SECOND LEVEL CLUSTER DEPENDENCIES:
A COMPARISON OF MODELING SOFTWARE AND MISSING DATA TECHNIQUES

A Dissertation

by

ROSS ALLEN ANDREW LARSEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2010

Major Subject: Educational Psychology

SECOND LEVEL CLUSTER DEPENDENCIES:

A COMPARISON OF MODELING SOFTWARE AND MISSING DATA TECHNIQUES

A Dissertation

by

ROSS ALLEN ANDREW LARSEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Victor Willson
	Bob Hall
Committee Members,	Mike Speed
	Suojin Wang
Head of Department,	Victor Willson

August 2010

Major Subject: Educational Psychology

ABSTRACT

Second Level Cluster Dependencies:

A Comparison of Modeling Software and Missing Data Techniques. (August 2010)

Ross Allen Andrew Larsen, B.S., Brigham Young University;

M.S., Brigham Young University

Co–Chairs of Advisory Committee: Dr. Victor Willson
Dr. Bob Hall

Dependencies in multilevel models at the second level have never been thoroughly examined. For certain designs first-level subjects are independent over time, but the second level subjects may exhibit nonzero covariances over time. Following a review of revelant literature the first study investigated which widely used computer programs adequately take into account these dependencies in their analysis. This was accomplished through a simulation study with SAS, and examples of analyses with Mplus and LISREL. The second study investigated the impact of two different missing data techniques for such designs in the case where data is missing at the first level with a simulation study in SAS.

The first study simulated data produced in a multiyear study varying the numbers of subjects in the first and second levels, the number of data waves, the magnitude of effects at both the first and second level, and the magnitude of the second level covariance. Results showed that SAS and the MULTILEV component in LISREL analyze such data well while Mplus does not.

The second study compared two missing data techniques in the presence of a second level dependency, multiple imputation (MI) and full information maximum likelihood (FIML). They were compared in a SAS simulation study in which the data was simulated with all the factors of the first study and the addition of missing data varied in amounts and patterns (missing completely at random or missing at random). Results showed that FIML is superior to MI because it produces lower bias and correctly estimates standard errors.

To Tatiana, Michael, and Charlotte Larsen

ACKNOWLEDGMENTS

I would like to thank my committee co-chairs, Dr. Willson and Dr. Hall, and my committee members, Dr. Speed, and Dr. Wang, for their guidance and support throughout the course of this research. Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience. Finally, thanks go to my wife for her patience and love and my two children who were very understanding that "daddy has to go to work."

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

TABLE Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

In education questions are asked such as "does intervention X improve students' performances?" Despite the deceptive simplicity of the question, it can be complicated and difficult to answer. To evaluate large-scale educational experiments it is necessary to consider the hierarchical nature of the data. Pupils are nested within classrooms, classrooms are nested within schools, schools are nested within districts, and districts are nested within counties, etc. Different information is available at the different levels. On the first level, which describes the individual students, there might be information on student quality, their previous work and their family background or SES. On the level describing teachers, there might be information on teacher quality, degree they earned, or whether the teacher has had specialized training. At the school level, information typically might be available about percent free-and-reduced lunch of the student body, whether a school is in a specialized program, or how much money the schools receive. Unfortunately, to analyze these variables on any of these levels separately can cause misleading results (Burstein, Kim, & Delandshere, 1989; Kreft, 1987). Thus, models that take all levels (student, classroom, school, district, county, etc) into account simultaneously are superior statistically. To deal with this problem researchers have attempted to handle these issues using hierarchical linear modeling (HLM), random coefficient modeling, or Bayesian linear modeling (Aitkin & Longford, 1986; Bryk & Raudenbush, 1992; Leeuw & Kreft, 1986; Fahrmeir, Tutz, & Hennevogl, 2001; Goldstein, 1987, 1993; Hox, 1995; Mason, Wong, & Entwisle, 1983). The approach that incorporates random coefficients is detailed below.

This dissertation follows the style of *Structural Equation Modeling*.

A. The Model

A typical multilevel model at the student level (Littell, Milliken, Stroup, & Wolfinger, 1996) is

$$(Score)_{ijt} = \beta_{0jt} + \beta_{1jt}(Pretest)_{ijt} + \epsilon_{ijt} \tag{1.1}$$

where i is the student, j is the teacher, and t is the year. *Score* is the result of an outcome test and *Pretest* is an antecedent covariate that may or may not be a parallel form to the outcome. The covariate is not required for the analysis; it merely reflects a common situation.

Next, we assume that the regression coefficients β_{0jt} and β_{1jt} arise from a model with nesting at the teacher level. Assuming initially that there are no teacher-level exogenous variables, the basic model is

$$\beta_{0jt} = \gamma_{00t} + \delta_{0jt} \text{ and} \tag{1.2}$$

$$\beta_{1jt} = \gamma_{10t} + \delta_{1jt}, \tag{1.3}$$

where the class level disturbance terms (δ_{0jt}, $\delta_{1j.}$) are assumed to be independent and identically distributed Gaussian variables with a zero mean and nonzero variance-covariance matrix.

Including a teacher-level covariate here termed *Quality*, a reasonable class level model then becomes

$$\beta_{0jt} = \gamma_{00t} + \gamma_{01}(Quality)_{jt} + \delta_{0jt} \text{ and} \tag{1.4}$$

$$\beta_{1jt} = \gamma_{10t} + \gamma_{11}(Quality)_{jt} + \delta_{1jt}. \tag{1.5}$$

Substituting these expressions for β_{0jt} and β_{1jt} into the student level model produces the following:

$$(Score)_{ijt} = \gamma_{00k} + \gamma_{01}(Quality)_{jt} + \gamma_{10}(Pretest)_{ijt} + \delta_{0jt} + \epsilon_{ijt} \tag{1.6}$$

where *Quality* is a teacher level exogenous variable for teacher j in year t. This model can written in general as

$$Y = X\beta + Z\mu + \epsilon, \tag{1.7}$$

where Y is a vector of the outcome; X and Z are known design matrices, both of which can include dummy variables and continuous covariates; β is the vector of fixed effects; μ is a vector of the random effects; and ϵ is the random error component. Then

$$Var(Y) = ZGZ' + R, \text{ where} \tag{1.8}$$

$$COV \begin{pmatrix} \mu \\ \epsilon \end{pmatrix} = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}. \tag{1.9}$$

1. The Independence Model

In the scenario where all the teachers' measurements are independent of each other and across the years, and thus there is no second level dependency (SLD). In this 'independence' model the G matrix in (1.8) and (1.9) can be written simply. For example, assume that there are two teachers across three years. Thus:

$$G = \begin{pmatrix} G_{11} & 0 \\ 0 & G_{22} \end{pmatrix}, \text{ and} \tag{1.10}$$

$$G_{11} = G_{22} = \begin{pmatrix} \sigma^2_{year} & 0 & 0 \\ 0 & \sigma^2_{year} & 0 \\ 0 & 0 & \sigma^2_{year} \end{pmatrix} \tag{1.11}$$

where G is a *j x t* matrix. The Z is a *n x 3* matrix which will be written as:

$$Z = \begin{pmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \end{pmatrix} \tag{1.12}$$

This relationship can be drawn in a path model with year as dummy coded variable as shown in Figure 1. The path model in Figure 1 will give equivalent results as when using (1.7) with the Z matrix defined in (1.12).

2. The Correlated Cluster Model

Moving from the independence model assumption to the correlated cluster model, in which the teacher's measurements may be dependent across the years, requires the SLD to be modeled in some way. A common modeling process that often captures the relationship well is

$$Average(Score)_{jt} = (1-\rho)\mu + \rho\{Average(Score_{j(t-1)}\} + \epsilon_{jt}, \tag{1.13}$$

an autoregressive process with one lag *AR(1)*, in which the average score for teacher j in year t is partly determined by the previous year's average score. G still follows

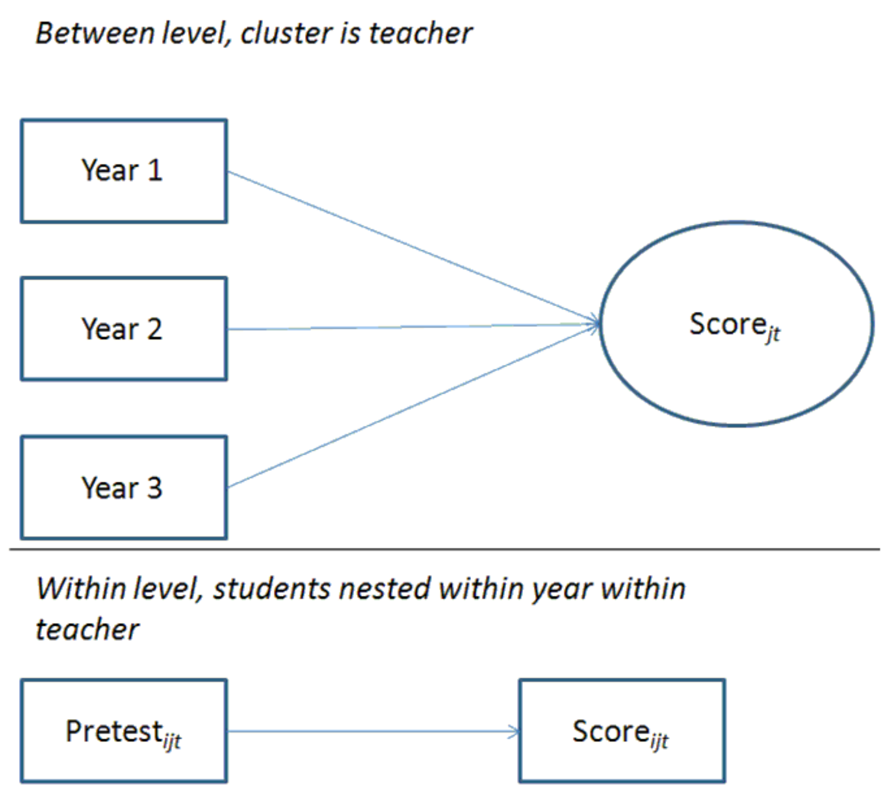

Figure 1 Path model for multiyear model with the second level variables independent.

the pattern shown in (1.10), but now

$$G_{11} = G_{22} = \sigma^2_{year} = \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho^2 \\ \rho^2 & \rho & 1 \end{pmatrix} \tag{1.14}$$

where ρ is the lag *AR(1)* coefficient (Ruppert, 2004). The Z matrix will still be defined as seen in (1.12). The equivalent path diagram is shown in Figure 2. This model can be generalized to have any covariate(s) at the second level given sufficient degrees of freedom with no loss of generalizability. More complex processes can be represented in a similar manner as that shown (1.14).

B. Missing Data

Missing data in educational research is almost always assured with student absence or mobility making it unlikely to have all test scores gathered on all students. Rubin (2004) made certain classifications for missing data and argued that missing data could be ignored if it is missing completely at random (MCAR) or missing at random (MAR). MCAR is defined as occuring when the probability that a data point Y is missing is independent from all other observed variables including Y itself. MAR is defined as occuring when the probability that a data point Y is missing can be determined from the other observed variables, or the variable itself. The typical default for ignoring missing data is listwise deletion. Allison (2002) argues that listwise deletion causes the standard error of the estimate to be inflated in the MCAR case and causes bias in the MAR case. Missing data is almost always assured in multiyear studies where SLD could occur. Improperly dealing with missing data could lead to biased estimates and underestimates of the standard errors (Chan, 1998).

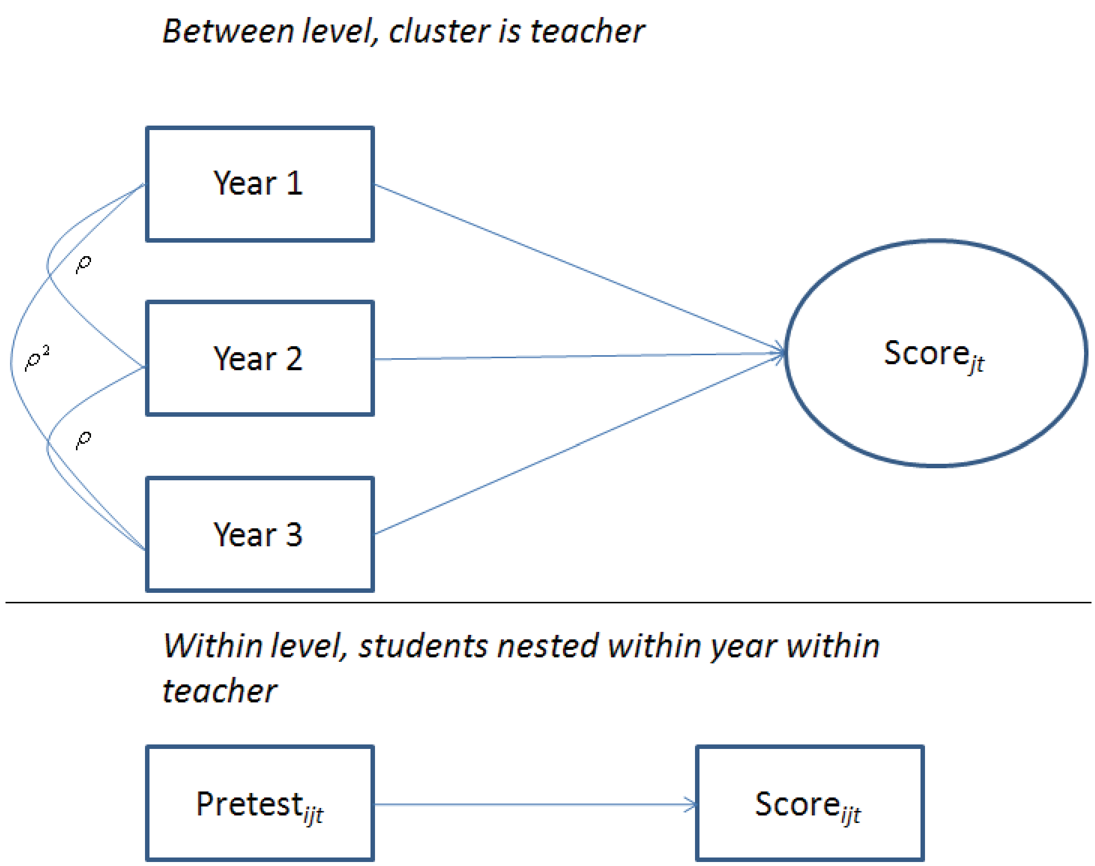

Figure 2 Path model for multiyear where the second level variable is correlated with an *AR(1)* structure.

