458	2.504
			30	3.639	3.619
			45	2.563	2.567
		0.4	15	2.705	2.653
			30	3.800	3.830
			45	2.734	2.740
0.3	No Missing	0	0	-0.033	-0.017
		0.4	0	-0.014	-0.007
	MCAR	0	15	-0.024	0.139
			30	-0.217	-0.257
			45	0.249	-0.193
		0.4	15	0.137	-0.046
			30	-0.108	-0.040
			45	0.162	-0.172
	MAR	0	15	0.135	-0.087
			30	-0.194	-0.251
			45	-0.264	-0.098
		0.4	15	-0.058	-0.210
			30	-0.208	-0.104
			45	0.048	-0.260
	MNAR	0	15	2.496	2.520
			30	3.513	3.583
			45	2.583	2.615
		0.4	15	2.675	2.662
			30	3.696	3.873
			45	2.634	2.743

Note. Absolute values |5%| are considered acceptable for relative bias for parameters.

Table 8

Relative percentage bias of γ_{010} with two feeder schools

Level-1 correlation	Type of missingness	Residual correlation	Missing percentage	Level-2 sample size 30	50
0.1	No Missing	0	0	0.206	-0.106
		0.4	0	0.072	0.057
	MCAR	0	15	0.007	-0.051
			30	0.094	-0.188
			45	-0.379	-0.015
		0.4	15	-0.014	0.110
			30	0.107	-0.366
			45	-0.241	-0.226
	MAR	0	15	0.127	0.277
			30	-0.175	-0.090
			45	0.366	-0.048
		0.4	15	-0.200	-0.303
			30	-0.086	0.109
			45	0.002	0.075
	MNAR	0	15	-0.063	-0.031
			30	0.010	0.053
			45	-0.039	0.085
		0.4	15	0.042	-0.022
			30	-0.024	0.098
			45	-0.075	0.024
0.3	No Missing	0	0	0.042	0.159
		0.4	0	-0.222	0.028
	MCAR	0	15	-0.119	-0.216
			30	-0.055	-0.177
			45	0.044	-0.140
		0.4	15	-0.120	-0.121
			30	0.110	0.158
			45	-0.561	0.026
	MAR	0	15	-0.096	0.012
			30	-0.167	-0.091
			45	0.072	-0.118
		0.4	15	0.133	-0.023
			30	-0.026	0.095
			45	-0.122	0.091
	MNAR	0	15	0.110	-0.025
			30	0.019	0.051
			45	-0.015	0.030
		0.4	15	0.170	0.035
			30	-0.006	0.185
			45	-0.026	0.002

Note. Absolute values below |5%| are considered acceptable for relative bias for parameters.

Table 9

Relative percentage bias of γ_{010} with three feeder schools

Level-1 correlation	Type of missingness	Residual correlation	Missing percentage	Level-2 sample size	
				30	50
0.1	No Missing	0	0	0.132	0.008
		0.4	0	0.068	0.014
	MCAR	0	15	-0.233	-0.035
			30	-0.172	0.270
			45	0.051	0.098
		0.4	15	-0.174	0.018
			30	-0.182	0.125
			45	-0.184	0.109
	MAR	0	15	0.088	-0.052
			30	0.012	-0.130
			45	-0.216	-0.086
		0.4	15	0.068	-0.180
			30	0.036	0.103
			45	-0.092	0.184
	MNAR	0	15	0.052	-0.012
			30	-0.025	0.046
			45	-0.067	0.128
		0.4	15	0.006	-0.036
			30	-0.122	-0.146
			45	-0.067	0.070
0.3	No Missing	0	0	0.033	0.002
		0.4	0	-0.010	-0.005
	MCAR	0	15	0.054	0.130
			30	-0.014	-0.132
			45	0.093	0.080
		0.4	15	0.358	-0.090
			30	0.166	0.391
			45	0.090	0.130
	MAR	0	15	0.030	-0.143
			30	0.269	0.055
			45	0.037	0.026
		0.4	15	-0.168	-0.164
			30	-0.071	-0.038
			45	0.011	-0.030
	MNAR	0	15	-0.010	-0.029
			30	0.021	-0.139
			45	-0.067	-0.027
		0.4	15	-0.016	-0.074
			30	0.052	0.101
			45	-0.027	0.048

Note. Absolute values below |5%| are considered acceptable for relative bias for parameters.

Table 10

Relative percentage bias of γ_{020} with two feeder schools

Level-1 correlation	Type of missingness	Residual correlation	Missing percentage	Level-2 sample size 30	50
0.1	No Missing	0	0	0.090	0.084
		0.4	0	0.246	-0.003
	MCAR	0	15	-0.027	0.103
			30	0.030	0.083
			45	0.207	-0.130
		0.4	15	0.080	-0.080
			30	-0.014	0.006
			45	-0.193	0.065
	MAR	0	15	-0.113	0.165
			30	-0.112	0.195
			45	-0.164	0.223
		0.4	15	0.114	0.092
			30	0.302	0.008
			45	0.082	0.178
	MNAR	0	15	-0.037	-0.040
			30	0.046	0.010
			45	-0.040	0.074
		0.4	15	0.044	-0.003
			30	0.073	0.000
			45	-0.024	0.032
0.3	No Missing	0	0	0.154	-0.010
		0.4	0	0.074	0.089
	MCAR	0	15	0.053	-0.114
			30	-0.214	0.112
			45	-0.080	-0.018
		0.4	15	0.131	-0.063
			30	-0.111	-0.053
			45	0.205	0.016
	MAR	0	15	-0.010	0.002
			30	-0.093	-0.087
			45	0.161	-0.167
		0.4	15	0.332	0.165
			30	0.014	0.028
			45	-0.149	0.056
	MNAR	0	15	0.037	-0.006
			30	0.058	0.002
			45	-0.052	-0.046
		0.4	15	0.010	0.005
			30	0.044	-0.002
			45	0.029	0.101

Note. Absolute values |5%| are considered acceptable for relative bias for parameters.

Table 11

Relative percentage bias of γ_{020} with three feeder schools

Level-1 correlation	Type of missingness	Residual correlation	Missing percentage	Level-2 sample size	
				30	50
0.1	No Missing	0	0	-0.144	0.019
		0.4	0	-0.113	-0.037
	MCAR	0	15	-0.163	-0.075
			30	0.149	0.068
			45	-0.190	-0.207
		0.4	15	0.167	0.035
			30	-0.099	0.030
			45	-0.165	0.183
	MAR	0	15	0.244	-0.096
			30	0.158	0.021
			45	0.068	0.091
		0.4	15	0.263	-0.140
			30	0.040	0.125
			45	0.061	0.053
	MNAR	0	15	0.037	0.004
			30	0.004	-0.021
			45	-0.017	-0.080
		0.4	15	0.001	0.054
			30	-0.052	0.099
			45	0.037	-0.078
0.3	No Missing	0	0	-0.088	-0.022
		0.4	0	-0.031	0.160
	MCAR	0	15	0.122	-0.206
			30	-0.154	-0.131
			45	0.135	0.176
		0.4	15	-0.247	0.196
			30	-0.103	0.345
			45	-0.003	0.086
	MAR	0	15	0.076	-0.041
			30	0.344	-0.029
			45	0.166	-0.177
		0.4	15	-0.046	0.199
			30	0.122	-0.004
			45	0.058	0.000
	MNAR	0	15	-0.024	-0.023
			30	0.067	-0.047
			45	-0.080	0.018
		0.4	15	0.005	-0.023
			30	0.037	-0.023
			45	-0.006	-0.068

Note. Absolute values |5%| are considered acceptable for relative bias for parameters.

Table 12

Relative standard error bias of γ_{000} with two feeder schools

Level-1 correlation	Type of missingness	Residual correlation	Missing percentage	Level-2 sample size	
				30	50
0.1	No Missing	0	0	1.870	-1.473
		0.4	0	0.985	-0.063
	MCAR	0	15	-1.686	-0.307
			30	-3.649	2.094
			45	-3.040	-7.181
		0.4	15	-2.394	-2.202
			30	3.052	-4.880
			45	-4.446	-7.420
	MAR	0	15	-0.856	-0.697
			30	-3.598	-3.728
			45	1.187	8.820
		0.4	15	-4.505	-7.143
			30	-2.730	-3.232
			45	-2.819	7.978
	MNAR	0	15	**-15.370**	**-10.508**
			30	**-18.228**	**-18.398**
			45	**-24.895**	**-27.978**
		0.4	15	**-10.094**	**-13.728**
			30	**-19.183**	**-16.616**
			45	**-30.673**	**-26.904**
0.3	No Missing	0	0	1.623	0.836
		0.4	0	0.296	0.488
	MCAR	0	15	-3.741	-1.083
			30	-2.978	-3.239
			45	-1.533	-5.530
		0.4	15	-2.131	-0.601
			30	-7.877	-6.677
			45	-4.260	-5.320
	MAR	0	15	-3.560	-1.852
			30	-3.311	-5.295
			45	-5.714	-1.325
		0.4	15	-1.957	-5.156
			30	-6.088	-2.521
			45	-5.874	9.004
	MNAR	0	15	-6.914	-8.046
			30	**-18.798**	**-21.982**
			45	**-24.108**	**-22.282**
		0.4	15	**-13.892**	-9.911
			30	**-23.429**	**-19.487**
			45	**-28.766**	**-28.917**

Note. Values below |10%| are considered acceptable for standard error bias.

Table 13

Relative standard error bias of γ_{000} with three feeder schools

Level-1 correlation	Type of missingness	Residual correlation	Missing percentage	Level-2 sample size	
				30	50
0.1	No Missing	0	0	-1.741	-2.777
		0.4	0	-2.060	-0.076
	MCAR	0	15	-4.274	-2.435
			30	-4.133	-2.732
			45	-2.660	-4.235
		0.4	15	-3.752	-3.451
			30	-3.408	-7.855
			45	-6.855	-6.480
	MAR	0	15	0.508	-1.091
			30	-4.951	-2.752
			45	-0.662	-1.870
		0.4	15	0.003	-4.661
			30	-4.072	-4.341
			45	-3.079	-1.748
	MNAR	0	15	-5.922	**-12.992**
			30	**-17.381**	**-16.184**
			45	**-24.206**	**-23.654**
		0.4	15	**-14.540**	**-10.959**
			30	**-13.437**	**-17.306**
			45	**-23.582**	**-26.121**
0.3	No Missing	0	0	-0.228	2.653
		0.4	0	-0.439	-1.902
	MCAR	0	15	-1.680	-1.222
			30	-1.336	-4.520
			45	-7.427	-3.835
		0.4	15	0.003	-5.977
			30	-4.036	-7.162
			45	-3.029	-7.743
	MAR	0	15	1.605	-3.643
			30	-2.934	-6.908
			45	-3.213	-6.124
		0.4	15	-2.882	-5.485
			30	-5.685	-1.980
			45	2.331	-7.147
	MNAR	0	15	-8.688	-7.550
			30	**-18.470**	**-13.243**
			45	**-23.792**	**-22.932**
		0.4	15	-9.266	**-11.832**
			30	**-16.643**	**-19.302**
			45	**-27.586**	**-27.406**

Note. Values below |10%| are considered acceptable for standard error bias.

Table 14

Relative standard error bias of γ_{100} with two feeder schools

Level-1 correlation	Type of missingness	Residual correlation	Missing percentage	Level-2 sample size	
				30	50
0.1	No Missing	0	0	2.563	1.703
		0.4	0	0.303	1.133
	MCAR	0	15	0.343	4.413
			30	0.320	1.954
			45	1.520	-2.120
		0.4	15	2.197	-1.465
			30	2.546	1.557
			45	2.246	-3.142
	MAR	0	15	-1.116	0.077
			30	-1.157	-1.402
			45	1.604	4.760
		0.4	15	1.537	7.235
			30	-0.645	0.373
			45	2.631	-2.821
	MNAR	0	15	-1.323	2.768
			30	2.408	-0.112
			45	-0.245	0.544
		0.4	15	0.615	2.279
			30	0.582	-0.325
			45	2.057	-0.586
0.3	No Missing	0	0	1.174	1.370
		0.4	0	-1.321	-0.433
	MCAR	0	15	-0.180	4.277
			30	2.987	-1.769
			45	-1.226	-2.171
		0.4	15	0.253	-3.435
			30	-0.558	-0.933
			45	-1.133	-2.313
	MAR	0	15	-1.287	0.806
			30	0.217	2.465
			45	0.984	-2.060
		0.4	15	0.928	-4.892
			30	0.399	-1.384
			45	-3.168	-1.820
	MNAR	0	15	6.238	5.264
			30	-1.839	0.375
			45	-1.926	-3.196
		0.4	15	1.613	2.420
			30	2.842	0.162
			45	-6.480	-3.726

Note. Values below |10%| are considered acceptable for standard error bias.

Table 15

Relative standard error bias of γ_{100} with three feeder schools

Level-1 correlation	Type of missingness	Residual correlation	Missing percentage	Level-2 sample size 30	50
0.1	No Missing	0	0	1.036	-1.461
		0.4	0	3.268	-0.685
	MCAR	0	15	0.719	-2.333
			30	-0.976	-2.128
			45	0.162	1.028
		0.4	15	-1.466	-1.781
			30	-1.974	0.480
			45	-0.220	0.051
	MAR	0	15	-1.372	-1.479
			30	-0.117	-1.693
			45	-2.532	3.677
		0.4	15	1.148	-2.940
			30	1.344	-1.356
			45	-0.631	-0.815
	MNAR	0	15	5.255	1.757
			30	0.671	-1.811
			45	1.375	-3.966
		0.4	15	2.873	3.686
			30	2.013	-2.560
			45	-0.917	-3.744
0.3	No Missing	0	0	-3.727	0.213
		0.4	0	-0.941	1.154
	MCAR	0	15	-2.669	-1.349
			30	1.021	0.253
			45	-4.414	-2.480
		0.4	15	0.291	-0.432
			30	0.716	0.116
			45	-0.954	0.380
	MAR	0	15	-1.833	-1.281
			30	0.249	0.171
			45	4.457	0.257
		0.4	15	1.128	-3.668
			30	-0.540	0.283
			45	4.799	0.640
	MNAR	0	15	-2.171	1.130
			30	-4.371	1.916
			45	1.037	-2.038
		0.4	15	0.905	0.464
			30	0.779	-0.786
			45	-4.504	0.390

Note. Values below |10%| are considered acceptable for standard error bias.

Table 16

Relative standard error bias of γ_{010} with two feeder schools

Level-1 correlation	Type of missingness	Residual correlation	Missing percentage	Level-2 sample size 30	Level-2 sample size 50
0.1	No Missing	0	0	-1.563	-1.087
		0.4	0	-1.260	-1.676
	MCAR	0	15	-5.398	-7.892
			30	-3.183	-4.728
			45	-6.086	-6.507
		0.4	15	-3.402	-5.322
			30	-6.258	-6.859
			45	-4.607	-9.208
	MAR	0	15	-6.517	-0.948
			30	-5.931	-4.248
			45	-3.601	7.175
		0.4	15	-2.633	-7.703
			30	-8.120	-4.033
			45	-8.453	7.179
	MNAR	0	15	**-15.622**	**-15.374**
			30	**-23.897**	**-24.853**
			45	**-32.952**	**-37.048**
		0.4	15	**-16.068**	**-14.780**
			30	**-23.871**	**-22.740**
			45	**-35.774**	**-32.808**
0.3	No Missing	0	0	0.484	-2.031
		0.4	0	-1.210	-2.356
	MCAR	0	15	-1.632	-3.084
			30	-5.133	-4.371
			45	-3.309	-3.947
		0.4	15	-0.682	-3.234
			30	-3.010	-6.062
			45	-6.027	**-12.021**
	MAR	0	15	-3.828	-6.408
			30	-7.887	-6.293
			45	-2.531	-3.624
		0.4	15	-7.753	-4.643
			30	-8.729	-4.991
			45	-2.492	5.595
	MNAR	0	15	**-12.256**	-9.415
			30	**-22.307**	**-24.760**
			45	**-32.676**	**-30.305**
		0.4	15	**-13.026**	**-12.200**
			30	**-22.301**	**-24.148**
			45	**-33.584**	**-33.001**

Note. Values below |10%| are considered acceptable for standard error bias.

Table 17

Relative standard error bias of γ_{010} with three feeder schools

Level-1 correlation	Type of missingness	Residual correlation	Missing percentage	Level-2 sample size 30	50
0.1	No Missing	0	0	-2.098	-0.074
		0.4	0	-0.319	-0.049
	MCAR	0	15	-5.502	-3.155
			30	-5.126	-3.724
			45	-7.290	-4.261
		0.4	15	-1.138	-3.997
			30	-5.319	-3.437
			45	-8.332	**-10.038**
	MAR	0	15	-3.552	-4.675
			30	-6.153	-3.331
			45	-6.364	-2.752
		0.4	15	-3.210	-5.371
			30	-5.427	-7.626
			45	-6.606	-8.876
	MNAR	0	15	**-13.334**	**-16.301**
			30	**-20.581**	**-22.497**
			45	**-32.653**	**-33.216**
		0.4	15	**-18.406**	**-13.778**
			30	**-23.806**	**-23.193**
			45	**-33.188**	**-30.634**
0.3	No Missing	0	0	-0.567	1.446
		0.4	0	-1.059	-0.700
	MCAR	0	15	-5.175	-6.468
			30	-2.916	-7.503
			45	-7.223	-3.418
		0.4	15	-5.484	-7.469
			30	-8.867	-6.007
			45	-6.701	**-10.394**
	MAR	0	15	-5.017	-6.978
			30	-6.536	-9.240
			45	-6.306	-5.126
		0.4	15	-7.809	-4.784
			30	-5.001	-4.682
			45	-5.579	-4.505
	MNAR	0	15	**-14.716**	-8.464
			30	**-23.646**	**-20.929**
			45	**-30.948**	**-34.245**
		0.4	15	**-13.696**	**-13.468**
			30	**-24.174**	**-24.450**
			45	**-31.799**	**-35.199**

Note. Values below |10%| are considered acceptable for standard error bias.