C. Dissertation

The purpose of this dissertation was 1) to test three widely-used computer programs for their utility in analyzing a SLD correctly, and 2) to test two missing data techniques in the presence of SLD. The first study compares SAS, LISREL, and Mplus. The data was generated in SAS and analyzed in PROC MIXED. A subset of the data was then exported to the other two programs. The bias and the standard errors across the different programs was compared. The second study varied the amount of missing data, the type of missing data (MAR or MCAR), and deals with the missing data through either multiple imputations or full information maximum likelihood (FIML). The results were compared to the full dataset results, comparing the bias, the standard deviation of the estimates, and the standard errors. The analysis was done in SAS.

CHAPTER II

LITERATURE REVIEW

Social science tries to capture reality in complex situations, leading to the evolution of advanced statistical methodologies such as hierarchical linear modeling (HLM) and structural equation modeling (SEM). Many studies in education are longitidinal in nature. The effect of teachers in a public school on students might be expected to exhibit covariation of their class means over time due to the kinds of students that are assigned to a teacher's classroom or to the teachers' own characteristics. All this teacher or classroom covariation is termed second level dependency (SLD), second level because the covariation is at the classroom level, and dependency because the effect is not independent across years or time periods. Missing data in such research questions is normally assured with student absence or mobility making it unlikely that all test scores will be gathered on all students. What follows is a review of the literature relating to the computer programs that analyze such data effectively and the missing data techniques that effectively handle longitudinal data with a multilevel structure.

A. Hierarchical Linear Modeling (HLM)

Kreft, De Leeuw, & Kim (1990) compare four different HLM programs: GENMOD, HLM, ML2, and VARCL3. All four programs use maximum likelihood (ML) estimation to calculate the within and between variance components. None of these programs handle missing data. All of the programs use an iterative procedure to calculate parameters. The main differences come from the convergence criteria and choice of algorithm to optimize these criteria.

GENMOD was originally written by Benjamin Hermalin and Albert F. Anderson at the Population Studies Center at the University of Michigan. The authors used the general model proposed by Wong & Mason (1989). According to Kreft, De Leeuw, & Kim (1990) the documentation is difficult to interpret, the learning curve is steep, and the program doesn't handle small data sets (where number of observations within groups is less than the number of variables), has slow convergence, and does not have any options for weighting the data. The program uses restricted maximum likelihood (REML). GENMOD was never easy to obtain and has more less completely disappeared since it was developed (De Leeuw & Kreft, 2001).

VARCL3 was started by Aitkin & Longford (1986). It allows multilevel data but does not allow interactions between slopes or covariates higher than level 1. It has average documentation, an easy learning curve, analyzes small datasets, has relatively handling of errors that occur during the compiling and has options for weighting the data. Although very popular in the early 1990's, VARCL3 is no longer actively developed or supported (De Leeuw & Kreft, 2001).

HLM version 2.1 was written by Bryk, Raudenbush, and Congdon Bryk, Raudenbush, Congdon, & Seltzer (1988). HLM has an easy-to-use interface and has output that contains significance tests, and model testing. Unfortunately, HLM does have a faulty singularity test affecting the software's handling of any errors in compiling. HLM's outcome variable can be normal, binary, poisson, multinormal categorical, or ordered categorical distributed. HLM has developed into a windows version with a graphical user interface (GUI), and has HLM/2L which handles two-level analysis and HLM/3L which handles three-level analysis. Version 4 saw an improvement in its alorithm. Now it can use Fisher scoring instead of just the EM algorithm. The documentation is user oriented but many of the choices in analysis are already made by the developers and so are denied to the user. The faulty singularity test was not mentioned again in the more recent reviews and so must be resolved in the current versions(De Leeuw & Kreft, 2001). HLM version 6 saw a further improvement in the windows interface as options are placed in a logical order and the graphing capabilities have been improved. The newer version contains more advanced modeling techniques such as cross-classified random effects models for linear models and nonlinear link functions. HLM version 6 has a greater capability for incorporating sample weights in complex designs such as cluster sampling (Michela, 2006).

ML2 is software for two-level analysis by Rabash, Prosser, and Goldstein. It is based on work by Goldstein (1987). ML2 has a lot of functionality, with the ability to do data exploration and preparation before and after modeling. The manual is clear and easy to use but the profusion of options makes learning the program difficult. ML2 can only handle two level data but a more recent product, ML3 (Prosser, Rasbach, & Goldstein, 1991), can do three level analysis as well. Eventually, ML3 evolved into MLN which could handle any number of levels and MLN evolved into MLWIN, the version that was created for windows by the same group. MLWIN has the ability to analyze data where the outcome variable is either normal, binary, poisson, multinomial categorical, or ordered categorical.

MLA version 4.1 is a program developed for two-level analysis only (Busing, Meijer, & van der Leeden, 2005). The interface for MLA is a little archaic as it does not use a GUI that many modern programs do. It has many options for data simulation, including options for bootstrapping multilevel models. It has simple estimation methods as opposed to complex iterative procedures used by other programs and has a fast algorithm for all model parameters. The program is not longer actively supported.

Other reviews also compare SAS PROC MIXED and the GLIMMIX macro (which later became PROC GLIMMIX) comparing it with HLM, MLN, MLWIN, and VARCL (Zhou, Perkins, & Hui, 1999). They found SAS PROC MIXED to be comparable to the other programs, albeit slower. MLN and MLWIN were the only two programs that performed second order approximations. Nevertheless, SAS has the most error distributions and link functions available. PROC MIXED is only capable of analyzing data where the outcome is normal, while PROC GLIMMIX is able to analyze data where the outcome variable is either normal, binary, poisson, multinomial categorical, or ordered categorical. A complete discussion of SAS PROC MIXED and how to get equivalent results from SAS as you do from many other programs can be found in Singer (1998).

B. Structural Equation Modeling (SEM)

Chantala & Suchindran (2006) compared Mplus, MLWIN, LISREL, PROC MIXED (SAS), LISREL, and GLLAMM (STATA) and their abilities to analyze HLM models. In addition to HLM capabilities LISREL, MPlus, and GLLAMM (STATA) have the ability to do SEM-type analyses. The vendors for Mplus, MLWIN , LISREL, and HLM all claim that their most recent upgrades produce similar results when analyzing complex sampling data.

Mplus has the ability to incorporate sampling weights that are necessary in HLM. In many cases it takes less than a minute to converge(Muthén & Muthén, 1998-2007). The manuals are all available online, and there is extensive literature in online forums that answers most common questions. Nevertheless, Mplus does have a learning curve associated with its software, sometimes convergence is a problem, and the error messages in the output are ambiguous at times. Mplus is the only program reviewed that is able to do subpopulation analysis and with LISREL shares the distinction of being able to adjust analyses for stratification. Mplus' outcome variables can be normal, binary, poisson, multinomial categorical and ordered categorical.

LISREL (Jöreskog, Sörbom, & Du Toit, 2000) also claims to have the ability to incorporate sampling weights that are necessary in HLM. Unfortunately, simulation results provide evidence that the sampling weights are not correctly integrated into the estimation of the standard errors of the parameters, as shown in differing results from HLM and Mplus(Chantala & Suchindran, 2006). The LISREL designers are working on the problem. LISREL can only handle outcome variables that are normal in nature. Additionally, LISREL has issues with its syntax, and the learning curve for certain situations is steep. The published help available at LISREL appears to be excellent.

GLLAMM (Stata) (Rabe-Hesketh, Skrondal, & Pickles, 2001) allows HLM weights and is able to do normal, binary, poisson, multinomial categorical and ordered categorical outcome variables. The results for GLLAMM are comparable to Mplus and MLWIN, but while the other programs converged in under a minute GLLAMM took over six hours for moderately complex models. This is perhaps do to approximations that work well for the normal model.

PROC MIXED in SAS is not able to use HLM weights with stratification properly, but PROC NLMIXED with the help of a macro does(Rabe-Hesketh, Skrondal, & Pickles, 2001). PROC NLMIXED will handle binary and poisson response variables while PROC MIXED only handles normal response variables. PROC GLIMMIX has all the capability of PROC MIXED but does allow non-normal outcome variables. Other PROCS such as PROC SURVEYSELECT may be more appropriate for complex survey data (Stapleton, 2006).

C. Error Structure

The programs that specifically model HLM generally assume that the errors at the second level are independent as in the case of MLWIN (Bryk, Raudenbush, Congdon, & Seltzer, 1988), or does not address the issue at the second level as in the case of HLM6 (Rasbash et al., 2000). PROC MIXED, on the other hand, has a rich array of specific options, such as autoregressive, compound symmetry, Toeplitz, or variance components, for specifying the structure of the error matrix (Singer, 1998).Manuals for VARCL3 and GENMOD were unavailable but as they are not currently maintained they can be assumed to be at the same technical level or lower as HLM6 or MLWIN.

According to the theory all SEM programs should be able to handle any error structure the user can conceive. LISREL has less options than PROC MIXED but allows the user to specify any error structure directly (Jöreskog & Sörbom, 2005). Mplus' approach to the error matrix is less intuitive and has to be modeled indirectly using the variables themselves (Muthén & Muthén, 1998-2007). It appears that GLLAMM does not approach this problem directly, but as GLLAMM is a SEM program it can be assumed that errors can be modeled by specifying the relationship between the variables themselves as is done in Mplus (Rabe-Hesketh, Skrondal, & Pickles, 2001).

D. Missing Data

Missing data can lead to large standard errors and even bias when data is missing in systematic ways (Chan, 1998). Approaches such as listwise deletion, pairwise deletion, full information maximum likelihood (FIML), and similar response pattern imputation (Jöreskog & Sörbom, 1996) have all been studied with regard to their estimation bias when the data is incomplete (Enders & Bandalos, 2001). Hot-deck imputation and mean imputation were studied by Brown (1994). Although all these techniques have been evaluated for MAR and MCAR data (Newman, 2003), and other work has suggested the need for investigating the effect that model misspecification can have on the results of multiple imputation (Duncan & Duncan, 1994), nonetheless there is little reported research on correctly specified and misspecified SLD data and their effects on missing data procedures. Lavori, Dawson, & Shera (1995) suggested using a Bayesian approach with propensity scoring to deal with the missing data. Others argued that simply using FIML approaches will serve well (Molenberghs et al., 2004).

CHAPTER III

MODELING SECOND LEVEL CLUSTER DEPENDENCIES

In the social sciences, multilevel designs have become increasingly complicated. Many times, in education, students are nested in classrooms and then analyzed as if the data was independent. This leads to incorrect results. Additionally, in education, studies are carried out over a period of several years with different students each year. Educational psychologists may be interested in the outcome variable at the student level, but they need to account for the fact that the same teachers are present each year. On the other hand, they may be interested in second level (classroom) outcomes associated with teachers over time. Teachers may be expected to covary on relevant variables such as their teaching quality over time, violating the usual assumption of independent units at the second level. This is a second level dependency (SLD). Hedges (2009) acknowledges this problem and proposes an adjustment to the t statistics post-hoc. Other work has been done to analytically calculate the SLD dependencies in certain meta-analyses where the structure is known (Stevens & Taylor, 2008).

One way to model this situation is to treat each teacher as a cluster at the third level with classroom means as data points at the second level. Unfortunately, this approach generally assumes independence at the second level. Muthen (1997), through simulation, shows the distortion effects on the standard errors and chi-square statistics when data are assumed to be independent when in fact there is dependencies at the first level. These results can be generalized to higher levels as well. Thus, to correctly model this situation with a three-level model would require the covariance matrix level at the second level to be unstructured or have some sort of structure that does not assume independence. Generally there are insufficient data points at the second level to obtain stable results as multilevel modeling requires the sample sizes for the lower levels to be large in order to have reliable results (Bentler, 1980; Maas & Hox, 2005). Thus, the three-level approach is generally unworkable.

Another way to model SLD is to use a mixed model with random effects (Littell, Milliken, Stroup, & Wolfinger, 1996). The random effect of year may have a nonzero correlation structure, to correctly account for the potential of teachers' effects on the student outcomes to be correlated across years. Many possible forms of the correlation structure may exist. One, an unstructured correlation matrix, allows all the elements in the correlation matrix to vary. This structure consumes many degrees of freedom and is cumbersome to interpret. Additionally, unstructured covariance estimation has been shown to lead to underestimates of standard errors (Kwok, West, & Green, 2007). Usually, if there is a theoretical basis for using a more constrained structure, it is preferred. For example, an autoregressive AR process: 1) may fit the theory behind the relationship; 2) consumes fewer degrees of freedom; and 3) may be easier to interpret. Our intent is to expand this to structural equation modeling (SEM) problems.

A. The Model

A general multilevel model can be defined as it is in SAS (Littell, Milliken, Stroup, & Wolfinger, 1996). Put into the context of an educational study the model can be written as

$$(Score)_{ijt} = \beta_{0jt} + \beta_{1jt}(Pretest)_{ijt} + \epsilon_{ijt} \tag{3.1}$$

where i is the student in the classroom, j is the teacher of the classroom, and t is the year. *Score* is the result. Generally, it is a standardized test, but any quantitative outcome variable can be used. *Pretest* is an antecedent covariate that may or may not be a parallel form to the outcome. Having the covariate will control for prior achievement but is not necessary, from a statistical point of view, to do the analysis.

According the framework that Littell, Milliken, Stroup, & Wolfinger (1996) uses the regression coefficients β_{0jt} and β_{1jt} arise from a model with nesting at the classroom or teacher level. Assuming initially that there are no teacher-level exogenous variables, the basic model is

$$\beta_{0jt} = \gamma_{00t} + \delta_{0jt} \text{ and} \tag{3.2}$$

$$\beta_{1jt} = \gamma_{10t} + \delta_{1jt}, \tag{3.3}$$

where the class level disturbance terms (δ_{0jt}, $\delta_{1j.}$) are independent and identically distributed Gaussian variables with a zero mean and have a nonzero variance-covariance matrix.