Table 18

Relative standard error bias of γ_{020} with two feeder schools

Level-1 correlation	Type of missingness	Residual correlation	Missing percentage	Level-2 sample size	
				30	50
0.1	No Missing	0	0	1.855	-1.201
		0.4	0	0.208	-2.931
	MCAR	0	15	3.467	-0.484
			30	-2.812	1.483
			45	-7.314	**-10.697**
		0.4	15	-4.776	-4.871
			30	0.664	-1.134
			45	-3.800	-3.714
	MAR	0	15	0.336	2.479
			30	-1.995	-5.992
			45	-3.090	8.133
		0.4	15	-4.023	-8.619
			30	-7.624	-2.058
			45	-6.736	4.221
	MNAR	0	15	**-14.893**	**-11.505**
			30	**-22.250**	**-21.218**
			45	**-28.673**	**-27.171**
		0.4	15	**-16.114**	**-16.792**
			30	**-24.132**	**-22.211**
			45	**-34.211**	**-32.105**
0.3	No Missing	0	0	0.109	2.433
		0.4	0	-1.030	-1.211
	MCAR	0	15	-3.728	0.609
			30	-4.290	-2.991
			45	-2.949	-4.504
		0.4	15	-1.171	-3.026
			30	-5.675	-8.955
			45	-6.167	-5.590
	MAR	0	15	-3.311	-3.247
			30	-3.040	-6.748
			45	-5.742	-0.869
		0.4	15	-5.345	-2.050
			30	-9.660	-2.310
			45	-6.335	8.204
	MNAR	0	15	**-12.181**	-7.025
			30	**-23.416**	**-19.137**
			45	**-29.309**	**-29.686**
		0.4	15	**-15.991**	**-14.801**
			30	**-25.627**	**-22.044**
			45	**-33.896**	**-34.094**

Note. Values below |10%| are considered acceptable for standard error bias.

Table 19

Relative standard error bias of γ_{020} with three feeder schools

Level-1 correlation	Type of missingness	Residual correlation	Missing percentage	Level-2 sample size	
				30	50
0.1	No Missing	0	0	-2.098	-0.074
		0.4	0	-0.319	-0.049
	MCAR	0	15	-5.548	-5.167
			30	-5.704	-2.417
			45	-5.895	-3.160
		0.4	15	-4.430	-2.376
			30	-6.443	-5.725
			45	-8.782	**-10.113**
	MAR	0	15	-3.115	-3.069
			30	-7.515	-3.930
			45	-0.212	-4.993
		0.4	15	-2.315	-1.764
			30	-6.896	-5.334
			45	-5.064	-3.264
	MNAR	0	15	**-11.985**	**-11.640**
			30	**-23.217**	**-20.234**
			45	**-26.568**	**-26.490**
		0.4	15	**-13.801**	**-13.554**
			30	**-21.518**	**-19.410**
			45	**-30.366**	**-30.187**
0.3	No Missing	0	0	0.567	1.446
		0.4	0	-1.059	-0.700
	MCAR	0	15	-2.991	-1.809
			30	-5.574	-3.001
			45	-8.000	-4.150
		0.4	15	-2.540	-4.569
			30	-6.313	-5.260
			45	-7.283	-8.627
	MAR	0	15	-0.300	-2.051
			30	-6.368	-9.700
			45	-4.473	-5.377
		0.4	15	-5.187	-1.417
			30	-5.176	-3.809
			45	-1.658	-8.199
	MNAR	0	15	**-11.734**	-9.778
			30	**-19.455**	**-18.757**
			45	**-25.125**	**-25.015**
		0.4	15	**-11.011**	**-12.201**
			30	**-20.331**	**-20.057**
			45	**-32.713**	**-30.419**

Note. Values below |10%| are considered acceptable for standard error bias.

9.0% (see Tables 12-13). However, when the data were MNAR, most cells exhibited substantial negative bias. For those cells, the bias ranged from -30.7% to -5.9%. There was no effect of sample size, level-1 correlation, and residual correlation on these cells. This bias; however, increased gradually as a function of the percentage of missing data. For the three levels of missing data (15%, 30% and 45%), the bias had a mean of -10.6%, -18.0%, and -25.8%, and a standard deviation of 2.8, 2.6 and 2.5, respectively, across all MNAR conditions. Across all MNAR conditions with two feeder schools, the mean bias was -19.1% (*SD* = 7.1), and this mean bias was only slightly smaller with three feeder schools, -17.2% (*SD* = 6.5). Additionally, the mean bias was -18.2% (*SD* = 6.4) for MNAR cells with a level-1 correlation of 0.1 and -18.1% (*SD* = 7.2) with a level-1 correlation of 0.3. When there was no residual correlation, the mean bias was -17.1% with a standard deviation of 6.6. In cells with a residual correlation (0.4), the mean of the bias was slightly higher at -19.1% (*SD* = 6.9).

In addition, there were a few cells with bias less than 10 percent in magnitude with MNAR data. Specifically, this was seen only in cells where 15% of data were missing and were more likely to occur when the residual correlation was 0 between the two cross-classified factors. In these cells, the negative bias ranged from 9.9% to 5.9%. All results are presented in Tables 12 and 13.

Relative standard error bias of γ_{100}. For the standard error of γ_{100} (the effect of the student level predictor), the magnitude of the bias for the baseline conditions was always within 10%, ranging from -3.7% to 3.3% (see Tables 14-15). For the estimates in cells with missing data, no substantial bias was found even when data were MNAR. In all cells with missing data, the bias ranged from – 6.5% to 7.2% (see Tables 14-15).

Relative standard error bias of γ_{010}. The relative bias of the standard error estimates of γ_{010} (the effect of the middle school predictor) in the baseline conditions (no missingness) was generally small in magnitude and was always considered acceptable according to Hoogland and Boomsma's criterion. The standard error bias in these cells ranged from -2.3% to 1.4%.

In cells with missing data, the standard error bias of γ_{010} was unacceptable under several conditions. When data were MCAR, only three cells had bias greater than 10% in magnitude and this bias was just slightly larger than 10% (never larger than 12.0% in magnitude). This bias occurred with a level-2 sample size of 50, a residual correlation of 0.4, and 45% of the level-1 predictor values missing. In addition, for the MCAR data, the cells with no substantial bias ranged from -9.2 to -0.1% (see Tables 16-17).

The bias in the standard errors of the MAR conditions was typically negative and generally modest in magnitude. There was no substantial bias in all cells, with the bias ranging from -9.2% to 7.1%. However, with the MNAR data, all conditions resulted in substantial bias, except two cells. In these two cells, the bias was just below 10% in magnitude and occurred with 15% of level-1 predictor values missing, a sample size of 50, and a residual correlation of 0.4. For all cells with MNAR data, the bias ranged between -37.04% and -8.5% (see Tables 16-17). In these cells, the bias magnitude increased as a function of the percentage of missing data with a mean of -13.8% (*SD* = 2.5) in cells with 15% of data missing, -23.2% (*SD* = 1.2) in cells with 30% of data missing, and -33.1% (*SD* = 1.8) with 45% of data missing. There was a slight difference in the mean for MNAR cells with a level-1 correlation of 0.1 (*M* = -24.1%, *SD* = 7.7) versus the mean in cells with a correlation of 0.3 (*M* = -22.7%, *SD* = 8.7). The means and standard deviations were otherwise quite comparable with two feeders (*M* = -23.5%, *SD* = 8.5) and with three feeders (*M* = -23.2%, *SD* = 8.1) and then with no residual correlation (*M* =-23.0%, *SD*= 8.5) versus with a residual correlation (*M* = -23.7%, *SD* = 7.9).

Relative standard error bias of γ_{020}. The standard error bias magnitude of γ_{020} (the effect of the high school predictor) in cells with no missingness never exceeded 10%. This bias ranged from -2.9% to 2.4% (see Tables 18-19).

In cells with missing data, there was substantial bias in two cells. In particular, when the data were MCAR, the level-2 sample size was 50, 45% of the level-1 predictor values were missing, and the level-1 correlation was 0.1, the bias magnitude that was above 10% ranged from -10.6% to -10.1%. In addition, for the rest of the cells (those below 10% in magnitude), the bias of the standard errors when data were MCAR was typically negative and generally modest in magnitude. For these cells, the bias ranged from -8.9% to 3.4%.

When the data were MAR, no substantial bias was found in any cells. In these cells, the bias ranged from -9.7% to 8.2%. Finally, similar to the standard errors of other parameters, there was substantial bias in all cells when the data were MNAR except in two cells. In those cells, the level-1 sample size was 50, the level-1 correlation was 0.3, the residual correlation was 0, and 15% of the level-1 predictor values were missing. In cells with substantial bias, the bias magnitude ranged between -34.2% and -11.01%. Again, similar to other parameters, the bias increased gradually as a function of the percentage of missing data (see Tables 18-19). For the three levels of missing data (15%, 30% and 45%), the bias magnitude had means of -12.8% (*SD* = 2.6), -21.4% (*SD* = 1.9), and -29.8% (*SD* = 3.1), respectively across all MNAR cells. There was mean bias of -22.4% (*SD* = 7.7), and -20.2% (*SD* = 7.1), respectively, across all MNAR cells for each of the feeder conditions (2 or 3). Furthermore, in cells with a level-1 correlations of 0.1 the mean bias was -21.6% (*SD* = 6.8) and -20.9% (*SD* = 8.1) for MNAR cells with a level-1 correlation of 0.3. Lastly, across all cells with no residual correlation, the bias had a mean of -19.8% (*SD* = 6.9), and a mean of -22.8% (*SD* = 7.8) with a correlation of 0.4.

CHAPTER 5

DISCUSSION

The purpose of this study was to evaluate the use of multiple imputation under three different missing data patterns for cross-classified random effects modeling. In addition, the effects of a correlated auxiliary variable were examined for MI's performance along with varying the number of feeder schools, percent of missing data, and correlation among the level-2 residuals. Based on the literature review, in which it was found that not much work has been done within the framework of missing data and cross classified models, it was hoped that this simulation study would provide some preliminary guidelines for researchers using CCREM with missing data.

The results showed that in general, MI met Hoogland and Boomsma's (1998) relative bias estimation criteria (less than 5% in magnitude) for parameter estimates under different types of missing data patterns. While the bias magnitude was never greater than 5%, the bias was slightly higher for MNAR cells than for cells with MCAR or MAR data only on the parameter associated with the student level predictor.

For the standard error estimates, substantial relative bias (defined by Hoogland and Boomsma as greater than 10%) was found in several conditions. While the majority of these cells occurred when the data were MNAR, there were several cells with MCAR data in which substantial bias (although always just slightly higher than 10% in magnitude) was found. These cells only occurred when 45% of the values on the level-1 predictor were missing. In other words, the high percentage of missingness tended to result in data sets more deviant from the original complete data sets and thus generate larger bias in the standard errors. In cells with MNAR data, substantial bias was consistently found for the standard errors of all parameters except for γ_{100} (the effect of the level-1 predictor). For this parameter, substantial standard error bias was never seen. Since the level-1 predictor was the variable with the missing values, it is interesting that its standard error did not show substantial bias. Instead, its parameter bias was slightly larger than that seen for the other parameters (but still below 5%), but otherwise the effect of those missing data all appeared in the relative bias of the standard errors for the other model parameters. For the standard errors of the other parameters, the bias was negative and increased in magnitude as a function of the percent of missing data. No study factor otherwise appeared to impact the bias magnitude (the number of feeder schools, the level-1 correlation, and the residuals correlation).

When comparing the results from cells with no missing data versus those with missing data, MI worked well with cross-classified data unless the data were MNAR. Future researchers should feel confident applying MI with a level-1 missing data of 30% with MCAR and MAR data. Since 45% of missing level-1 predictor values produced substantial standard error bias in some conditions, MI should be applied cautiously to cross-classified data with more than 30% of the data missing. This result also supports the research conducted by Zhang (2005) in which he examined multilevel modeling and multilevel structural equation modeling with missing data. In particular, he investigated the influence of non-normality and performance of multiple imputation utilizing the expectation maximization (EM) algorithm. He found that a higher proportion of missing data tended to produce more bias in standard error estimates. Applied cross-classified research in education may benefit from this information as it is frequently the case that the missingness in a level-1 predictor is 30% or less.

When data were MNAR, there were very few cells without substantial bias. When no substantial bias occurred the degree of missingness was always 15% thus emphasizing that the effect of MNAR data is increasingly worse for higher degrees of missing data. MNAR creates a situation that is difficult to handle in a study. Because the missingness of the values is related to the variable itself, there is usually no information to access the missing observations.

One of the benefits of MI is the potential use of an auxiliary variable to improve the quality of the imputed data. Unexpectedly in this study, however, the degree of correlation between the auxiliary variable and the level-1 predictor (the variable with missingness) did not impact the results for MI. While it was assumed that the greater the correlation between the auxiliary variable and the level-1 predictor, the better the performance of MI. However, this finding was not seen. It is unknown as to why this correlation did not impact the findings. This result should be investigated further to determine if there are other scenarios where the auxiliary variable might improve the performance for MI with cross-classified data.

Limitations and Directions for Future Research

Although this study has findings that add to the body of literature in the area of missing data with cross-classified data, there are limitations in the study design and study conditions which in turn support future research. The use of different sample sizes at level-2, different auxiliary variable correlations with the level-1 variable, different level-2 residual correlations, and varying percentages of missing data were used to mirror closely conditions that are present during real data situations and in the literature of simulation studies. The generalizability of the simulated results, however, is restricted to those conditions studied. This section discusses the limitations and directions for future research.

The first limitation of this study was the CCREM model utilized. The model described in this study is a simple model in which not all possible random effects were included. For instance, the effect of a variable (here, size) for one of the cross-classified factors (middle school) could vary across levels of the other cross-classified factor (high school). Future research should include models in which these random effects are included and missing data are present.

The second limitation of this study occurred through the way multiple imputation was utilized. There are several ways of imputing the data when using multiple imputation. In this study, the Markov Chain Monte Carlo (MCMC) method (Schafer, 1997) was used. Future research should study different imputation methods such as the regression method and propensity score methods to compare their performance with the MCMC method.

The third limitation of this study was that the missingness only occurred at level-1. In real data situations, however, it is possible that the missingness can also occur at level-2. While this is a limitation of this study, it is quite realistic in that it is common for applied datasets to have little to no missingness for the variables for each of the level-2 cross-classified factors since variables at the level of the school or program are easier to collect than are data at the individual level (indeed, many relevant variables describing a school can now be found on a school's website).

The fourth limitation of the study was that the auxiliary variable did not have an impact on the multiple imputation results. Possibly, the correlations that were included in this study were too small. Future research should explore different levels of auxiliary variable correlations with level-1 predictor.

The last limitation of this study was the manipulated sample sizes. As the present study only studied two different sample sizes at level-2 (and these were constrained to be equal for each of the cross-classified factors), more combinations of sample sizes at level-1 and level-2 should be investigated. In addition, although in this study middle and high school sample sizes were kept equal, unequal sample sizes are also possible in real data situations. Future research should examine MI's performance under varying sample sizes.

Educational Importance and Conclusion

In educational research, it is very possible that students are nested within some type of cross-classification. The enactment of the No Child Left Behind Act (NCLB; Public Law 107-110, 2001) has also led to potential cross-classified data scenarios. Under this act, if a school fails to meet the state standards, parents have the choice of identifying a better school in the school district and transferring their child to that school. This change is problematic for traditional statistical designs, especially in longitudinal designs. The use of CCREM can help remedy that scenario and allow for students who have changed schools to have that cross-classification modeled.

This research also could contribute to educational evaluation and policy analysis; in particular, evaluation of after school programs. Again, under the NCLB Act, many state departments of education apply for after school programs funding from the federal government and then they partition this money out to the local education agencies in their state resulting in a number of programs across the state. The use of program theory-based evaluation enables evaluators to provide feedback about the state-level program planning and implementation as it affects program performance at the school level. In these programs, cross-classification can occur where students from different schools attend the same after school program centers and where students from two different after school programs come from the same school. If researchers are interested in investigating the impact of characteristics of schools on after-school program outcomes, they need to consider the cross-classification of the data.

Moreover, the presence of missing values is very common in datasets analyzed in educational research. The results of this simulation study suggest that MI can be utilized to handle missing data as long as the data are MCAR or MAR. If the percent of missing data is greater than 30%, then MI should be used somewhat cautiously, however. It is hoped that the results from the present study will be helpful to any researcher who is performing CCREM models in which missing data are a problem. More generally, it is hoped that this work will be useful and worthwhile to researchers wishing to gain a better understanding of how to use MI, and how it performs with cross-classified data.

References

Allison, P. D. (2001). *Missing data*. Thousand Oaks, CA: Sage Publications.

Bentler, P. M. (1989). *EQS Structural Equations Program.* Los Angeles, CA: BMPD Statistical Software.

Browne, W. J., Goldstein, H., & Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modeling, 1*, 103-124.

Bryk, A.S., & Raudenbush, S.W. (1988). Toward a more conceptualization of research on school effects: A three-level hierarchical linear model. *American Journal of Education, 97* (1)*, 65-108.*

Clayton, D., & Rasbash, J. (1999). Estimation in large random effects models by data augmentation. *Journal of the Royal Statistics Society, 162*(3), 425-436.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. New Jersey: Lawrence Erlbaum.

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology, 57*, 505-528.

Collins, L. M., Schafer, J. L., & Kam, C.M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*(4), 330-351.