Including a covariate here termed *Quality*, which captures the quality of teaching in a classroom, a reasonable class level model then becomes

$$\beta_{0jt} = \gamma_{00t} + \gamma_{01}(Quality)_{jt} + \delta_{0jt} \text{ and} \tag{3.4}$$

$$\beta_{1jt} = \gamma_{10t} + \gamma_{11}(Quality)_{jt} + \delta_{1jt}. \tag{3.5}$$

These expressions for β_{0jt} and β_{1jt} will be substituted into the full model, as shown in (3.1), resulting in:

$$(Score)_{ijt} = \gamma_{00k} + \gamma_{01}(Quality)_{jt} + \gamma_{10}(Pretest)_{ijt} + \delta_{0jt} + \epsilon_{ijt} \tag{3.6}$$

where *Quality* is a classroom level exogenous variable for classroom or teacher j in year t. This model can written in general as

$$Y = X\beta + Z\mu + \epsilon, \tag{3.7}$$

here Y is a vector of the quantitative outcome; X and Z are design matrices which are none, both of which can include dummy variables and continuous covariates; β is the vector of fixed effects; μ is a vector of the random effects; and ϵ is the random error component. Then

$$Var(Y) = ZGZ' + R, \text{ where} \tag{3.8}$$

$$COV \begin{pmatrix} \mu \\ \epsilon \end{pmatrix} = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}. \tag{3.9}$$

1. The Independence Model

In the scenario where all the teachers' measurements are independent of each other and across the years there is no SLD, and thus the G matrix in (3.8) and (3.9) can be written simply. For example, assume that there are two teachers across three years. Thus:

$$G = \begin{pmatrix} G_{11} & 0 \\ 0 & G_{22} \end{pmatrix}, \text{ and} \tag{3.10}$$

$$G_{11} = G_{22} = \begin{pmatrix} \sigma^2_{year} & 0 & 0 \\ 0 & \sigma^2_{year} & 0 \\ 0 & 0 & \sigma^2_{year} \end{pmatrix} \tag{3.11}$$

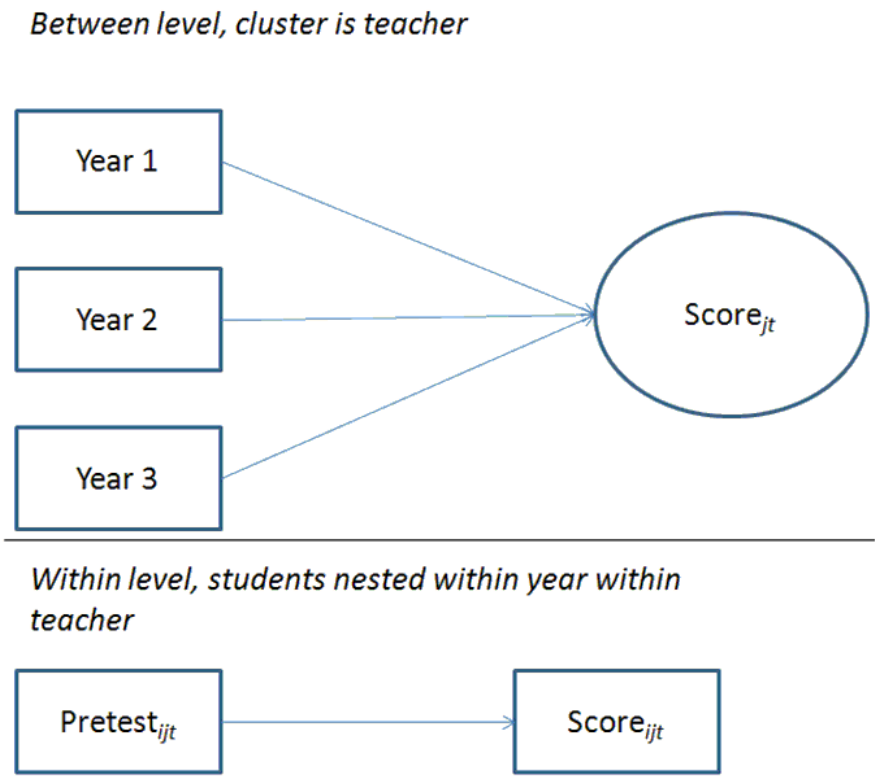

Figure 3 Path model for multiyear model with teacher variable independent across years.

where G is a $j\ x\ t$ matrix. The Z is a $n\ x\ 3$ matrix which will be written as:

$$Z = \begin{pmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \end{pmatrix} \tag{3.12}$$

An equivalent path model is shown in Figure 3. This path model will give equivalent results as when using (3.7) with the Z matrix defined in (3.12).

2. The Correlated Cluster Model

The 'independence' model is unrealistic in many educational settings where the same teachers are followed over the course of several years. A common modeling process that reflects this fact is

$$Average(Score)_{jt} = (1-\rho)\mu + \rho\{Average(Score_{j(t-1)}\} + \epsilon_{jt}, \tag{3.13}$$

which is an autoregressive process with one lag *AR(1)*. The average score for classroom or teacher j in year t is partly determined by the previous year's average score. G still follows the pattern shown in (1.10), but now

$$G_{11} = G_{22} = \sigma^2_{year} \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho^2 \\ \rho^2 & \rho & 1 \end{pmatrix} \tag{3.14}$$

where the *AR(1)* parameter is ρ (Ruppert, 2004). The Z matrix is (3.12). This model has its equivalent form in a path model as shown in Figure 4. The path model shown assumes only one covariate at the classroom or teacher level, but additional covariates can be added with no ill results given sufficient degrees of freedom.

B. Method

Many software packages analyze the zero SLD multilevel model reasonably well: Mplus (Muthén & Muthén, 1998-2007), the statistical application MULTILEV of LISREL for Windows (Jöreskog & Sörbom, 2005), SAS 9.2, and R (Ihaka & Gentleman, 1996), for example. We conducted a simulation study to evaluate the adequacy of estimation (bias) of the point estimate, the bias of the standard errors and stan-

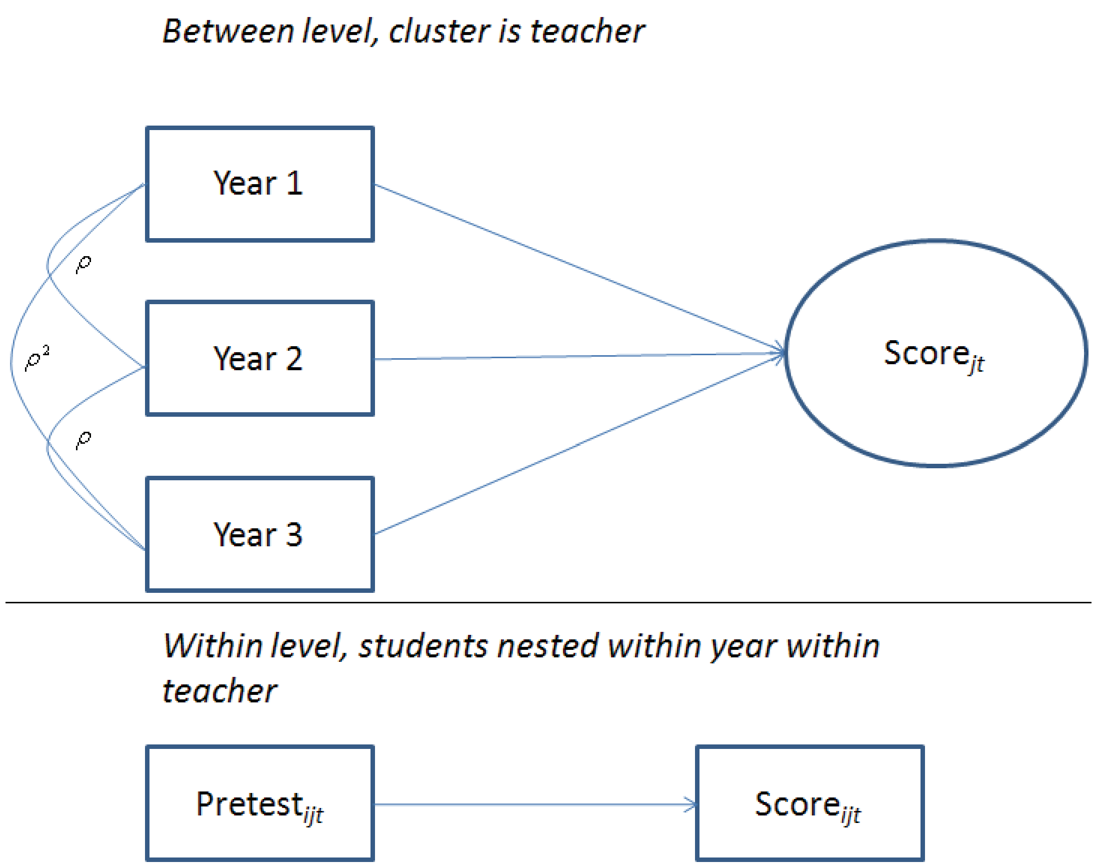

Figure 4 Path model for multiyear where teacher variable is correlated with an *AR(1)* structure.

dard deviations of the point estimates in a nonzero covariance SLD for a 2x2x2x3x2 factorial design for the first three programs. The data was simulated in SAS PROC IML. Table 1 shows the design conditions. To represent a more realistic design, we included predictors at both first and second levels.

Table 1

Number of Teachers, Students, Years, and the Different AR(1) Conditions in the Simulation Study

Variable	Number	Effect Size (σ)
Teachers	20,35	(0,1)[a]
Students	20	(0,1)[b]
Year	3,5,9	0
AR(1)	2	0,.5(ρ)

[a]Represents teacher's 'quality.' [b]Represents student's 'ability.'

We varied the number of 'teachers' as either 20 or 35, reflecting a medium to large study. Teacher quality effect size referred to two scenarios, one where teachers are homogeneous (effect size 0) and thus their 'quality' scores will be a random draw from a standard normal distribution $N(0,1)$. The second scenario occurs when the effect size is equal to '1'. In that case, half of the teachers have a superior teaching quality, with their scores drawn from a $N(1,1)$, while the normal teachers are drawn from a $N(0,1)$. This is mirrored at the student level, where either the students are homogeneous and their pretest scores are drawn from $N(0,1)$, or half of the students in each class have pretest scores drawn from a $N(1,1)$ distribution and half from $N(0,1)$.

The year row refers to how many waves of data (2, 4, or 8) were generated, with the assumption that there is no improvement or decline in the teacher's exogenous quality scores during the study's duration. The $AR(1)$ column refers to how teachers' quality scores covary between years of the study. In the '0' value for condition, the teacher's quality scores are independent. In the 0.5 condition, the teacher's quality scores are drawn according to an *AR(1)* process:

$$Quality_{jt} = (1-\rho)\mu + \rho(Quality)_{j(t-1)} + \epsilon_{jt} \tag{3.15}$$

as shown in equation (3.15) (Ruppert, 2004), with the $AR(1)$ parameter, ρ, of 0.5. Each of these conditions was simulated with 100 iterations. Notice in (3.15) that 'Quality' simply captures the AR structure. Without an exogenous variable, the AR process is carried by the second level error covariance structure $\epsilon_j + (1-\rho)\mu + \rho\epsilon_{t-1}$.

The dependent variable in the regression will be the predicted test scores of the simulated students. The regression will produce estimates of the parameters which will be averaged across all the iterations to produce the bias, standard errors, and standard deviation of the estimates of the parameters of both first (student 'ability') and second (teacher 'quality') levels. Bias is defined as the averaged estimated value subtracted from the true parameter value. Additionally, the software programs ML-WIN (Rasbash et al., 2000) and WINBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) documentation will be examined to discover their ability to analyze SLD data but no analysis using simulated data will be attempted.

C. Results

In order to have a gold standard of comparison, a sample of the data is analyzed with SAS PROC MIXED using a selected condition that reflected a typical study. The sample is then analyzed in Mplus (Muthén & Muthén, 1998-2007) and MULTILEV of LISREL for Windows (Jöreskog & Sörbom, 2005) to evaluate program-specific compatibility and estimation problems. The data generated in SAS was then evaluated in MULTILEV and Mplus. Because of the difficulty of analyzing many different scenarios in Mplus and MULTILEV, one typical simulation was selected from the simulation with ten iterations run at both four and eight years. The conditions selected for the test case are: number of teachers, 35; number of students, 20; the $AR(1)$ component, .5; and teacher 'quality' and student 'ability' effect sizes of 1. Table 2 summarizes the findings for the results in SAS. The 'Student level bias' and 'Teacher level bias' columns show how much bias was in the estimation of the parameters 'quality' (at the classroom or teacher level) and 'ability' (at the student level). As can be seen, the bias was minimal. The σ column shows the variance of the estimate. The 'Average SE' column shows the average standard error for the estimate.

Table 2

Bias, Standard Deviation of Estimate, and Average Standard Error of Simulated Data Analyzed by PROC MIXED (Students=20, simulations=10,000)

				Teachers=20					
				Student Level Estimates			Teacher Level Estimates		
ρ	ability	quality	years	bias	$\hat{\sigma}$	$\overline{\text{S.E.}}$	bias	$\hat{\sigma}$	$\overline{\text{S.E.}}$
0.0	0	0	2	−0.003	0.058	0.061	−0.018	0.245	0.274
0.5	0	0	2	0.003	0.059	0.061	−0.080	0.378	0.354
0.0	1	0	2	0.008	0.058	0.062	−0.007	0.225	0.266
0.5	1	0	2	−0.002	0.061	0.060	−0.066	0.339	0.354
0.0	0	1	2	−0.002	0.053	0.062	−0.038	0.249	0.272
0.5	0	1	2	0.001	0.056	0.060	−0.015	0.353	0.360
0.0	1	1	2	0.005	0.050	0.062	−0.021	0.296	0.269
0.5	1	1	2	0.005	0.063	0.061	−0.069	0.383	0.347
0.0	0	0	4	0.001	0.045	0.047	−0.015	0.192	0.211
0.5	0	0	4	0.004	0.053	0.047	0.002	0.269	0.305
0.0	1	0	4	0.002	0.044	0.048	−0.050	0.227	0.211
0.5	1	0	4	−0.001	0.042	0.048	0.057	0.281	0.307
0.0	0	1	4	−0.007	0.046	0.047	−0.022	0.218	0.214
0.5	0	1	4	0.004	0.049	0.048	−0.044	0.341	0.303
0.0	1	1	4	−0.002	0.042	0.047	−0.011	0.202	0.207
0.5	1	1	4	0.002	0.046	0.047	−0.036	0.293	0.308
0.0	0	0	8	−0.001	0.036	0.035	−0.002	0.158	0.164
0.5	0	0	8	−0.008	0.033	0.035	−0.006	0.251	0.246
0.0	1	0	8	0.003	0.033	0.035	−0.038	0.156	0.160

Table 2 (Continued)

				Student Level Estimates			Teacher Level Estimates		
ρ	ability	quality	years	bias	$\hat{\sigma}$	$\overline{\text{S.E.}}$	bias	$\hat{\sigma}$	$\overline{\text{S.E.}}$
0.5	1	0	8	−0.002	0.031	0.036	0.002	0.207	0.248
0.0	0	1	8	0.005	0.034	0.035	−0.009	0.166	0.162
0.5	0	1	8	0.000	0.031	0.035	−0.016	0.247	0.243
0.0	1	1	8	0.000	0.028	0.035	−0.001	0.162	0.164
0.5	1	1	8	−0.007	0.038	0.035	−0.039	0.251	0.243

				Teachers=35					
				Student Level Estimates			Teacher Level Estimates		
ρ	ability	quality	years	bias	$\hat{\sigma}$	$\overline{\text{S.E.}}$	bias	$\hat{\sigma}$	$\overline{\text{S.E.}}$
0.0	0	0	2	0.001	0.043	0.046	0.027	0.169	0.207
0.5	0	0	2	0.005	0.044	0.046	0.049	0.278	0.276
0.0	1	0	2	0.004	0.041	0.046	0.044	0.198	0.201
0.5	1	0	2	0.001	0.047	0.046	−0.033	0.276	0.264
0.0	0	1	2	0.003	0.047	0.046	−0.004	0.195	0.204
0.5	0	1	2	−0.008	0.040	0.046	0.002	0.246	0.270
0.0	1	1	2	0.000	0.040	0.045	0.000	0.191	0.205
0.5	1	1	2	−0.001	0.039	0.046	−0.038	0.272	0.265
0.0	0	0	4	0.003	0.032	0.036	0.007	0.150	0.158
0.5	0	0	4	0.000	0.032	0.035	0.360	0.240	0.231
0.0	1	0	4	−0.001	0.037	0.035	0.011	0.164	0.159
0.5	1	0	4	−0.001	0.033	0.035	−0.024	0.198	0.233
0.0	0	1	4	0.003	0.036	0.035	0.035	0.141	0.163

Table 2 (Continued)

				Student Level Estimates			Teacher Level Estimates		
ρ	ability	quality	years	bias	$\hat{\sigma}$	$\overline{\text{S.E.}}$	bias	$\hat{\sigma}$	$\overline{\text{S.E.}}$
0.5	0	1	4	−0.004	0.034	0.035	−0.012	0.228	0.227
0.0	1	1	4	−0.002	0.030	0.035	−0.007	0.158	0.162
0.5	1	1	4	−0.002	0.035	0.046	0.034	0.226	0.227
0.0	0	0	8	−0.001	0.022	0.026	−0.003	0.085	0.120
0.5	0	0	8	0.003	0.025	0.026	−0.002	0.198	0.185
0.0	1	1	8	−0.001	0.026	0.026	−0.001	0.144	0.119
0.5	1	0	8	0.003	0.027	0.026	−0.030	0.185	0.182
0.0	0	1	8	−0.003	0.023	0.026	0.009	0.098	0.120
0.5	0	1	8	−0.003	0.027	0.027	0.008	0.174	0.186
0.0	1	1	8	0.002	0.024	0.026	−0.011	0.127	0.119
0.5	1	1	8	−0.002	0.020	0.026	0.000	0.207	0.183

Table 3 shows the results averaged over the variables. As can be seen, the standard errors behave as one would expect, decreasing with additional teachers and years.

1. Parameter Estimation

The results in the fixed parameter estimation comparing the programs SAS, Mplus, and MULTILEVEL are given in Table 4 for ten simulations.

SAS's bias was low for the fixed parameters and the *AR(1)* component. SAS also had the property of consistency, as the number of years increased the bias decreased.

Table 3

Results of Simulation Using PROC MIXED Averaged Across the Variables

	Student Level Estimate			Teacher Level Estimate		
Variable	bias	$\hat{\sigma}$	$\overline{\text{S.E.}}$	bias	$\hat{\sigma}$	$\overline{\text{S.E.}}$
$\rho = 0$	0.0006	0.0385	0.0419	−0.0052	0.1786	0.1880
$\rho = 0.5$	−0.0003	0.0402	0.0418	−0.0133	0.2634	0.2644
Ability=0	−0.0002	0.0399	0.0418	−0.0046	0.2196	0.2272
Ability=1	0.0005	0.0389	0.0418	−0.0138	0.2224	0.2252
Quality=0	0.0009	0.0399	0.0419	−0.0058	0.2160	0.2265
Quality=1	−0.0006	0.0388	0.0418	−0.0127	0.2261	0.2259
Teachers=20	0.0003	0.0455	0.0479	−0.0225	0.2537	0.2580
Teachers=35	0.0000	0.0322	0.0358	0.0041	0.1883	0.1944
Years=2	0.0013	0.0500	0.0534	−0.0166	0.2683	0.2741
Years=4	−0.0001	0.0398	0.0413	−0.0024	0.2206	0.2267
Years=8	−0.0008	0.0283	0.0308	−0.0086	0.1742	0.1779
Overall	0.0001	0.0394	0.0418	−0.0092	0.2210	0.2262

Table 4

Bias, Standard Deviation of Point Estimate, and the Average Standard Error for SAS, MULTILEV, and Mplus. Number of Teachers=35, Number of Students=20, ρ=.5, teacher 'Quality'=1, student 'Ability'=1 (simulations=10)

		Years=4								
		SAS			Mplus			MULTILEV		
Variable	Value[a]	bias	$\hat{\sigma}$	$\overline{\text{S.E.}}$	bias	$\hat{\sigma}$	$\overline{\text{S.E.}}$	bias	$\hat{\sigma}$	$\overline{\text{S.E.}}$
Ability	1.0	−0.02	0.04	0.04	0.03	0.03	−0.02	0.04	0.04	0.03
Quality	1.0	−0.18	0.24	0.25	0.13	0.19	−0.19	0.24	0.24	0.13
ρ	0.5	0.06	0.25	0.13	0.15	0.09	0.00	0.18	0.14	0.15

		Years=8								
		SAS			Mplus			MULTILEV		
Variable	Value[a]	bias	σ	$\overline{\text{S.E.}}$	bias	σ	$\overline{\text{S.E.}}$	bias	σ	$\overline{\text{S.E.}}$
Ability	1.0	−0.01	0.03	0.03	0.01	0.02	0.01	−0.01	0.03	0.03
Quality	1.0	−0.13	0.21	0.20	0.15	0.08	0.09	−0.13	0.21	0.18
ρ	0.5	0.01	0.07	0.06	0.18	0.11	0.04	−0.02	0.12	0.10

[a]Population Parameter Value

MULTILEV and Mplus both had a larger bias for the *AR(1)* component. MULTILEV's bias decreased as the years increased while Mplus' bias actually increased as the years increased for the estimation of the 'quality' term and the *AR(1)* component.

The 2nd level correlation matrix estimated by SAS averaged across the 10 iterations can be seen in Table 5 (4 years) and Table 6 (8 years). SAS correctly estimates the correlation according to an AR(1) structure with no covariance estimates dipping below 0 and all estimates very close to the theorectical value with differences that are due to sampling.

Table 5

The 2nd Level Correlation Matrix as Estimated by SAS, Averaged over 10 Iterations (Years=4)

	Year 1	Year 2	Year 3	Year 4
Year 1	1.00			
Year 2	0.44	1.00		
Year 3	0.22	0.44	1.00	
Year 4	0.11	0.22	0.44	1.00

The 2nd level correlation matrix estimated by MULTILEV averaged across the 10 simulations can be seen in Table 7 (4 years) and Table 8 (8 years). As can be seen, for MULTILEV the covariance parameter does not mirror an *AR(1)* structure. Notice in the year 8 scenario there are negative estimates for the correlation which is impossible with a positive *AR(1)* value. This is due to the fact that is impossible to specify an exact *AR(1)* covariance structure in MULTILEV. Nevertheless, the es-

Table 6

The 2nd Level Correlation Matrix as Estimated by SAS, Averaged over 10 Simulations (Years=8)

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8
Year 1	1.00							
Year 2	0.49	1.00						
Year 3	0.25	0.49	1.00					
Year 4	0.13	0.25	0.49	1.00				
Year 5	0.06	0.13	0.25	0.49	1.00			
Year 6	0.03	0.06	0.13	0.25	0.49	1.00		
Year 7	0.02	0.03	0.06	0.13	0.25	0.49	1.00	
Year 8	0.01	0.02	0.03	0.06	0.13	0.25	0.49	1.00

Table 7

The 2nd Level Correlation Matrix as Estimated by MULTILEV, Averaged over 10 Simulations (Years=4)

	Year 1	Year 2	Year 3	Year 4
Year 1	1.00			
Year 2	0.47	1.00		
Year 3	0.22	0.47	1.00	
Year 4	0.04	0.22	0.47	1.00

Table 8

The 2nd Level Correlation Matrix as Estimated by MULTILEV, Averaged over 10 Simulations (Years=8)

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8
Year 1	1.00							
Year 2	0.50	1.00						
Year 3	0.25	0.50	1.00					
Year 4	0.09	0.25	0.50	1.00				
Year 5	0.01	0.09	0.25	0.50	1.00			
Year 6	−0.03	0.01	0.09	0.25	0.50	1.00		
Year 7	−0.03	−0.03	0.01	0.09	0.25	0.50	1.00	
Year 8	0.02	−0.03	−0.03	0.01	0.09	0.25	0.50	1.00

timates are close, with errors due to sampling and the fact that only 10 simulations were used to get these estiamtes.

The 2nd level correlation matrix estimated by Mplus averaged across the 10 iterations can be seen in Table 9 (4 years) and Table 10 (8 years). Mplus does not come anywhere close to the theorectical values in its estimation. In the 4 year scenario as seen in Table 9 the off diagonals do not appear to follow a *AR(1)* structure, where the (2,1) coordinate would be .5, the (3,1) coordinate would be .25 and the (4,1) coordinate would be .0625. Additionally, the off-diagonal elements vary from each other as they should not. Coordinates (2,1) and (3,2) should be the same but they are not. Table 10 shows that the problems observed in the 4 year scenario are not resolved with more years added to the study. In fact the bias becomes more extreme; the (8,1) coordinate should be very close to 0 but Mplus estimates the value to be 0.26.

Table 9

The 2nd Level Correlation Matrix as Estimated by Mplus, Averaged over 10 Iterations (Years=4)

	Year 1	Year 2	Year 3	Year 4
Year 1	1.00			
Year 2	0.58	1.00		
Year 3	0.46	0.64	1.00	
Year 4	0.32	0.46	0.66	1.00

Table 10

The 2nd Level Correlation Matrix as Estimated by Mplus, Averaged over 10 Iterations (Years=8)

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8
Year 1	1.00							
Year 2	0.58	1.00						
Year 3	0.46	0.64	1.00					
Year 4	0.32	0.46	0.66	1.00				
Year 5	0.27	0.35	0.47	0.64	1.00			
Year 6	0.26	0.33	0.37	0.40	0.60	1.00		
Year 7	0.30	0.33	0.31	0.32	0.47	0.65	1.00	
Year 8	0.26	0.22	0.15	0.21	0.30	0.43	0.61	1.00

2. MLWIN and WINBUGS

MLWIN (Rasbash et al., 2000) and WINBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) were examined for their capabilities in analyzing data with SLD as well. Neither program will allow the user to specify the structure of the 2nd level covariance matrix. Instead, an unstructured covariance matrix is estimated. This is an inefficient solution as the unstructured covariance structure will consume many degrees of freedom and any theory of the structure of the data cannot be applied to the analysis directly. Thus, power is expected to be reduced with these programs, although no bias would be expected.

D. Discussion

So which program should be used? The answer is, it depends on what the researcher is trying to accomplish. In the cases of a PATH analysis with no mediating variables, SAS's PROC MIXED reigns supreme. If latent structures or mediating variables are of interest, PROC CALIS would be used. PROC CALIS has structural equation capability, but does not permit inclusion of random effects; thus SLD or other complex correlated structures cannot be correctly estimated. Mplus has the capability to analyze complex SEM frameworks, but does not handle SLD. The best alternative for an SEM with latent variables and/or mediating relationships with a SLD is MULTILEV. This fact is important when choices are made on which software programs to invest time and money on and which software packages should be taught in the classroom. Many frustrating hours of programming can be sidestepped if the proper software with the capabilities needed can be chosen early. As every software program has its quirks, learning curve, and different capabilities it is important to know how to get equivalent results from different programs (Albright & Park, 2008;

Chantala & Suchindran, 2006; Kenny, 2007, May). This is true even within a program using the program's different procedures (Gao, Thompson, Xiong, & Miller, 2009). In the educational literature regarding longitudinal studies model misspecification with missing independent variables can be a problem (Dewey, Husted, & Kenny, 2000), as well as misspecification of the structure of the model which the fit indices may not detect (Fan & Sivo, 2005, 2007). To add on top of these common problems an SLD that is not correctly accounted for can lead to further bias in the parameter and standard error estimations. Thus, it behooves the researcher to use the software that has the capability to model their research question as closely to reality as possible.

CHAPTER IV

MISSING DATA IMPUTATION VERSUS FULL INFORMATION MAXIMUM LIKELIHOOD IN THE PRESENCE OF A SECOND LEVEL DEPENDENCY

Social science and in studying educational attempts to understand reality in nested situations that lead to complex statistical techniques such as mulilevel or hierarchial linear modeling (HLM). In the educational setting it is common for a study to persist over several years where classrooms with the same teachers, but different students, are observed over time. In such situations, the researcher may be interested in student growth or change as measured by standardized tests. Because the teachers are the same across time it would be inappropiate to treat each classroom as independent across the years. Instead, some kind of covariance structure that takes into account this dependency needs to be employed. This dependency exists only at the classroom or teacher level which is the 2nd level, thus such dependencies are referred to as second level dependencies (SLD). Few structural equation model (SEM) programs handle this characteristic of the data well (Larsen & Willson, unpublished). To further complicate matters, it is almost assured that not all test scores will be gathered for all students leading to missing data.

Missing data can either be missing completely at random (MCAR), or missing at random (MAR) (Rubin, 1976). MCAR is defined as when the probability that a data point Y is missing is independent from all other observed variables including Y itself. In the case of MCAR, Rubin (1976) argues that the fact that some of the data is missing can be safely ignored. This is not the case of MAR data where the probability that Y is missing depends on the other observed variables.