Enders, C. K. (2003). Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. *Psychological Methods, 8*(3)*, 322-337.*

Enders, C. K. (2001). The impact of nonnormality on full information maximum likelihood estimation for structural equation models with missing data. *Psychological Methods, 6,* 352-370.

Garner, C. L., & Raudenbush, S. W. (1991). Neighborhood effects on educational attainment: A multilevel analysis. *Sociology of Education, 64*(4), 251-262.

Gibson, N. M., & Olejnik, S. (2003). Treatment of missing data at the second level of hierarchical linear models. *Education and Psychological Measurement, 63*(2)*,* 204.

Graham, J. W.., & Schafer, J. L. (1999). On the performance of multiple imputation for multilevel data with small sample size. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp.1-29). Thousand Oaks, CA: Sage.

Goldstein, H., & Fielding, A. (2006). *Cross-classified and multiple membership structures in multilevel models: An introduction and Review* (Research Report). London, UK: University of Birmingham University of Bristol.

Goldstein, H., & Sammons, P. (1997). The influence of secondary and junior schools on sixteen year examination performance: A cross-classified multilevel analysis. *School Effectiveness & School Improvement, 8*(2)*,* 219-232.

Goldstein, H. (1995). Hierarchical data modeling in the social sciences. *Journal of Educational and Behavioral Statistics, 20*(2)*,* 201-204.

Goldstein, H. (1987). *Multilevel models in educational and social research.* London: New York:Griffin; Oxford Univeristy Press.

Hagenaars, J. A., & Van de Pol, A. (2002). *Applied latent class analysis (MPLUS).* Cambridge, UK: Cambridge University Press.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods and Research, 26,* 329-367.

Hox, J.J. (2002). *Multilevel analysis: Techniques and applications.* Mahmah, NJ: Lawrence Erlbaum Associates.

Kreft, L. & de Leeuw, J. (1998). *Introducing Multilevel Modeling.* Thousand Oaks, CA: Sage Publications, Inc.

Laird, N. M., & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics, 38*, 963-974.

Lepkowski, J. M., Landis, R.J., & Stehouwer, S.A. (1987). Strategies for the analysis of imputed data in a sample survey. *Medical Care, 16,* 705-715

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika, 74*, 817-827.

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. New York: Guilford Press.

Meyers, J. L. & Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research, 41*(4), 473-497.

Meyers, J. L. (2005). The impact of the inappropriate modeling of cross-classified data structures. *Dissertation Abstracts International, 65, 8A.*.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74*(4), 525-556.

Rasbash, J., & Browne, W. J. (2001). Modeling non-hierarchical structures. In A. H. Leyland & H. Goldstein (Eds.), *Multilevel Modeling of Health Statistics*, 93-105 Chichester: John Wiley & Sons.

Rasbash, J., Browne, W. J., & Goldstein, H. (2000). *A user's guide to MLwin*. Unpublished manuscript, Institude of Education, London.

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

Raudenbush, Bryk, A. S., & Congdon, R. (2005). HLM 6. Chicago: Scientific Software International.

Rubin, D. B, (1976). Inference and missing data. *Biometrika, 63,* 581-592

SAS Institute (2005). *SAS/IML software: Changes and enhancements, through release 9.1.* Cary, NC: SAS Institute.

Schafer, J.L. (1997). *Analysis of incomplete multivariate data* (1st ed.). London; New York: Chapman & Hall.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33*(4), 545.

Schafer, J. L. (1999). NORM (Version 2.03 for Windows) [Computer software]. Retrieved March 1, 2008, from http://www.stat.psu.edu/~jls/misoftwa.html.

Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147-177.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23*(4), 323-355.

Tate, R.L., & Wongbundhit, Y. (1983). Random versus nonrandom coefficient models for multilevel analysis. *Journal of Educational Statistics, 8*(2)*,* 103-120.

SPSS (2006). *Statistical Package for the Social Sciences.* Chicago, IL: SPSS 15.0

Zhang, D. (2005). A Monte Carlo investigation of robustness to nonnormal incomplete data of multilevel modeling. *Dissertation Abstracts International, 67, 8A.*

APPENDIXES

Appendix A

SAS Program for Two Feeder Schools

```
/*This SAS Code is adapted from Meyers, J. (2004) dissertation with an
inclusion of missing data part
/*Generating middle school and high school residuals based on a known
correlation*/

%macro generate;
%let seed1 = 12541;
%let seed2 = 93045;
%let seed3 = 26889;
%let seed4 = 74621;
%let seed5 = 96519;
%let seed6 = 42197;
%let seed7 = 24775;
%let seed8 = 88367;
%let seed9 = 20569;
%let seed10 = 68243;
%let seed11 = 53197;
%let level2n = 30; /*either 30 or 50*/
%let lev1corr = .10; /*either .10 or .30*/
%let rescorr = .0; /*either .4 or .0*/
%let missper=.4;/*either .15. .30 or .40*/

%do i =1 %to 1;/*Identifies the number of replications*/
*PROC PRINTTO LOG='C:\log.TXT' new;
*PROC PRINTTO PRINT='C:\output.TXT' new;

/*CONDITION:Number of levels of the cross-classified factors*/
proc iml;
resid=normal(j(&level2n,2,&seed11));/*change this to 30/50 in other
condition*/
/*CONDITION: CORRELATION OF RESIDUALS*/
/* schools (high & middle the schools only correlated on .40 or .00*/
corr={1.00 &rescorr,
         &rescorr 1.00};
corr_root=root(corr); /*cholesky decomposition*/
fin_resid=resid*corr_root; /*correlating residuals*/
fin_resid2=4#fin_resid;
create sasresid from fin_resid2; /*new variable*/
append from fin_resid2;
quit;

/*out of IML*/
```

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

proc sort data=sasresid; by ms_res; /*sorted by the middle school
residual, in ascending order*/
run;

/*main feeder*middle or high schools*/
data one; set sasresid;
retain counter 0;
counter=counter+1;
ms_id=counter;
hs_id=counter;
ms_size= int(30+2*rannor(&seed1)); /*number of students fixed at 30.
the middle school sample size drawn from a normal distribution with a
known mean
(with values 30), sd=2)*/

cell_size = int(.60*ms_size); /*number of students*/
size= 50+10*rannor(&seed2); /*mean=50, sd=10 */
middleschool_effect= .50*size + ms_res;
free=50+10*rannor(&seed3); /* generating normal distributed data*/
highschool_effect=.50*free + hs_res;
run;

/*secondary feeder*/
data onemore; set one;
keep ms_id hs_id cell_size;
ms_id=counter; /*counter means if it is countable*/
hs_id=counter+1;
cell_size=int (.40*ms_size);  /*max. cell size*/
if cell_size = 0 then cell_size = 1;
if hs_id gt &level2n then hs_id=1;
ms_res=ms_res;
run;

proc sort data=onemore; by ms_id;run; /*sorted by the middle school
residual, in ascending order*/

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;

proc sort data=onemore;by hs_id;

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;

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;

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

/*creating student observations-Level 1*/
data student; set grand;
do student= 1 to cell_size;
x1 = 50+10*rannor(&seed4);
x2 = 50+10*rannor(&seed5);
output; end;


proc iml;
use student; read all VAR {x1 x2}into studmatrix;
corr={1.00 &lev1corr,
         &lev1corr 1.00};
corr_root=root(corr); /*cholesky decomposition*/
corrdata=studmatrix*corr_root; /*correlating variables*/
CNAME={"X1" "X2"}; *names variables;
create corrl1 from corrdata [COLNAME=CNAME]; /*new variable*/
append from corrdata;
quit;

data combine; merge student corrl1;
student_effect=.50*x1 + 4*rannor(&seed6);
testsc=100+student_effect+middleschool_effect+highschool_effect;
iter=&i;
run;
proc corr;
var x1 x2 testsc;

/*MCAR, creating MCAR pattern**********************************/

/* GENERATE uniform random number between 0 and 1 for each X1*/

data combine1;set combine;
mcaruni= ranuni(&seed7);
run;

/*Generate MCAR missingness*/

data MCAR; set combine1;
if mcaruni <= &missper then x1=.;run; /*percentage missingness .15, .30
or .45*/

/*MCAR ends here******************************************************/

/*MAR**********************************************************/
/*Generate MAR missingness */

data combine2; set combine;
```

```
maruni= ranuni (&seed8);run;
proc sort data=combine2;/*sort auxiliary variable, x2, by ascending
order*/
 by x2; run;

proc rank percent out=data1;/*Assign a percentage rank for each x2*/
var x2;
ranks x2P;
run;

/* Assign a deletion probability for each x2(auxiliary variable)*/
data combine3; set work.data1;
x2Por=x2P/100;
X2DP= 1-x2Por;
run;