Listwise deletion, a common missing data technique, leads to inflated standard errors for the parameter estimates in the case of MCAR and bias in the parameter estiamtes in the MAR case (Allison, 2002; Chan, 1998). Other approaches such as: pairwise deletion, full information maximum likelihood (FIML), hot-deck imputation, mean imputation, similar response pattern imputation, have all been studied with regard to their performance in estimating unbiased estimates from both MCAR and MAR data (Enders & Bandalos, 2001; Jöreskog & Sörbom, 1996; Brown, 1994; Newman, 2003). Additionally, Duncan & Duncan (1994) suggest the need for investigating model misspecification and its effects on multiple imputation techniques.

Despite this rich literature on missing data techniques there is little reported research on correctly specified and misspecified SLD data and their effects on missing data procedures. There has been some work suggesting a Bayesian approach would be effective (Lavori, Dawson, & Shera, 1995) and that FIML works well in such cases (Molenberghs et al., 2004). This paper compares two missing data techniques, multiple imputations (MI) and the full information maximum likelihood (FIML) to see which one performs the best in the presence of SLD with varying degrees of missing data, covariation, and effect sizes.

A. The Model

The model underlying the simulated data is discussed at length elsewhere (Larsen & Willson, unpublished); a typical multilevel is

$$(Score)_{ijt} = \gamma_{00k} + \gamma_{01}(Teacher\ Covariate)_{jt} + \gamma_{10}(Student\ Covariate)_{ijt} + \delta_{0jt} + \epsilon_{ijt} \tag{4.1}$$

where i is the student, j is the teacher, and t is the year. *Score* is the result of an outcome test, *Student Covariate* is any covariate(s) of interest at the first level or student level, and *Teacher Covariate* is any covariate(s) at the second or teacher level. While the covariates are not necessary for this paper, they are included as typical for two-level designs.

1. Correlated Cluster Model

A common modeling process that captures the SLD relationship well is an autoregressive process with one lag *AR(1)*:

$$Average(Score)_{jt} = (1-\rho)\mu + \rho\{Average(Score)_{j(t-1)}\} + \epsilon_{jt}, \tag{4.2}$$

where ρ is the lag coefficient (Ruppert, 2004). The path diagram shown in Figure 5. This model can be generalized to have 'teacher' mean scores at the second level with any number of dependent variables or any number of exogenous predictors with no loss of generalizability given sufficient degrees of freedom under SEM. I limit our discussion here to the univariate dependent variable case.

2. Estimation Methods for Missing Data

Given missing data in the correlated cluster model, this paper explores the use of MI or FIML. The MI approach has been well documented and has exhibited good qualities when a wide variety of models is correctly specified (Rubin, 1996, 2004). FIML is a popular approach for analyzing hierarchical data (Hartley & Rao, 1967). FIML has been compared favorably with pairwise deletion, listwise deletion, and mean imputation in a single level SEM (Enders, 2001; Enders & Bandalos, 2001). Given the disparate set of findings, I wished to determine if there was a clear preference in the SLD case.

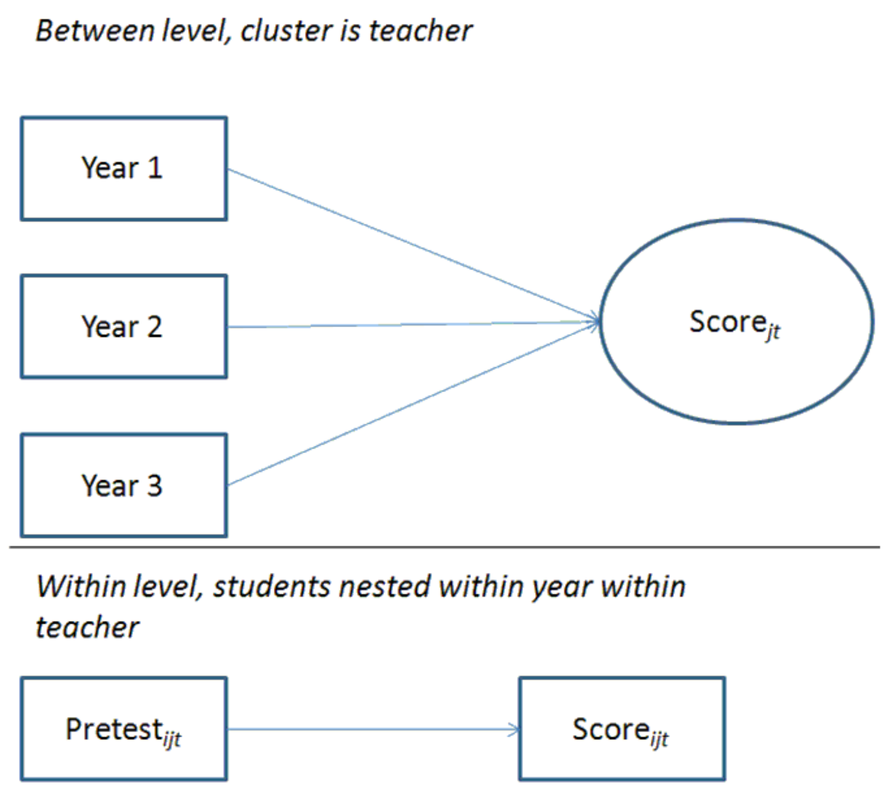

Figure 5 Path model for model with several years with teacher variable independent across years.

Table 11

Variables in the Simulation Study and How They Vary Across the Simulation

Variable	Setting	Effect Size (σ)
Teachers	20,35	'Quality' 0,1
Students	20	'Ability' 0,1
Years	3,5,9	0
Percent Missing	0,25,50	NA
Missing Data Conditions	MCAR, MAR	NA
Missing Data Technique	MI, FIML	NA

B. Method

The author conducts a simulation study regarding the behavior of the bias and standard errors for a 2x2x2x3x2x3x2 factorial design. The data was simulated in SAS PROC IML (SAS, 1999). Table 11 shows the design conditions. To represent a more realistic design predictors at both first and second levels are included.

The first factor in the study is the number of 'teachers' as either 20 or 35, reflecting a medium to large study. The second factor is teacher quality effect size as either effect size '0', where teachers are homogeneous and thus their 'quality' scores are a random draw from a standard normal distribution *N(0,1)* or the effect size of '1' which refers to the case when half of the teachers have a superior teaching quality, drawn from *N(1,1)*, while the rest of the teachers are drawn from *N(0,1)*. The third factor is the 'student' ability with pretest scores drawn from *N(0,1)*, or half of the students in each class pretest scores drawn from a *N(1,1)* distribution and half from *N(0,1)*. The fourth factor is the number of waves of data generated: 3, 5, or 9, with the assumption that there is no improvement or decline in the teacher's exogenous quality scores during the study's duration. The fifth factor is the *AR* condition, with teachers' quality scores covaring between years of the study at '0' or '.5', with the teacher's quality scores drawn according to an *AR(1)* process:

$$(\mathit{Teacher\ Quality})_{jt} = (1-\rho)\mu + \rho(\mathit{Teacher\ Quality})_{j(t-1)} + \epsilon_{jt} \qquad (4.3)$$

as shown in equation 4.3 (Ruppert, 2004), with the *AR(1)* parameter, ρ, of .5. Notice in equation 4.3 'Quality' simply captures the *AR* structure. Without an exogenous variable, the *AR* process is carried by the second level error covariance structure and can be expressed: $\epsilon_t + (1-\rho)\mu + \rho\epsilon_{t-1}$. The sixth factor of the simulation study is a random percent of data taken out at the first level. This comprised the missing data section. Either 0 percent was removed reflecting full data, or 25 percent, or 50 percent removed to show moderate or extreme missing quantities. I also altered the way the data was missing, either MCAR or MAR. I simulated MAR data by correlating high quality teaching with a probability of being missing. Specifically, a worst case scenario was simulated with all those teachers with higher quality scores having all the missing data.

The seventh factor of the simulation study is not a characteristic of the generated data but how the missing data is dealt with, either the MI or the FIML approach. For FIML the PROC MIXED estimation method maximum likelihood in SAS is used. The Restricted Maximum Likelihood (REML) approach is also examined as a subcategory of FIML to see if there were any advantages to either FIML or REML. For the MI approach PROC MI in SAS are run with the covariates at the teacher, and student level included, but year is ignored, which would be appropriate only if the data was independent. This was done because of the limitations that are inherent in PROC MI and other MI programs where random effects and advanced covariance structures cannot be specified. Nevertheless, it has been asserted that the model used in the imputation procedure does not have to be exactly the model used in analysis as long as the multiple imputation model is richer (Meng, 1994; Rubin, 1996). Thus, the richest model possible that made sense was used in the MI step.

One hundred imputations were generated for the 50% missing case, analyzed with PROC MIXED and then combined with PROC MIANALYZE. Forty imputations were used for the 25% missing case. The choice to use a large number of imputations is based on work by Bodner (2008) that concludes that the the earlier suggestions for a low number of imputations are actually inappropriate with a large percentage of missing data. Each of these conditions is simulated with 100 iterations for 0% missing, 50 iterations for 25% missing and 25 iterations for the 50% missing condition because of computational time limitations.

To study the effects of the different missing data techniques the bias, defined as the distance from the estimated statistic to the parameter, is recorded. It was decided to not divide the bias by the true parameter value as is commonly done as many times the parameter value is '0'. The standard deviation of the point estimate, and the average standard error for both the first and second level are also recorded across the differing conditions of the parameters.

C. Results

When considering the results of the study it is first necessary to compare those datasets that have missing data with the dataset that is complete. If the missing data procedures are effective the results should be similar to the full dataset results. The results show that using either FIML or REML estimation techniques had nearly identical results, thus only the FIML results are shown.

First, the variation in the parameter estimates for the covariates is considered. The first level covariate is the ‘ability’ of the students, the second level covariate is the ‘quality’ of the teachers. Table 12 and Table 13 record the average bias, average standard deviation of the estimates, and the average standard errors of the estimates of the covariates as the covariate condition (homogeneous or heterogeneous) itself varies. The tables show that bias, standard deviation, and standard error are all independent of the condition of the covariates. There were no statistically significant interactions.

Table 12

Average Bias, Average Standard Deviation of the Point Estimate, and Average Standard Error for both First and Second Level Parameters Across both MI and FIML Techniques under MCAR and MAR Conditions Varying the Second Level Parameter (Teacher ‘Quality’)

	MCAR: Multiple Imputations					
	Teacher Level ‘Quality’			Student Level ‘Ability’		
’Quality’	bias	σ	S.E.	bias	σ	S.E.
0	0.13	0.16	0.18	0.00	0.04	0.06
1	0.14	0.17	0.18	0.00	0.04	0.06

	MCAR: FIML					
	Teacher Level ‘Quality’			Student Level ‘Ability’		
’Quality’	bias	σ	S.E.	bias	σ	S.E.
0	0.02	0.22	0.23	0.00	0.05	0.05

Table 12 (Continued)

	Teacher Level ‘Quality’			Student Level ‘Ability’		
’Quality’	bias	σ	S.E.	bias	σ	S.E.
1	0.02	0.22	0.23	0.00	0.05	0.05

MAR: Multiple Imputations

	Teacher Level ‘Quality’			Student Level ‘Ability’		
’Quality’	bias	σ	S.E.	bias	σ	S.E.
0	0.06	0.20	0.21	0.00	0.04	0.05
1	0.06	0.20	0.21	0.00	0.04	0.05

MAR: FIML

	Teacher Level ‘Quality’			Student Level ‘Ability’		
’Quality’	bias	σ	S.E.	bias	σ	S.E.
0	0.03	0.23	0.23	0.00	0.04	0.05
1	0.03	0.23	0.23	0.00	0.04	0.05

Table 13

Average Bias, Average Standard Deviation of the Point Estimate, and Average Standard Error for both First and Second Level Parameters Across both MI and FIML Techniques under MCAR and MAR Conditions Varying the First Level Parameter (Student 'Ability')

	MCAR: Multiple Imputations					
	Teacher Level 'Quality'			Student Level 'Ability'		
'Ability'	bias	σ	S.E.	bias	σ	S.E.
0	0.13	0.16	0.18	0.12	0.04	0.06
1	0.14	0.17	0.18	−0.12	0.04	0.06

	MCAR: FIML					
	Teacher Level 'Quality'			Student Level 'Ability'		
'Ability'	bias	σ	S.E.	bias	σ	S.E.
0	0.02	0.22	0.23	0.00	0.05	0.05
1	0.03	0.22	0.23	0.00	0.05	0.05

	MAR: Multiple Imputations					
	Teacher Level 'Quality'			Student Level 'Ability'		
'Ability'	bias	σ	S.E.	bias	σ	S.E.
0	NA	NA	NA	NA	NA	NA
1	0.06	0.20	0.21	0.00	0.04	0.06

Table 13 (Continued)

	MAR: FIML					
	Teacher Level 'Quality'			Student Level 'Ability'		
'Ability'	bias	σ	S.E.	bias	σ	S.E.
0	NA	NA	NA	NA	NA	NA
1	0.03	0.23	0.23	0.00	0.04	0.05

Another variable of interest is the number of repeated measures or years the study went on. The theory being that as years increase both the standard deviation of the point estimate and the standard errors of the point estimates will, on average, decrease. Bias should be unaffected.

Table 14

Average Bias, Average Standard Deviation of the Point Estimate, and Average Standard Error of First and Second Level Parameter Estimates as Years of the Simulation Increases

	MCAR: Multiple Imputations					
	Teacher Level 'Quality'			Student Level 'Ability'		
Years	bias	σ	S.E.	bias	σ	S.E.
3	0.14	0.20	0.22	0.00	0.05	0.07
5	0.14	0.16	0.18	0.00	0.04	0.06
9	0.13	0.13	0.14	0.00	0.03	0.04

	MCAR: FIML					
	Teacher Level 'Quality'			Student Level 'Ability'		
Years	bias	σ	S.E.	bias	σ	S.E.
3	0.03	0.27	0.28	0.00	0.06	0.06
5	0.02	0.22	0.23	0.00	0.05	0.05
9	0.02	0.17	0.18	0.00	0.04	0.04

	MAR: Multiple Imputations					
	Teacher Level 'Quality'			Student Level 'Ability'		
Years	bias	σ	S.E.	bias	σ	S.E.
3	0.07	0.24	0.25	0.00	0.05	0.07
5	0.07	0.20	0.21	0.00	0.04	0.05
9	0.05	0.15	0.16	0.00	0.03	0.04

Table 14 (Continued)

	MAR: FIML					
	Teacher Level 'Quality'			Student Level 'Ability'		
Years	bias	σ	S.E.	bias	σ	S.E.
3	0.03	0.28	0.28	0.00	0.06	0.06
5	0.03	0.22	0.23	0.00	0.04	0.04
9	0.02	0.19	0.18	0.00	0.03	0.03

Table 14 shows that the bias of the covariates was not affected at either the teacher or student level as the number of years increased in the study. The standard deviation and the standard error, on the other hand, decreased as the numbers of years increased. There were no statistically significant interactions of the results when analyzed in PROC MIXED. This follows the large sample theory that as the amount of data increases that estimates become more precise. Bias of the point estimate was not affected.