/*Deletion process will begin w/choosing the lowest x2
until desired portion of missing data reached:
15%,30% or 45%*/
proc iml;
use combine3; read all var{x1 maruni x2dp} into marmat [colname =
vars];
count=0;
permar=0;

      do i=1 to nrow(marmat) until (permar>=&missper);
            if marmat[i,2] < marmat[i,3] then marmat[i,1]=.;
            if marmat[i,1]=. then count=count+1;
            permar=count/nrow(marmat);
      end;
      CNAME={"x1" "maruni" "x2dp"};

CREATE mardata FROM marmat [COLNAME=CNAME];
APPEND FROM marmat;
quit;

data mar; merge combine3 mardata;

/*MAR ends here******************************************************/

/*MNAR***************************************************************/
/*Generate MNAR missingness mnaruni=v*/

data combine4;set combine;
mnaruni= ranuni (&seed9);run;
proc sort data=combine4; /*sort x1 by asecending order*/
by x1;run;

proc rank percent out=data2;/*Assign a percentige rank for each x2*/
var x1;
ranks x1P;
run;

/* Assign a deletion probability for each x1*/
data combine5; set work.data2;
x1Por=x1P/100;
X1DP= 1-x1Por;
```

```
run;

/*Deletion process will begin w/choosing the lowest x1
until desired portion of missing data reached:
15%,30% or 45%*/
/*Deletion process will begin w/choosing the lowest x2
until desired portion of missing data reached:
15%,30% or 45%*/
proc iml;
use combine5; read all var{x1 mnaruni x1dp} into mnarmat [colname =
vars];
count=0;
permar=0;

      do i=1 to nrow(mnarmat) until (permar>=&missper);
            if mnarmat[i,2] < mnarmat[i,3] then mnarmat[i,1]=.;
            if mnarmat[i,1]=. then count=count+1;
            permar=count/nrow(mnarmat);
      end;
      CNAME={"x1" "mnaruni" "x1dp"};

CREATE mnardata FROM mnarmat [COLNAME=CNAME];
APPEND FROM mnarmat;
quit;

data mnar; merge combine5 mnardata;
/*MNAR ends here*******************************************************/

/*Using Listwise deletion for handling missing data*/
data LD; set MCAR; /*change set option to MNAR and MAR for other types
of missingness*/
  if x1=. then delete;
RUN;
proc corr;
var x1 x2 testsc;
title 'listwise';

/* MULTIPLE IMPUTTAION PROCESS, USE AUXILIARY VARIBLE "X2" IN
IMPUTATION STAGE*/
/* condition: data change to MAR or MNAR*/

PROC MI DATA=MCAR OUT=IMPUTE_MCAR NIMPUTE=5 SEED=&seed10; /*change data
option to MAR and MNAR*/
VAR x1 x2 testsc free size;
RUN;
quit;
proc corr;
var x1 x2 testsc; by _imputation_;
title 'mi'; run;

ods output solutionf = fixed_effects;

/* Fitting the cross-classified model for imputed data */
proc mixed data= impute_mcar noclprint covtest noinfo noitprint;
class ms_id hs_id;
model testsc= free size x1/s;
random int/sub= ms_id;
```

```
random int/sub=hs_id;
by _imputation_;
run;

ods output solutionf = fixed_effectsLD;
/* Fitting the cross-classified model for LD data */
proc mixed data= LD noclprint covtest noinfo noitprint;
class ms_id hs_id;
model testsc= free size x1/s;
random int/sub= ms_id;
random int/sub=hs_id;
run;

/*transposing fixed effects for imputed data*/
proc transpose data = fixed_effects out= fixed_trans;
var estimate StdErr tvalue Probt;
id effect;
by _imputation_;
run;

/*renaming fixed effects for imputed data*/
data fixedeff; set fixed_trans; by _imputation_;
retain intercept_est free_est size_est SAT_est  intercept_std free_std
size_std SAT_std  intercept_t free_t size_t SAT_t  intercept_probt
free_probt size_probt SAT_probt _imputation_;
if _NAME_ = 'Estimate' then do;
intercept_est = intercept;
free_est = free;
size_est=size;
SAT_est = x1;
end;
if _NAME_='StdErr' then do;
intercept_std = intercept;
free_std = free;
size_std=size;
SAT_std = x1;
end;
if _NAME_='tValue' then do;
intercept_t = intercept;
free_t = free;
size_t=size;
SAT_t = x1;
end;
if _NAME_='Probt' then do;
intercept_probt = intercept;
free_probt = free;
size_probt=size;
SAT_probt = x1;
end;
if _NAME_='Probt';
keep intercept_est free_est  size_est SAT_est intercept_std free_std
size_std SAT_std intercept_t free_t size_t SAT_t intercept_probt
free_probt size_probt SAT_probt  _imputation_;
if last._imputation_;
run;
```

```
/*LD
PART///////////////////////////////////////////////////////////////////
//////////*/
/*transposing fixed effects for LD data*/
proc transpose data = fixed_effectsLD out= fixed_transLD;
var estimate StdErr tvalue Probt;
id effect;
run;

/*renaming fixed effects for LD data*/
data fixedeffLD; set fixed_transLD;
retain intercept_estLD free_estLD size_estLD SAT_estLD  intercept_stdLD
free_stdLD
size_stdLD SAT_stdLD  intercept_tLD free_tLD size_tLD SAT_tLD
intercept_probtLD
free_probtLD size_probtLD SAT_probtLD;
if _NAME_ = 'Estimate' then do;
intercept_estLD = intercept;
free_estLD = free;
size_estLD=size;
SAT_estLD = x1;
end;
if _NAME_='StdErr' then do;
intercept_stdLD = intercept;
free_stdLD = free;
size_stdLD=size;
SAT_stdLD = x1;
end;
if _NAME_='tValue' then do;
intercept_tLD = intercept;
free_tLD = free;
size_tLD=size;
SAT_tLD = x1;
end;
if _NAME_='Probt' then do;
intercept_probtLD = intercept;
free_probtLD = free;
size_probtLD=size;
SAT_probtLD = x1;
end;
if _NAME_='Probt';
keep intercept_estLD free_estLD size_estLD SAT_estLD  intercept_stdLD
free_stdLD
size_stdLD SAT_stdLD  intercept_tLD free_tLD size_tLD SAT_tLD
intercept_probtLD
free_probtLD size_probtLD SAT_probtLD;
run;

/*combine all imputed data;*/


ods output parameterestimates=miparms;

proc mianalyze data=fixedeff;
modeleffects intercept_est free_est size_est  SAT_est;
stderr  intercept_std free_std size_std SAT_std; run;
```

```
proc transpose data=miparms out=estimates_trans;
var estimate StdErr;
id Parm;
run;

data estall; set estimates_trans;
retain intercept_est1 free_est1  size_est1 SAT_est1 intercept_std
free_std
size_std SAT_std ;

if _NAME_ = 'Estimate' then do;
intercept_est1=intercept_est;
free_est1=free_est;
size_est1=size_est;
SAT_est1=SAT_est;
end;

if _Label_='Std Error' then do;
intercept_std=intercept_est;
free_std=free_est;
size_std=size_est;
SAT_std=SAT_est;
end;

if _Label_='Std Error';

keep intercept_est1 free_est1  size_est1 SAT_est1 intercept_std
free_std
size_std SAT_std ;
run;


data comb_seed; set estall;set fixedeffLD;
seed1=&seed1; seed2=&seed2; seed3=&seed3;seed4=&seed4;seed5=&seed5;
seed6=&seed6;seed7=&seed7;seed8=&seed8;seed9=&seed9;seed10=&seed10;
seed11=&seed11; iter=&i;
level2n=&level2n; lev1corr=&lev1corr; rescorr=&rescorr;
missper=&missper;

proc append base=all_MCAR data=comb_seed; run;/*change MCAR to MAR and
MNAR*/

%LET SEED1=&SEED1+2;
%LET SEED2=&SEED2+2;
%LET SEED3=&SEED3+2;
%LET SEED4=&SEED4+2;
%LET SEED5=&SEED5+2;
%LET SEED6=&SEED6+2;
%LET SEED7=&SEED7+2;
%LET SEED8=&SEED8+2;
%LET SEED9=&SEED9+2;
%LET SEED10=%eval(&SEED10+2);
%LET SEED11=&SEED11+2;
%end;
%mend;
%generate;
```

```
/*Calculating Relative Bias of the Parameter Estimates*/

PROC MEANS DATA= all_MCAR; VAR intercept_est1 free_est1 size_est1
SAT_est1 /*change all_MCAR to All_MAR and all_MNAR*/
intercept_estLD free_estLD size_estLD SAT_estLD
intercept_std free_std size_std SAT_std intercept_stdLD free_stdLD
size_stdLD SAT_stdLD;

OUTPUT OUT=PARM1 MEAN= Sintercept_est1 Sfree_est1 Ssize_est1 SSAT_est1
Sintercept_estLD Sfree_estLD Ssize_estLD SSAT_estLD
intercept_std free_std size_std SAT_std intercept_stdLD free_stdLD
size_stdLD SAT_stdLD;
OUTPUT OUT=SE1 STD= ssintercept_est1 ssfree_est1 sssize_est1 ssSAT_est1
ssintercept_estLD ssfree_estLD sssize_estLD ssSAT_estLD ;


DATA BIAS; SET PARM1;drop _type_ _freq_;
PBIAS1_MI=((Sintercept_est1-100)/100)*100;
PBIAS2_MI= ((Sfree_est1-.5)/.5)*100;
PBIAS3_MI=((Ssize_est1-.5)/.5)*100;
PBIAS4_MI=((SSAT_est1-.5)/.5)*100;
PBIAS5_LD=((Sintercept_estLD-100)/100)*100;
PBIAS6_LD= ((Sfree_estLD-.5)/.5)*100;
PBIAS7_LD=((Ssize_estLD-.5)/.5)*100;
PBIAS8_LD=((SSAT_estLD-.5)/.5)*100;

PROC PRINT; VAR PBIAS1_MI PBIAS2_MI PBIAS3_MI PBIAS4_MI
                        PBIAS5_LD PBIAS6_LD PBIAS7_LD PBIAS8_LD; RUN;

/*Calculating Relative Bias of the Standard Errors*/
DATA SEM1; SET  PARM1;  SET SE1; drop _type_ _freq_;

SBIAS1_MI=(((intercept_std-ssintercept_est1)/ssintercept_est1))*100;
SBIAS2_MI=(((free_std-ssfree_est1)/ssfree_est1))*100;
SBIAS3_MI=(((size_std-sssize_est1)/sssize_est1))*100;
SBIAS4_MI=(((SAT_std-ssSAT_est1)/ssSAT_est1))*100;
SBIAS5_LD=(((intercept_stdLD-
ssintercept_estLD)/ssintercept_estLD))*100;
SBIAS6_LD=(((free_stdLD-ssfree_estLD)/ssfree_estLD))*100;
SBIAS7_LD=(((size_stdLD-sssize_estLD)/sssize_estLD))*100;
SBIAS8_LD=(((SAT_stdLD-ssSAT_estLD)/ssSAT_estLD))*100;
PROC PRINT; VAR SBIAS1_MI SBIAS2_MI SBIAS3_MI SBIAS4_MI
                SBIAS5_LD SBIAS6_LD SBIAS7_LD SBIAS8_LD;  RUN;

data MCAR;set bias;set sem1;run;/*change MCAR to MAR and MNAR*/
```

Appendix B

SAS Program for Three Feeder Schools

```
/*This SAS Code is adapted from Meyers, J. (2004) dissertation with an
inclusion of missing data part
/*Generating middle school and high school residuals based on a known
correlation*/

%macro generate;
%let seed1 = 46191;
%let seed2 = 30121;
%let seed3 = 40579;
%let seed4 = 85197;
%let seed5 = 31035;
%let seed6 = 35163;
%let seed7 = 12493;
%let seed8 = 49569;
%let seed9 = 44889;
%let seed10 = 19455;
%let seed11 = 16981;
%let level2n = 50; /*either 30 or 50*/
%let lev1corr = .10; /*either .10 or .30*/
%let rescorr = .0; /*either .4 or .0*/
%let missper=.15;/*either .15. .30 or .40*/

%do i =1 %to 1000;/*Identifies the number of replications*/
PROC PRINTTO LOG='C:\SUKI.TXT' new;
PROC PRINTTO PRINT='C:\STUFF.TXT' new;

/*CONDITION:Number of levels of the cross-classified factors*/
proc iml;
resid=normal(j(&level2n,2,&seed11));/*change this to 30/50 in other
condition*/
/*CONDITION: CORRELATION OF RESIDUALS*/
/* schools (high & middle the schools only correlated on .40 or .00*/
corr={1.00 &rescorr,
         &rescorr 1.00};
corr_root=root(corr); /*cholesky decomposition*/
fin_resid=resid*corr_root; /*correlating residuals*/
fin_resid2=5#fin_resid;
create sasresid from fin_resid2; /*new variable*/
append from fin_resid2;
quit;

/*out of IML*/
data sasresid; set sasresid;
rename COL1= ms_res;
rename COL2= hs_res;
```

```
run;

proc sort data=sasresid; by ms_res; /*sorted by the middle school
residual, in ascending order*/
run;

/*main feeder*middle or high schools*/
data one; set sasresid;
retain counter 0;
counter=counter+1;
ms_id=counter;
hs_id=counter;
ms_size= int(30+2*rannor(&seed1)); /*number of students fixed at 30.
the middle school sample size drawn from a normal distribution with a
known mean
(with values 30), sd=2)*/

cell_size = int(.60*ms_size); /*number of students*/
size= 50+10*rannor(&seed2); /*mean=50, sd=10 */
middleschool_effect= .50*size + ms_res;
free=50+10*rannor(&seed3); /* generating normal distributed data*/
highschool_effect=.50*free + hs_res;
run;

/*secondary feeder*/
data onemore; set one;
keep ms_id hs_id cell_size;
ms_id=counter; /*counter means if it is countable*/
hs_id=counter+1;
cell_size=int (.15*ms_size);  /*max. cell size*/
if cell_size = 0 then cell_size = 1;
if hs_id gt &level2n then hs_id=1;
ms_res=ms_res;
run;

/*three feeder condition

/*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 &level2n then ms_id=1; /*change to 30 in other condition*/
run;


proc sort data=onemore; by ms_id;run; /*sorted by the middle school
residual, in ascending order*/



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-Level 1*/
data student; set grand;
do student= 1 to cell_size;
x1 = 50+10*rannor(&seed4);
x2 = 50+10*rannor(&seed5);
output; end;


proc iml;
use student; read all VAR {x1 x2}into studmatrix;
corr={1.00 &lev1corr,
         &lev1corr 1.00};
corr_root=root(corr); /*cholesky decomposition*/
corrdata=studmatrix*corr_root; /*correlating variables*/
CNAME={"X1" "X2"}; *names variables;
create corrl1 from corrdata [COLNAME=CNAME]; /*new variable*/
append from corrdata;
quit;
```

```
data combine; merge student corrl1;
student_effect=.50*x1 + 5*rannor(&seed6);
testsc=100+student_effect+middleschool_effect+highschool_effect;
iter=&i;
run;

/*MCAR, creating MCAR pattern**********************************/

/* GENERATE uniform random number between 0 and 1 for each X1*/

data combine1;set combine;
mcaruni= ranuni(&seed7);
run;

/*Generete MCAR missingness*/

data MCAR; set combine1;
if mcaruni <= &missper then x1=.;run; /*percentage missingness .15, .30
or .45*/

/*MCAR ends here******************************************************/

/*MAR*************************************************************/
/*Generate MAR missingness maruni=v*/

data combine2; set combine;
maruni= ranuni (&seed8);run;
proc sort data=combine2;/*sort auxiliary variable, x2, by ascending
order*/
 by x2; run;

proc rank percent out=data1;/*Assign a percentage rank for each x2*/
var x2;
ranks x2P;
run;

/* Assign a deletion probability for each x2(auxiliary variable)*/
data combine3; set work.data1;
x2Por=x2P/100;
X2DP= 1-x2Por;
run;

/*Deletion process will begin w/choosing the lowest x2
until desired portion of missing data reached:
15%,30% or 45%*/
proc iml;
use combine3; read all var{x1 maruni x2dp} into marmat [colname =
vars];
count=0;
permar=0;

      do i=1 to nrow(marmat) until (permar>=&missper);
            if marmat[i,2] < marmat[i,3] then marmat[i,1]=.;
            if marmat[i,1]=. then count=count+1;
            permar=count/nrow(marmat);
      end;
```

```
      CNAME={"x1" "maruni" "x2dp"};

CREATE mardata FROM marmat [COLNAME=CNAME];
APPEND FROM marmat;
quit;

data mar; merge combine3 mardata;

/*MAR ends here*******************************************************/

/*MNAR*****************************************************************/
/*Generate MNAR missingness mnaruni=v*/

data combine4;set combine;
mnaruni= ranuni (&seed9);run;
proc sort data=combine4; /*sort x1 by ascending order*/
by x1;run;

proc rank percent out=data2;/*Assign a percentage rank for each x2*/
var x1;
ranks x1P;
run;

/* Assign a deletion probability for each x1*/
data combine5; set work.data2;
x1Por=x1P/100;
X1DP= 1-x1Por;
run;

/*Deletion process will begin w/choosing the lowest x1
until desired portion of missing data reached:
15%,30% or 45%*/
proc iml;
use combine5; read all var{x1 mnaruni x1dp} into mnarmat [colname =
vars];
count=0;
permar=0;

      do i=1 to nrow(mnarmat) until (permar>=&missper);
            if mnarmat[i,2] < mnarmat[i,3] then mnarmat[i,1]=.;
            if mnarmat[i,1]=. then count=count+1;
            permar=count/nrow(mnarmat);
      end;
      CNAME={"x1" "mnaruni" "x1dp"};

CREATE mnardata FROM mnarmat [COLNAME=CNAME];
APPEND FROM mnarmat;
quit;

data mnar; merge combine5 mnardata;
/*MNAR ends here******************************************************/