Another variable of interest is the number of teachers in the study. As with years, the theory states that the standard deviation of the point estimate, and the standard error of the point estimate should decrease as the number of clusters (teachers) increases.

Table 15

Average Bias, Average Standard Deviation of the Point Estimate, and Average Standard Error for both First and Second Level Parameters Across both MI and FIML Techniques under MCAR and MAR Conditions Varying the Number of Teachers in the Simulated Data

	MCAR: Multiple Imputations					
	Teacher Level 'Quality'			Student Level 'Ability'		
Teachers	bias	σ	S.E.	bias	σ	S.E.
20	0.14	0.19	0.20	0.00	0.04	0.06
35	0.13	0.14	0.15	0.00	0.03	0.05

	MCAR: FIML					
	Teacher Level 'Quality'			Student Level 'Ability'		
Teachers	bias	σ	S.E.	bias	σ	S.E.
20	0.03	0.25	0.26	0.00	0.05	0.06
35	0.02	0.19	0.20	0.00	0.04	0.04

	MAR: Multiple Imputations					
	Teacher Level 'Quality'			Student Level 'Ability'		
Teachers	bias	σ	S.E.	bias	σ	S.E.
20	0.06	0.23	0.23	0.00	0.04	0.06
35	0.06	0.18	0.18	0.00	0.03	0.04

Table 15 (Continued)

	MAR: FIML					
	Teacher Level 'Quality'			Student Level 'Ability'		
Teachers	bias	σ	S.E.	bias	σ	S.E.
20	0.03	0.27	0.26	0.00	0.05	0.05
35	0.03	0.19	0.19	0.00	0.04	0.04

Table 15 shows the standard errors and standard deviation of the estimate decreases as the number of teachers increase; bias is unaffected.

The *AR(1)* component ρ is also of interest. The author predicted a priori that the FIML technique would produce estiamates biased towards 0 when the missing data were either MCAR or MAR in its standard error estimation.

Table 16

Average Bias, Average Standard Deviation of the Point Estimate, and Average Standard Error for both First and Second Level Parameters Across both MI and FIML Techniques under MCAR and MAR Conditions Varying the *AR(1)* Component ρ

	MCAR: Multiple Imputations					
	Teacher Level 'Quality'			Student Level 'Ability'		
ρ	bias	σ	S.E.	bias	σ	S.E.
0.0	0.13	0.14	0.15	0.00	0.04	0.06
0.5	0.14	0.20	0.21	0.00	0.04	0.06

Table 16 (Continued)

	MCAR: FIML					
	Teacher Level 'Quality'			Student Level 'Ability'		
ρ	bias	σ	S.E.	bias	σ	S.E.
0.0	0.02	0.18	0.19	0.00	0.05	0.05
0.5	0.03	0.26	0.27	0.00	0.05	0.05

	MAR: Multiple Imputations					
	Teacher Level 'Quality'			Student Level 'Ability'		
ρ	bias	σ	S.E.	bias	σ	S.E.
0.0	0.06	0.17	0.17	0.00	0.04	0.05
0.5	0.07	0.24	0.24	0.00	0.04	0.05

	MAR: FIML					
	Teacher Level 'Quality'			Student Level 'Ability'		
ρ	bias	σ	S.E.	bias	σ	S.E.
0.0	0.02	0.19	0.19	0.00	0.05	0.05
0.5	0.03	0.27	0.19	0.00	0.04	0.04

Table 16 shows the effect of varying the AR time component ρ had on the estimates of the covariates of interest. The student level 'ability' was robust with respect to a nonzero AR parameter, which follows intuitively as the student scores are independent across the years. The standard error of the teacher level 'quality' increased with a nonzero AR. This is expected because there will be less information available to calculate estimates if the data are correlated. Surprisingly, the estimates produced by FIML of the standard error when the missing data is MAR were not biased toward zero but actually increased as they should.

One of the main purposes of this study was to observe the effect that missing data has on the estimation of the parameters in a multilevel parameter when FIML or multiple imputation techniques were used. The ideal case would be for the missing data technique to correctly estimate the parameter (low bias), and to have the same standard errors that are estimated when no data is missing.

Table 17

Average Bias, Average Standard Deviation of the Point Estimate, and Average Standard Error of First and Second Level Parameter Estimates as the Percent of Missing Data Increases

	MCAR: Multiple Imputations					
	Teacher Level 'Quality'			Student Level 'Ability'		
% missing	bias	σ	S.E.	bias	σ	S.E.
0%	0.02	0.22	0.24	0.00	0.04	0.04
25%	0.13	0.17	0.18	0.00	0.04	0.06
50%	0.25	0.11	0.13	0.00	0.04	0.07

	MCAR: FIML					
	Teacher Level 'Quality'			Student Level 'Ability'		
% missing	bias	σ	S.E.	bias	σ	S.E.
0%	0.02	0.22	0.23	0.00	0.04	0.04

Table 17 (Continued)

	Teacher Level 'Quality'			Student Level 'Ability'		
% missing	bias	σ	S.E.	bias	σ	S.E.
25%	0.03	0.22	0.23	0.00	0.04	0.05
50%	0.03	0.22	0.23	0.00	0.06	0.06

MAR: Multiple Imputations

	Teacher Level 'Quality'			Student Level 'Ability'		
% missing	bias	σ	S.E.	bias	σ	S.E.
0%	0.02	0.23	0.23	0.00	0.04	0.04
25%	0.06	0.20	0.20	0.00	0.04	0.05
50%	0.10	0.18	0.19	0.00	0.04	0.06

MAR: FIML

	Teacher Level 'Quality'			Student Level 'Ability'		
% missing	bias	σ	S.E.	bias	σ	S.E.
0%	0.02	0.23	0.23	0.00	0.04	0.04
25%	0.03	0.23	0.23	0.00	0.04	0.04
50%	0.03	0.23	0.23	0.00	0.05	0.05

Table 17 shows the effect that missing data had on the estimates of the covariates. The student level 'ability' estimate was fairly robust against increasing missing data. Nevertheless, there was a slight increase in the standard errors as percent of missing data increased. This is expected as there should be extra variability in the estimates to reflect the less certain nature of the data with a higher percentage of the data missing. The teacher level ‘quality’ showed a troubling increase in bias, and a decrease in the standard errors as compared to the full data case. The fact that the cases with more missing data had a smaller standard error shows than estimates with a SLD and a high degree of missingness estimated with multiple imputations have inappropriately high power for hypothesis testing.

Another important purpose to discuss is what, if any, are the interactions between the variables. Using PROC MIXED on the results, one interaction was found. The estimate of the standard errors depended on both the percent of missing data and the value of the *AR(1)* or ρ component as seen in Table 18.

Table 18

Average Standard Error of the First and Second Level Parameter Estimate as both Percent Missing and the ρ Parameter Vary

	MCAR: Multiple Imputations			
	Teacher Level 'Quality'		Student Level 'Ability'	
% missing	ρ=0	ρ=0.5	ρ=0	ρ=0.5
0%	0.19	0.26	0.04	0.04
25%	0.15	0.20	0.06	0.06
50%	0.11	0.15	0.07	0.07

	MCAR: FIML			
	Teacher Level 'Quality'		Student Level 'Ability'	
% missing	ρ=0	ρ=0.5	ρ=0	ρ=0.5
0%	0.19	0.26	0.04	0.04

Table 18 (Continued)

	Teacher Level 'Quality'		Student Level 'Ability'	
% missing	ρ=0	ρ=0.5	ρ=0	ρ=0.5
25%	0.19	0.26	0.04	0.04
50%	0.19	0.26	0.04	0.04

MAR: Multiple Imputations

	Teacher Level 'Quality'		Student Level 'Ability'	
% missing	ρ=0	ρ=0.5	ρ=0	ρ=0.5
0%	0.19	0.26	0.04	0.04
25%	0.17	0.24	0.05	0.05
50%	0.16	0.21	0.06	0.06

MAR: FIML

	Teacher Level 'Quality'		Student Level 'Ability'	
% missing	ρ=0	ρ=0.5	ρ=0	ρ=0.5
0%	0.19	0.27	0.04	0.04
25%	0.19	0.27	0.04	0.04
50%	0.19	0.27	0.05	0.05

This problem of an inappropriately high power when multiple imputations are used to deal with missing data is exacerbated with a higher ρ component as can be seen in Table 18 and in Figure 6. Table 17 shows that the student level 'ability' estimate is free from an interaction but the teacher level 'quality' estimate is not.

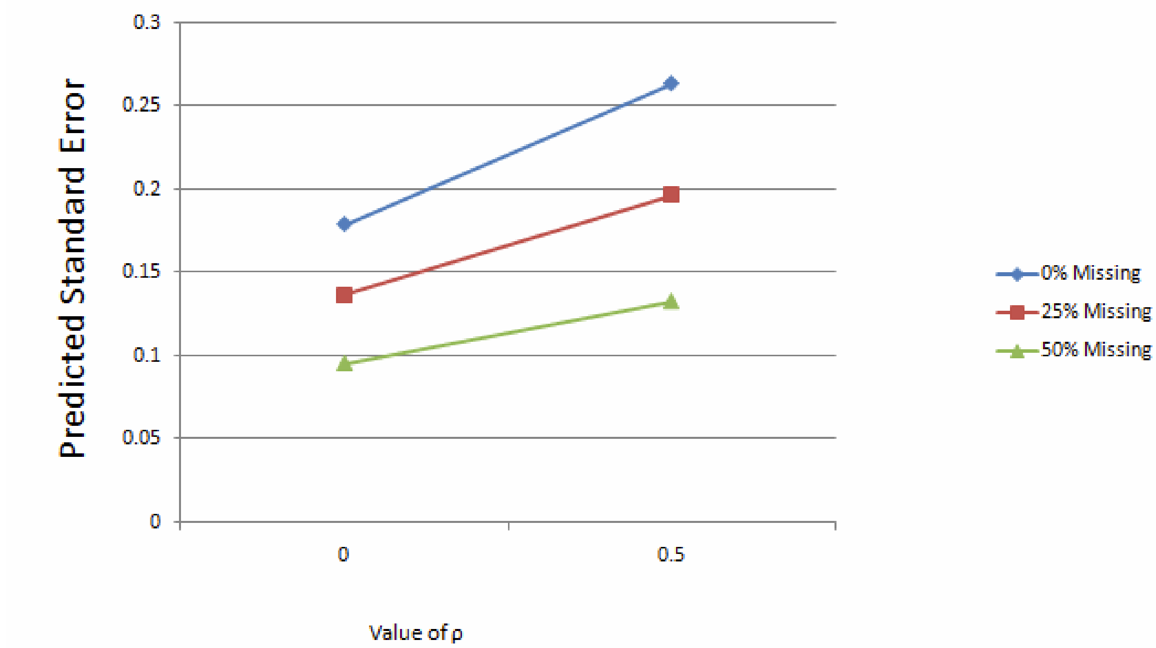

Figure 6 Standard error of second level parameter (teacher 'quality') when multiple imputations are used to estimate missing data, graphed across ρ and percent missing.

The 0% line in Figure 6 shows the best approximation and any deviations from it can be thought of as design-dependent. There is a straight forward reduction in the estimate of standard errors in the case where ρ=0. This difference becomes wider as ρ increases to 0.5. Thus, even without an SLD the MI procedure does not handle imputing the data appropriately. This is due to the fact that PROC MI does not include random effects, so that all the student level data is assumed to come from the same classroom when running PROC MI. This problem becomes worse if there is any SLD. This situation is not mirrored when using the FIML approach either with MAR or MCAR data. Thus, it seems clear that FIML in these situations is superior to MI techniques as implemented in SAS.

D. Discussion

The results of this study indicate that it is very important to realize that the MI approach for dealing with missing data may give underestimates for the standard error and increase the bias of the estimate. This could seriously impact the results of studies with small effect sizes such as those common in the social sciences. Thus, as many other research conditions have indicated (Enders & Bandalos, 2001), the FIML approach is robust to the SLD deficiencies that MI exhibits. This is probably due to the fact that when using the FIML of PROC MIXED it is possible to specify the correct model while in PROC MI it is only possible to generate an approximation for that model. As long as the model used in the MI is more general than the model used in analyzing the data good results should be produced (Schafer, 2005, Nov). The model specified in the MI step could be argued as a more general model as it does not take year into account, but as noted previously this generalization does not take into account the particular covariance effects. It should be noted that PROC MIXED, while correctly handling MAR data as shown, will throw out any data missing any of the values for the covariates, which assumes the data is MCAR. This will also lead to incorrect estimates if the data is in fact MAR. Further work needs to be done to explore that scenario using PROC MIXED to analyze data. Otherwise, Larsen & Willson (unpublished) have shown that LISREL competes favorably with SAS in its estimation of models with SLD. Therefore, in the highly probable event of missing data for the covariates as well as the response, LISREL should be used with FIML in analyzing SEM data. In the future if MI techniques are still going to be useful in the missing data field, additional error structures, random effects, and a wider array of models need to be engineered into common MI procedures.

CHAPTER V

CONCLUSION

The best alternative for an SEM with latent variables and/or mediating relationships with a SLD is MULTILEV. This fact is important when choices are made on which software programs to invest time and money on and which software packages to choose for classroom instruction. Choosing the right software can save time and resources which would be better allocated to study design and theory. Additionally, because the differences between programs, it is important to know how to get equivalent results from different programs (Albright & Park, 2008; Chantala & Suchindran, 2006; Kenny, 2007, May) or from different procedures within the same programs (Gao, Thompson, Xiong, & Miller, 2009) so the results can be validated.

In the educational literature with longitudinal studies model misspecification with missing independent variables can be a problem (Dewey, Husted, & Kenny, 2000), as well as misspecification of the structure of the model which the fit indices may not detect (Fan & Sivo, 2005, 2007). To add on top of these common problems an SLD that is not correctly accounted for can lead to further bias in the parameter and standard error estimations. Thus, it behooves the researcher to use the software that has the capability to model reality as close as possible. Because practical and statistical significance are big concerns in educational research (Fan, 2001), correct standard errors are absolutely essential for making correct policy decisions. Correct standard errors are also necessary for sample size calculations (Snijders & Bosker, 1993).

Missing data and missing data techniques are still being studied (Acock, 2005), and SLD is not heavily emphasized in the literature. Thus, an incorrect approach can cause standard error deflation and lead to inappropiate conclusions and poor educational policy decisions. The bias that occurs if a SLD is not modeled correctly and the bias that occurs by using incorrect missing data techniques may be small in an individual study but as many policy decisions are made on results of meta-analysis (Kavale & Forness, 2000), small errors will add up to inflated effect sizes and incorrect expectations.

A. Limitations and Future Research

This dissertation has several limitations. First, only three computer programs are compared explicitly. Other programs may have the same functionality of SAS and the same flexibility of LISREL. Other limitations have to deal with the data structure itself. There was no attempt to have variation at higher levels that would represent the school such as the district. Additionally, the study assumed that the teacher's quality remains constant through time which denies the fact that teachers can improve in time, and ignored any analytic techniques that takes improvement into account. Any trends in class assignment such as giving poorly performing students to younger teachers are ignored. Class assignment is assumed to be perfectly random which is not reality. Also, the whole simulation study was not replicated in the other programs. If it had been, perhaps other idiosyncrasies in Mplus and LISREL might have been discovered.

The missing data study is limited insomuch that only MCAR and MAR missing data patterns are considered and one MAR situation is considered. Also, other missing data techniques besides MI and FIML are not considered. Other programs who claim to use FIML like Mplus was not tested.

Both simulation studies assumed normal data which in reality may not be true as data are often binary, ordinal, multinomial etc. Future research needs to be done with missing data and a SLD in the absence of the normality assumption.

All these limitations are opportunities for future research. It would be interesting to simulate data in other programs such as Mplus or LISREL and see if the results are constant across all conditions. Additionally, other multiple imputation programs besides PROC MI in SAS could be considered for use in dealing with missing data.

REFERENCES

Acock, A. (2005). Working with missing values. *Journal of Marriage and Family*, *67*(4), 1012–1028.

Aitkin, M., & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society. Series A (General)*, *149*(1), 1–43.

Albright, J., & Park, H. (2008). *Confirmatory factor analysis using Amos, LISREL, Mplus, SAS/STAT CALIS*. Technical Working Paper. The University Information Technology Service (UITS) Center for Statistical and Mathematical Computing, Bloomington, IN: Indiana University.
URL `http://www.nanoogi.org/documents/cfa.pdf`

Allison, P. (2002). *Missing data*. Thousand Oaks, CA: Sage.

Bentler, P. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology*, *31*(1), 419–456.

Bodner, T. (2008). What improves with increased missing data imputations? *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(4), 651–675.

Brown, R. (1994). Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods. *Structural Equation Modeling: A Multidisciplinary Journal*, *1*(4), 287–316. DOI: 10.1080/10705519409539983.

Bryk, A., & Raudenbush, S. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.

Bryk, A., Raudenbush, S., Congdon, R., & Seltzer, M. (1988). *Introduction to HLM: computer program & user's guide*. Chicago: University of Chicago.

Burstein, L., Kim, K., & Delandshere, G. (1989). *Multilevel investigations of systematically varying slopes: Issues, alternatives, and consequences*. San Diego, CA: Academic Press.

Busing, F., Meijer, E., & van der Leeden, R. (2005). *Software for multilevel analysis of data with two levels users guide for Version 4.1*. Leiden, The Netherlands: Leiden University.

Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods*, *1*(4), 421–483. DOI: 10.1177/109442819814004.

Chantala, K., & Suchindran, C. (2006). Adjusting for unequal selection probability in multilevel models: A comparison of software packages. *Proceedings of the Survey Research Methods Section*, (pp. 2815–2824). 2006, Seattle, WA: American Statistical Association.

De Leeuw, J., & Kreft, I. (2001). Software for multilevel analysis. ln H. Goldstein (Eds.) *Multilevel Modeling of Health Statistics.* (pp. 187–204). New York: Wiley Chichester.

Dewey, J., Husted, T., & Kenny, L. (2000). The ineffectiveness of school inputs: a product of misspecification? *Economics of Education Review*, *19*(1), 27–45. DOI: 10.1016/S0272-7757(99)00015-1.

Duncan, S., & Duncan, T. (1994). Modeling incomplete longitudinal substance use data using latent variable growth curve methodology. *Multivariate Behavioral Research*, *29*(4), 313–338.

Enders, C. (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement*, *61*(5), 713.

Enders, C., & Bandalos, D. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(3), 430–457.

Fahrmeir, L., Tutz, G., & Hennevogl, W. (2001). *Multivariate statistical modeling based on generalized linear models*. New York: Springer-Verlag.

Fan, X. (2001). Statistical significance and effect size in education research: Two sides of a coin. *The Journal of Educational Research*, *94*(5), 275–282.

Fan, X., & Sivo, S. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(3), 343–367.

Fan, X., & Sivo, S. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, *42*(3), 509–529.

Gao, F., Thompson, P., Xiong, C., & Miller, J. (2009). Analyzing multivariate longitudinal data using SAS. *Proceedings of the Thirty-first Annual SAS Users Group International Conference*. Paper 209-31, March 2009, Cary, NC: SAS Institute Inc. URL `http://tiny.cc/gnxxt`

Goldstein, H. (1987). *Multilevel models in education & social research*. London: Griffin.

Goldstein, H. (1993). *Multilevel statistical models*. New York: Halstead Press.

Hartley, H. O., & Rao, J. N. K. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*, *54*(1/2), 93–108.

Hedges, L. (2009). Adjusting a significance test for clustering in designs with two levels of nesting. *Journal of Educational and Behavioral Statistics*, *34*(4), 464–490. DOI: 10.3102/1076998609337251.

Hox, J. (1995). *Applied multilevel analysis*. Amsterdam, The Netherlands: TT-Publikaties.

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, *5*(3), 299–314.

Jöreskog, K., & Sörbom, D. (1996). *PRELIS 2: User's reference guide*. Lincolnwood, IL: Scientific Software International.

Jöreskog, K., & Sörbom, D. (2005). *LISREL for windows*. Lincolnwood, IL: Scientific Software International.

Jöreskog, K., Sörbom, D., & Du Toit, S. (2000). *LISREL 8: New statistical features*. Lincolnwood, IL: Scientific Software International.

Kavale, K., & Forness, S. (2000). *Policy decisions in special education: The role of meta-analysis*. Mahwah, NJ: Erlbaum.

Kenny, D. (2007, May). Conventional multilevel modeling: SAS and SPSS. Paper presented at the *Intergroup Data as Modeled Using the Social Relations Model Conference* (pp. 1-22). Storrs CT: National Science Foundation. URL `http://davidakenny.net/doc/kenny.doc`

Kreft, I. G. (1987). Models and methods for the measurement of school effects. Unpublished doctoral dissertation. Amsterdam, The Netherlands: University of Amsterdam.

Kreft, I. G., De Leeuw, J., & Kim, K.-S. (1990). Comparing four different statistical packages for hierarchical linear regression: GENMOD, HLM, ML2, and VARCL. Tech. rep., Los Angeles, CA: CSE Dissemination Office, UCLA Graduate School of Education.

Kwok, O., West, S., & Green, S. (2007). The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A monte carlo study. *Multivariate Behavioral Research*, *42*(3), 36. DOI: 10.1080/00273170701540537.

Larsen, R., & Willson, V. (unpublished). Modeling second level cluster dependency in multilevel models. Submitted for Publication.

Lavori, P., Dawson, R., & Shera, D. (1995). A multiple imputation strategy for clinical trials with truncation of patient data. *Statistics in Medicine*, *14*(17), 1913–1925. DOI: 10.1002/sim.4780141707.

Leeuw, J. D., & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, *11*(1), 57–85.

Littell, R., Milliken, G., Stroup, W., & Wolfinger, R. (1996). *SAS system for mixed models*. Cary, NC: SAS Institute.

Lunn, D., Thomas, A., Best, N., & Spiegelhalter, D. (2000). Winbugs-a bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*(4), 325–337. DOI: 10.1023/A:1008929526011.

Maas, C., & Hox, J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *1*, 86–92.

Mason, W. M., Wong, G. Y., & Entwisle, B. (1983). Contextual analysis through the multilevel linear model. *Sociological Methodology*, *14*, 72–103.

Meng, X. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science*, *9*(4), 538–558.

Michela, J. (2006). Software review: HLM 6. *Organizational Research Methods*, *9*(1), 119.

Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M. G., Mallinckrodt, C., & Carroll, R. J. (2004). Analyzing incomplete longitudinal clinical trial data. *Biostat*, *5*(3), 445–464.

Muthen, B. (1997). Latent variable modeling of longitudinal and multilevel data. *Sociological Methodology*, *27*(1), 453–480.

Muthén, L., & Muthén, B. (1998-2007). *Mplus user's guide. (5th ed.)*. Los Angeles, CA: Muthén and Muthén.

Newman, D. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods*, *6*(3), 328–362. DOI: 10.1177/1094428103254673.

Prosser, R., Rasbach, J., & Goldstein, H. (1991). *ML3 software for two-level analysis: Users guide*. London: Institute of Education.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2001). *GLLAMM manual.* Tech. rept. 2001/01. Department of Biostatistics and Computing, Institute of Psychiatry, Kings College, University of London.
URL `http://www.gllamm.org`

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Woodhouse, G., Healy, M., Draper, D., Langford, I., & Lewis, T. (2000). *A user's guide to MLwiN*. London: Multilevel Models Project, Institute of Education, University of London.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. DOI: 10.1093/biomet/63.3.581.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*(434), 473–89.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Ruppert, D. (2004). *Statistics and finance: An introduction*. New York: Springer.

SAS (1999). *SAS Manual, release 8*. Cary, NC: SAS.

Schafer, J. (2005, Nov). *Missing data in longitudinal studies: A review.* Paper presented at the AAPS, Nov 2005, Nashville, Tennessee.
URL `http://www.stat.psu.edu///~jls/aaps/_schafer.pdf`

Singer, J. (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. DOI: 10.3102/10769986023004323.

Snijders, T., & Bosker, R. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational and Behavioral Statistics*, *18*(3), 237–259. DOI:10.3102/10769986018003237.

Stapleton, L. (2006). An assessment of practical solutions for structural equation modeling with complex sample data. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(1), 28–58.

Stevens, J., & Taylor, A. (2008). Hierarchical dependence in meta-analysis. *Journal of Educational and Behavioral Statistics*, *34*(1), 46–73. DOI: 10.3102/1076998607309080.

Wong, G. Y., & Mason, W. M. (1989). *Ethnicity, comparative analysis, and generalization of the hierarchical normal linear model for multilevel analysis*. Research Rep. No. 89-138. Ann Arbor, MI: Population Studies Center, University of Michigan.

Zhou, X.-H., Perkins, A. J., & Hui, S. L. (1999). Comparisons of software packages for generalized linear multilevel models. *The American Statistician*, *53*(3), 282–290.