/*Using Listwise deletion for handling missing data*/
data LD; set MCAR; /*change set option to MNAR and MAR for other types
of missingness*/
  if x1=. then delete;
RUN;
```

```
proc corr;
var x1 x2 testsc;
title 'listwise';

ods output solutionf = fixed_effects;

/* Fitting the cross-classified model for imputed data */
proc mixed data= impute_mcar noclprint covtest noinfo noitprint;
/*change data option to MAR and MNAR*/
class ms_id hs_id;
model testsc= free size x1/s;
random int/sub= ms_id;
random int/sub=hs_id;
by _imputation_;
run;

ods output solutionf = fixed_effectsLD;
/* Fitting the cross-classified model for LD data */
proc mixed data= LD noclprint covtest noinfo noitprint;
class ms_id hs_id;
model testsc= free size x1/s;
random int/sub= ms_id;
random int/sub=hs_id;
run;

/*transposing fixed effects for imputed data*/
proc transpose data = fixed_effects out= fixed_trans;
var estimate StdErr tvalue Probt;
id effect;
by _imputation_;
run;

/*renaming fixed effects for imputed data*/
data fixedeff; set fixed_trans; by _imputation_;
retain intercept_est free_est size_est SAT_est  intercept_std free_std
size_std SAT_std  intercept_t free_t size_t SAT_t  intercept_probt
free_probt size_probt SAT_probt _imputation_;
if _NAME_ = 'Estimate' then do;
intercept_est = intercept;
free_est = free;
size_est=size;
SAT_est = x1;
end;
if _NAME_='StdErr' then do;
intercept_std = intercept;
free_std = free;
size_std=size;
SAT_std = x1;
end;
if _NAME_='tValue' then do;
intercept_t = intercept;
free_t = free;
size_t=size;
SAT_t = x1;
end;
if _NAME_='Probt' then do;
intercept_probt = intercept;
```

```
free_probt = free;
size_probt=size;
SAT_probt = x1;
end;
if _NAME_='Probt';
keep intercept_est free_est  size_est SAT_est intercept_std free_std
size_std SAT_std intercept_t free_t size_t SAT_t intercept_probt
free_probt size_probt SAT_probt  _imputation_;
if last._imputation_;
run;

/*LD
PART///////////////////////////////////////////////////////////////////
//////////*/
/*transposing fixed effects for LD data*/
proc transpose data = fixed_effectsLD out= fixed_transLD;
var estimate StdErr tvalue Probt;
id effect;
run;

/*renaming fixed effects for LD data*/
data fixedeffLD; set fixed_transLD;
retain intercept_estLD free_estLD size_estLD SAT_estLD  intercept_stdLD
free_stdLD
size_stdLD SAT_stdLD  intercept_tLD free_tLD size_tLD SAT_tLD
intercept_probtLD
free_probtLD size_probtLD SAT_probtLD;
if _NAME_ = 'Estimate' then do;
intercept_estLD = intercept;
free_estLD = free;
size_estLD=size;
SAT_estLD = x1;
end;
if _NAME_='StdErr' then do;
intercept_stdLD = intercept;
free_stdLD = free;
size_stdLD=size;
SAT_stdLD = x1;
end;
if _NAME_='tValue' then do;
intercept_tLD = intercept;
free_tLD = free;
size_tLD=size;
SAT_tLD = x1;
end;
if _NAME_='Probt' then do;
intercept_probtLD = intercept;
free_probtLD = free;
size_probtLD=size;
SAT_probtLD = x1;
end;
if _NAME_='Probt';
keep intercept_estLD free_estLD size_estLD SAT_estLD  intercept_stdLD
free_stdLD
size_stdLD SAT_stdLD  intercept_tLD free_tLD size_tLD SAT_tLD
intercept_probtLD
free_probtLD size_probtLD SAT_probtLD; run;
```

```
/*combine all imputed data;*/


ods output parameterestimates=miparms;

proc mianalyze data=fixedeff;
modeleffects intercept_est free_est size_est  SAT_est;
stderr  intercept_std free_std size_std SAT_std; run;

proc transpose data=miparms out=estimates_trans;
var estimate StdErr;
id Parm;
run;

data estall; set estimates_trans;
retain intercept_est1 free_est1  size_est1 SAT_est1 intercept_std
free_std
size_std SAT_std ;

if _NAME_ = 'Estimate' then do;
intercept_est1=intercept_est;
free_est1=free_est;
size_est1=size_est;
SAT_est1=SAT_est;
end;

if _Label_='Std Error' then do;
intercept_std=intercept_est;
free_std=free_est;
size_std=size_est;
SAT_std=SAT_est;
end;

if _Label_='Std Error';

keep intercept_est1 free_est1  size_est1 SAT_est1 intercept_std
free_std
size_std SAT_std ;
run;


data comb_seed; set estall;set fixedeffLD;
seed1=&seed1; seed2=&seed2; seed3=&seed3;seed4=&seed4;seed5=&seed5;
seed6=&seed6;seed7=&seed7;seed8=&seed8;seed9=&seed9;seed10=&seed10;
seed11=&seed11; iter=&i;
level2n=&level2n; lev1corr=&lev1corr; rescorr=&rescorr;
missper=&missper;

proc append base=MCAR3feeder data=comb_seed; run;/*change MCAR3feeder
to MAR3feeder and MNAR3feeder*/

%LET SEED1=&SEED1+2;
%LET SEED2=&SEED2+2;
%LET SEED3=&SEED3+2;
%LET SEED4=&SEED4+2;
%LET SEED5=&SEED5+2;
```

```
%LET SEED6=&SEED6+2;
%LET SEED7=&SEED7+2;
%LET SEED8=&SEED8+2;
%LET SEED9=&SEED9+2;
%LET SEED10=%eval(&SEED10+2);
%LET SEED11=&SEED11+2;
%end;
%mend;
%generate;

/*Calculating Relative Bias of the Parameter Estimates*/
/*change MCAR3feeder to MAR3feeder and MNAR3feeder*/
PROC MEANS DATA= MCARfeeder; VAR intercept_est1 free_est1 size_est1
SAT_est1
intercept_estLD free_estLD size_estLD SAT_estLD
intercept_std free_std size_std SAT_std intercept_stdLD free_stdLD
size_stdLD SAT_stdLD;

OUTPUT OUT=PARM1 MEAN= Sintercept_est1 Sfree_est1 Ssize_est1 SSAT_est1
Sintercept_estLD Sfree_estLD Ssize_estLD SSAT_estLD
intercept_std free_std size_std SAT_std intercept_stdLD free_stdLD
size_stdLD SAT_stdLD;
OUTPUT OUT=SE1 STD= ssintercept_est1 ssfree_est1 sssize_est1 ssSAT_est1
ssintercept_estLD ssfree_estLD sssize_estLD ssSAT_estLD ;


DATA BIAS; SET PARM1;drop _type_ _freq_;
PBIAS1_MI=((Sintercept_est1-100)/100)*100;
PBIAS2_MI= ((Sfree_est1-.5)/.5)*100;
PBIAS3_MI=((Ssize_est1-.5)/.5)*100;
PBIAS4_MI=((SSAT_est1-.5)/.5)*100;
PBIAS5_LD=((Sintercept_estLD-100)/100)*100;
PBIAS6_LD= ((Sfree_estLD-.5)/.5)*100;
PBIAS7_LD=((Ssize_estLD-.5)/.5)*100;
PBIAS8_LD=((SSAT_estLD-.5)/.5)*100;

PROC PRINT; VAR PBIAS1_MI PBIAS2_MI PBIAS3_MI PBIAS4_MI
                        PBIAS5_LD PBIAS6_LD PBIAS7_LD PBIAS8_LD; RUN;

/*Calculating Relative Bias of the Standard Errors*/
DATA SEM1; SET  PARM1;  SET SE1; drop _type_ _freq_;

SBIAS1_MI=(((intercept_std-ssintercept_est1)/ssintercept_est1))*100;
SBIAS2_MI=(((free_std-ssfree_est1)/ssfree_est1))*100;
SBIAS3_MI=(((size_std-sssize_est1)/sssize_est1))*100;
SBIAS4_MI=(((SAT_std-ssSAT_est1)/ssSAT_est1))*100;
SBIAS5_LD=(((intercept_stdLD-
ssintercept_estLD)/ssintercept_estLD))*100;
SBIAS6_LD=(((free_stdLD-ssfree_estLD)/ssfree_estLD))*100;
SBIAS7_LD=(((size_stdLD-sssize_estLD)/sssize_estLD))*100;
SBIAS8_LD=(((SAT_stdLD-ssSAT_estLD)/ssSAT_estLD))*100;
PROC PRINT; VAR SBIAS1_MI SBIAS2_MI SBIAS3_MI SBIAS4_MI
                SBIAS5_LD SBIAS6_LD SBIAS7_LD SBIAS8_LD;  RUN;


data MCAR3feeder;set bias; set sem1; run;/*change MCAR3feeder to
MAR3feeder and MNAR3feeder*/
```