APPENDIX A

CODE FOR SIMULATING DATA IN SAS PROC IML

```
%macro Sim1(iter=, tmax=,smax=,ymax=,cmax=
,qmax= ,amax= ,yemax=, cemax=,missing=,);


proc iml;
nit=&iter;
tmax=&tmax;
smax=&smax;
ymax=&ymax;
cmax=&cmax;
qmax=&qmax;
amax=&amax;

yemax=&yemax;
cemax=&cemax;
miss=&missing;
/*setting up parameters for model*/
total=tmax*smax*3*cmax*qmax+1*amax+1*yemax+1*cemax+1;
data=j(smax,14,-9);
final=j(1,14,-9);
scenario=1;
group=0;
achieve=0;
iteration=0;
do iteration=1 to nit;
do timeeffect=0 to 1;
yeareffect=rand('normal',timeeffect,1);

do ability=0 to amax;
do quality=0 to qmax;
do teacher=1 to tmax;
if teacher<(tmax+1)/2 then group=0;
else group=1;
do year=0 to ymax;
if year=0 then
teachereffect=rand('normal',0,1)+(quality*group);
else teachereffect=(teachereffect-(quality*group))
*timeeffect/2+(quality*group)
+rand('normal',0,(1-(timeeffect/2)**2)**(1/2));

do student=1 to smax;
if student<smax/2 then achieve=0;
```

```
else achieve=1;
data[student,1]=rand('normal',0,1)+(ability*achieve);
data[student,2]=ability;
data[student,3]=teacher;
data[student,4]=teachereffect;
data[student,5]=data[student,1]+data[student,4];
data[student,6]=quality;
data[student,7]=year;
data[student,8]=yeareffect;
data[student,9]=timeeffect/2;
data[student,10]=teachereffect*rand('normal',0,1);
data[student,11]=group;
data[student,12]=achieve;
data[student,13]=iteration;
data[student,14]=student;

end;
final=final//data;

end;

end;

end;

end;

end;

print iteration;

end;

create data from final;
append from final;
quit;

data data;
set data;
if _n_ ne 1;
run;

data analyze (rename= (COL1=student COL2=ability
COL3=teacher COL4=teachereffect COL5=score COL6=quality COL7=year
COL8=yeareffect COL9=timeeffect COL10=scenario
 COL11=group COL12=achieve COL13=iteration COL14=ID));
set data;
run;

data analyze&tmax&smax&ymax&missing;
set analyze;
```

```
run;


proc sort data=analyze;
by  quality ability timeeffect iteration ;
run;


ODS output ConvergenceStatus=Converge;
ODS output SolutionF=fit;
ods listing close;

proc mixed data=analyze;
by iteration timeeffect ability quality NOTSORTED;
class teacher year ;
model score= achieve group/ solution;
random intercept achieve/  sub=teacher ;
random year / sub=teacher type=AR(1) ;
ods output SolutionF=mxparms;
run;
ods listing;



/*Calculating the mean bias for achieve*/

data meanbiasachieve;
set fit;
if Effect ^= 'achieve' then delete;
run;

ODS output Moments=AchieveMeanBias;
ods listing close;
proc univariate data=meanbiasachieve;
by  timeeffect ability quality NOTSORTED;
var estimate Stderr;
run;
ods listing;


data achieveBias&tmax&smax&ymax&missing;
set AchieveMeanBias;
if Label1 = 'Mean'  then select=1;
if Label2 ='Variance' then select=1;
if select ^=1 then delete;
calcvalue=cvalue1;
Bias=cvalue1-ability;
iterations=&iter;
teachers=&tmax;
students=&smax;
years=&ymax;
```

```
missing=&missing;
keep Varname iterations timeeffect ability
quality teachers students years  calcvalue Bias Label1;
run;

/*Calculating the mean bias for group
which is teacher's ability*/

data meanbiasgroup;
set fit;
if Effect ^= 'group' then delete;
run;

ODS output Moments=GroupMeanBias;
ods listing close;
proc univariate data=meanbiasgroup;
by  timeeffect ability quality NOTSORTED;
var estimate Stderr;
run;
ods listing;


data GroupBias&tmax&smax&ymax&missing;
set GroupMeanBias;
if Label1 = 'Mean'  then select=1;
if Label2 ='Variance' then select=1;
if select ^=1 then delete;
calcvalue=cvalue1;
Bias=cvalue1-quality;
iterations=&iter;
teachers=&tmax;
students=&smax;
years=&ymax;
missing=&missing;
keep Varname iterations timeeffect ability
quality teachers students years  calcvalue Bias Label1;
run;



/*Coverage bias*/

ODS output Moments=MeanConverge;
ods listing close;
proc univariate data=Converge;
by  timeeffect ability quality NOTSORTED;
var status pdG pdH;
run;
ods listing;
```

```
ODS output Moments=MeanConverge;
ods listing close;
proc univariate data=Converge;
by  timeeffect ability quality NOTSORTED;
var status pdG pdH;
run;
ods listing;


data MeanConverge&tmax&smax&ymax&missing;
set MeanConverge;
if Label1 ^= 'Mean' then delete;
percent_convergence=(1-cvalue1)*100;
iterations=&iter;
teachers=&tmax;
students=&smax;
years=&ymax;
missing=&missing;
keep Varname iterations timeeffect ability
quality teachers students years percent_convergence ;
run;
quit;
%mend Sim1;
/**/
%Sim1 (iter=100,tmax=20,smax=20,ymax=1,cmax=2
,qmax=1, amax=1,yemax=1, cemax=1);
%Sim1 (iter=100,tmax=20,smax=20,ymax=3,cmax=2,
qmax=1, amax=1,yemax=1, cemax=1);
%Sim1 (iter=100,tmax=20,smax=20,ymax=7,cmax=2,
qmax=1, amax=1,yemax=1, cemax=1);
%Sim1 (iter=100,tmax=35,smax=20,ymax=1,cmax=2,
qmax=1, amax=1,yemax=1, cemax=1);
%Sim1 (iter=100,tmax=35,smax=20,ymax=3,cmax=2,
qmax=1, amax=1,yemax=1, cemax=1);
%Sim1 (iter=100,tmax=35,smax=20,ymax=7,cmax=2,
qmax=1, amax=1,yemax=1, cemax=1);
```

APPENDIX B

CODE FOR ANALYZING DATA IN MULTILEV, LISREL, MPLUS, AND SAS

SAS Code:

```
proc mixed data=analyze;
class teacher year ;
model score= achieve group/ solution;
random intercept achieve/  sub=teacher;
random year / sub=teacher type=AR(1);
run;
```

Mplus Code:

```
TITLE:  Simulation

  DATA:
    FILE IS "G:\ESPY\Simulation\URGEN\Data_for_MPLUS.dat";

  VARIABLE:
    NAMES ARE Teacher ID group achieve y1 y2 y3 y4;

    USEVARIABLES ARE Teacher  group achieve y1-y4;
    CLUSTER IS Teacher;
  within=achieve;
  between=group;

  ANALYSIS:
    TYPE IS TWOLEVEL;
    ESTIMATOR IS ML;
    ITERATIONS = 1000;
    CONVERGENCE = 0.00005;

    Model:
    %WITHIN%
    iw | y1@0 y2@1 y3@2 y4@3;
                    y1-y4 (1);
         Y1 WITH Y4@0;
         Y2 WITH Y3@0;
         Y2 WITH Y4@0;
         Y3 WITH Y4@0;
         iw on achieve;


  %BETWEEN%
  ib | y1@0 y2@1 y3@2 y4@3;
```

```
          Y1-Y4 (2);
          y1 with y2 (3);
          y2 with y3 (3);
          y3 with y4 (3);
  ib ON group
output:     sampstat stand;
```

MULTILEV Code:

```
  OPTIONS OLS=YES CONVERGE=0.001000 MAXITER=100 EFFECTS=YES
  OUTPUT=STANDARD ;
  TITLE=Multilevel;
  SY='G:\ESPY\Simulation\Eightyear Lisrel\EYDA.psf';
  ID2=teacher;
  RESPONSE=score;
  FIXED=intcept group achieve;
  DUMMY=year;
  RANDOM1=intcept achieve;
  RANDOM2=dummy1 dummy2 dummy3 dummy4 dummy5 dummy6 dummy7 dummy8;
 COV2PAT=
1
 2 1
 4 2 1
 8 4 2 1
 16 8 4 2 1
 32 16 8 4 2 1
 64 32 16 8 4 2 1
128 64 32 16 8 4 2 1;
```

APPENDIX C

CODE FOR SIMULATING DATA, SIMULATING MISSING DATA CONDITIONS (MCAR AND MAR) AND ANALYZING RESULTS IN SAS PROC IML AND SAS PROC MIXED

```
%macro Sim1(iter=, tmax=,smax=,ymax=,cmax=,qmax= ,
amax= ,yemax=, cemax=,missing=,);


proc iml;
nit=&iter;
tmax=&tmax;
smax=&smax;
ymax=&ymax;
cmax=&cmax;
qmax=&qmax;
amax=&amax;

yemax=&yemax;
cemax=&cemax;
miss=&missing;
/*setting up parameters for model*/
total=tmax*smax*3*cmax*qmax+1*amax+1*yemax+1*cemax+1;
data=j(smax,13,-9);
final=j(1,13,-9);
scenario=1;
group=0;
achieve=0;
iteration=0;
do iteration=1 to nit;
do timeeffect=0 to 1;
yeareffect=rand('normal',timeeffect,1);

do ability=0 to amax;
do quality=0 to qmax;
do teacher=1 to tmax;
if teacher<(tmax+1)/2 then group=0;
else group=1;
do year=0 to ymax;
if year=0 then
teachereffect=rand('normal',0,1)+(quality*group);
else teachereffect=(teachereffect-(quality*group))
*timeeffect/2+(quality*group)
+rand('normal',0,(1-(timeeffect/2)**2)**(1/2));
```

```
do student=1 to smax;
if student<smax/2 then achieve=0;
else achieve=1;
data[student,1]=rand('normal',0,1)+(ability*achieve);
data[student,2]=ability;
data[student,3]=teacher;
data[student,4]=teachereffect;
r=RAND('UNIFORM');
if r<(1-1/miss) then
data[student,5]=data[student,1]+data[student,4];
else data[student,5]=-9999;
data[student,6]=quality;
data[student,7]=year;
data[student,8]=yeareffect;
data[student,9]=timeeffect/2;
data[student,10]=teachereffect*rand('normal',0,1);
data[student,11]=group;
data[student,12]=achieve;
data[student,13]=iteration;

end;
final=final//data;

end;

end;

end;

end;

end;

print iteration;

end;

create data from final;
append from final;
quit;

data data;
set data;
if _n_ ne 1;
run;

data analyze1 (rename= (COL1=student COL2=ability
COL3=teacher COL4=teachereffect COL5=score
COL6=quality COL7=year
COL8=yeareffect COL9=timeeffect COL10=scenario
```

```
COL11=group COL12=achieve COL13=iteration ));
set data;
run;


data analyze2;
set analyze1;
if score=-9999 then
call missing(score);
run;



proc mi data=analyze2  out=analyze NIMPUTE=100 method=ML;
by iteration timeeffect ability quality NOTSORTED;
mcmc chain=multiple displayinit initial=em(itprint);
      var  achieve group score;
   run;


proc sort data=analyze;
by  quality ability timeeffect iteration ;
run;
/**/
/*proc print data=analyze;*/
/*run;*/
/*ODS output ConvergenceStatus=Converge;*/
/*ODS output SolutionF=fit;*/


proc mixed data=analyze method=ML;
by iteration timeeffect ability
quality NOTSORTED _Imputation_;
class teacher year ;
model score= achieve group/ solution;
random intercept achieve/  sub=teacher;
random year / sub=teacher type=AR(1);
ods output SolutionF=mxparms;
run;


ODS output ParameterEstimates=fit1;

  proc mianalyze parms(classvar=full)=mxparms;
  by iteration timeeffect ability quality NOTSORTED;
      class teacher;
      modeleffects Intercept achieve group;
   run;
```

```
data fit;
set fit1;
Effect=PARM;
drop PARM;
run;


/*Calculating the mean bias for achieve*/

data meanbiasachieve;
set fit;
if Effect ^= 'achieve' then delete;
run;

ODS output Moments=AchieveMeanBias;
ods listing close;
proc univariate data=meanbiasachieve;
by  timeeffect ability quality NOTSORTED;
var estimate;
run;
ods listing;

data achieveBias&tmax&smax&ymax&missing;
set AchieveMeanBias;
if Label1 ^= 'Mean' then delete;
calcvalue=cvalue1;
Bias=cvalue1-ability;
iterations=&iter;
teachers=&tmax;
students=&smax;
years=&ymax;
missing=&missing;
keep Varname iterations timeeffect ability quality
teachers students years missing calcvalue Bias ;
run;

/*Calculating the mean bias for
group which is teacher's ability*/

data meanbiasgroup;
set fit;
if Effect ^= 'group' then delete;
run;

ODS output Moments=GroupMeanBias;
ods listing close;
proc univariate data=meanbiasgroup;
by  timeeffect ability quality NOTSORTED;
var estimate;
run;
```

```
ods listing;


data GroupBias&tmax&smax&ymax&missing;
set GroupMeanBias;
if Label1 ^= 'Mean' then delete;
calcvalue=cvalue1;
Bias=cvalue1-quality;
iterations=&iter;
teachers=&tmax;
students=&smax;
years=&ymax;
missing=&missing;
keep Varname iterations timeeffect
 ability quality teachers
students years missing calcvalue Bias ;
run;



/*Coverage bias*/

/*ODS output Moments=MeanConverge;*/
/*ods listing close;*/
/*proc univariate data=Converge;*/
/*by  timeeffect ability quality NOTSORTED;*/
/*var status pdG pdH;*/
/*run;*/
/*ods listing;*/
/**/
/**/
/*ODS output Moments=MeanConverge;*/
/*ods listing close;*/
/*proc univariate data=Converge;*/
/*by  timeeffect ability quality NOTSORTED;*/
/*var status pdG pdH;*/
/*run;*/
/*ods listing;*/
/**/
/**/
/*data MeanConverge&tmax&smax&ymax&missing;*/
/*set MeanConverge;*/
/*if Label1 ^= 'Mean' then delete;*/
/*percent_convergence=(1-cvalue1)*100;*/
/*iterations=&iter;*/
/*teachers=&tmax;*/
/*students=&smax;*/
/*years=&ymax;*/
/*missing=&missing;*/
/*keep Varname iterations timeeffect
ability quality teachers students years
```

```
missing percent_convergence ;*/
/*run;*/
quit;
%mend Sim1;

%Sim1 (iter=25,tmax=20,smax=20,ymax=4,cmax=2,
qmax=1, amax=1,yemax=1, cemax=1,missing=2);
%Sim1 (iter=25,tmax=20,smax=20,ymax=8,cmax=2,qmax=1,
amax=1,yemax=1, cemax=1,missing=2);
%Sim1 (iter=25,tmax=35,smax=20,ymax=2,cmax=2,qmax=1,
amax=1,yemax=1, cemax=1,missing=2);
%Sim1 (iter=25,tmax=35,smax=20,ymax=4,cmax=2,qmax=1,
 amax=1,yemax=1, cemax=1,missing=2);
%Sim1 (iter=25,tmax=35,smax=20,ymax=8,cmax=2,qmax=1,
amax=1,yemax=1, cemax=1,missing=2);
%Sim1 (iter=100,tmax=20,smax=20,ymax=2,cmax=2,
qmax=1, amax=1,yemax=1, cemax=1,missing=4);
%Sim1 (iter=100,tmax=20,smax=20,ymax=4,cmax=2,
qmax=1, amax=1,yemax=1, cemax=1,missing=4);
%Sim1 (iter=100,tmax=20,smax=20,ymax=8,cmax=2,
qmax=1, amax=1,yemax=1, cemax=1,missing=4);
%Sim1 (iter=100,tmax=35,smax=20,ymax=2,cmax=2,
qmax=1, amax=1,yemax=1, cemax=1,missing=4);
%Sim1 (iter=100,tmax=35,smax=20,ymax=4,cmax=2,
qmax=1, amax=1,yemax=1, cemax=1,missing=4);
%Sim1 (iter=100,tmax=35,smax=20,ymax=8,
cmax=2,qmax=1, amax=1,yemax=1, cemax=1,missing=4);
```

VITA

Ross Allen Andrew Larsen

Department of Educational Psychology, College of Education & Human Development, Texas A&M University, 704 Harrington Tower, MS 4225, College Station, TX 77843-4225.

Ph.D. in Educational Psychology. Texas A&M, College Station, TX. Research interests: Research and Measurement Second level dependencies in SEM models, missing data problems, professional development of principals and teachers. Chair: Dr. Victor Willson. August 2010.

Master of Science. Brigham Young University, Provo, UT. Research interests: Statistics, Efficacy of maximum likelihood techniques, genetic algorithms. Chair: Dr. Bruce Schaalje. Aug 2006

Bachelor of Science. Brigham Young University, Provo UT. Major: Biostatistics, minor psychology. Deans List. Aug 2004

Larsen, R., & Warne, R. (2010). Estimating confidence intervals for eigenvalues in exploratory factor analysis. *Behavior Research Methods* (in press)

Larsen, R., Schaalje, G., & Lawson, J. (2009). Food shelf life: Estimation and optimal design. *Journal of Statistical Computation and Simulation*, 80(2),1-15.