

MISSING DATA TREATMENTS AT THE SECOND LEVEL OF HIERARCHICAL LINEAR MODELS

Suzanne W. St. Clair, B.S., M.P.H.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

August 2011

APPROVED:

Robin K. Henson, Major Professor

Robert Pavur, Minor Professor

J. Kyle Roberts, Committee Member

Endia Lindo, Committee Member

Abbas Tashakkori, Chair of the Department of
Educational Psychology

Jerry Thomas, Dean of the College of
Education

James D. Meernik, Acting Dean of the
Toulouse Graduate School

St. Clair, Suzanne W., Missing data treatments at the second level of hierarchical linear models.

Doctor of Philosophy (Educational Research), August 2011, 63 pp., 18 tables, references, 58 titles.

The current study evaluated the performance of traditional versus modern MDTs in the estimation of fixed-effects and variance components for data missing at the second level of an hierarchical linear model (HLM) model across 24 different study conditions. Variables manipulated in the analysis included, (a) number of Level-2 variables with missing data, (b) percentage of missing data, and (c) Level-2 sample size. Listwise deletion outperformed all other methods across all study conditions in the estimation of both fixed-effects and variance components. The model-based procedures evaluated, EM and MI, outperformed the other traditional MDTs, mean and group mean substitution, in the estimation of the variance components, outperforming mean substitution in the estimation of the fixed-effects as well. Group mean substitution performed well in the estimation of the fixed-effects, but poorly in the estimation of the variance components. Data in the current study were modeled as missing completely at random (MCAR). Further research is suggested to compare the performance of model-based versus traditional MDTs, specifically listwise deletion, when data are missing at random (MAR), a condition that is more likely to occur in practical research settings.

Copyright 2011

by

Suzanne W. St. Clair

TABLE OF CONTENTS

	Page
LIST OF TABLES.....	iv
MISSING DATA TREATMENTS AT THE SECOND LEVEL OF HIERARCHICAL LINEAR MODELS.....	1
Introduction.....	1
Method.....	14
Results.....	21
Discussion.....	33
Conclusion.....	40
APPENDIX	
R CODE.....	42
REFERENCES.....	59

LIST OF TABLES

	Page
1. HLM Model Parameter Estimates.....	16
2. Factor Levels.....	18
3. Mean Estimates of the MEANSES Fixed-Effects by MDT and Number of Missing Variables.....	22
4. Effect Size Estimates (Cohen's <i>d</i>) of MDTs on the Fixed-Effects for MEANSES.....	23
5. Mean Estimates of MANOVA 3 Intercept by MDT and Number of Missing Variables.....	24
6. Mean Estimates of MANOVA 3 CSES by MDT and Number of Missing Variables.....	24
7. Mean Estimates of MANOVA 3 MEANSES x CSES by MDT and Number of Missing Variables.....	25
8. Mean Estimates of MANOVA 3 SECTOR x CSES by MDT and Number of Missing Variables.....	25
9. Mean Estimates of MANOVA 4 Intercept by MDT and % Missing.....	27
10. Mean Estimates of MANOVA 4 SECTOR by MDT and % Missing.....	27
11. Mean Estimates of MANOVA 4 SECTOR x CSES by MDT and % Missing.....	28
12. Mean Estimates of Intercept Variance by MDT and % Missing.....	29
13. Mean Estimates of Covariance by MDT and % Missing.....	29
14. Mean Estimates of Slope Variance by MDT and Number of Missing Variables.....	30
15. Mean Estimates of Intercept Variance by MDT and Number of Missing Variables.....	30
16. Mean Estimates of Level-1 Variance by MDT and Number of Missing Variables.....	31
17. Mean Estimates of Covariance by MDT and Number of Missing Variables.....	32
18. Mean Estimates of Intercept Variance by MDT and Level-2 Sample Size.....	33

MISSING DATA TREATMENTS AT THE SECOND LEVEL OF HIERARCHICAL LINEAR MODELS

Introduction

A common problem in educational and survey research is the absence of observations for individuals on one or more variables of interest, which is generally referred to as missing data. Missing data can result from unanswered questions, student absences, movement of students out of a school, or any situation involving the failure to collect an observation for a given variable. Generally, missing data can result in two basic problems for most research designs: (a) a decrease in statistical power due to a loss of information, and (b) the possibility of biased estimates for parameters (Roth, 1994). Power refers to the ability of a test to detect a statistically significant effect. Bias refers to the under or overestimation of a parameter (Roth & Switzer, 1995).

Reduction to Statistical Power

A reduction in statistical power translates to an increased Type II error rate, which means that a statistically significant effect, if present, will be more difficult to detect (Delucchi, 1994). Because most statistical procedures require complete data for a case for inclusion in analyses, cases with incomplete data are discarded by default (Gibson & Olejnik, 2003). For multivariate procedures, the likelihood of a cases being discarded increases as the number of variables increases (Gleason & Staelin, 1975). The resulting decreased sample size decreases the power of statistical tests to detect effects (Gibson & Olejnik, 2003).

Monte Carlo studies have demonstrated that 18.3% of cases from a data set may be lost to analyses when 2% of the data are missing randomly and entire cases with missing data are deleted (Roth & Switzer, 1995). Many research design choices are conditioned on how missing data will be treated. In determining sample size to provide adequate statistical power, the researcher will collect data from more

than the required sample to adjust for the expected amount of missing data. Even with best efforts, however, researchers often find significant levels of missing data (Fichman & Cummings, 2003) and may not have large enough sample or enough power to detect a statistically significant effect.

Biased Parameter Estimates

Missing data may also result in biased parameter estimates. If data are missing systematically, parameter estimates will be biased (Davey, Shanahan, & Schafer, 2001). In education for example, if low performers have a greater tendency to be absent, they will more likely be missed for measurement on a given variable. In this situation, measures of central tendency such as the mean will be biased upward because of the absence of lower scores. Model parameters such as regression coefficients will also be biased if the variable with missing data is related to the dependent variable. The bias also affects the amount of dispersion or variance around parameters (Roth & Switzer, 1995) and in this example may result in a reduction of variance for a given variable. Reducing variance effectively restricts the range of variable values and leads to attenuation or reduction of correlations between variables with reduced variance. Eliminating entire cases with missing data may increase the variance around the parameter estimate because of the smaller N divisor in the variance equation (Roth & Switzer, 1995) leading to the calculation of unreasonably large ranges for confidence intervals. Simply stated, bias leads to a degradation of accuracy in the estimation of parameter estimates and precision in the estimation of parameter variance and confidence intervals.

Missing Data Mechanisms

Bias is heavily influenced by the missing data mechanism (MDM) underlying a given dataset. A MDM is the process that prevents us from observing our intended data (Xiao-Li Meng, 2000), or a description of the probability distribution of the pattern of missing observations (Heitjan, 1997). Simply

stated, the MDM provides an answer to the question, “why are the data missing?” (Hedecker & Gibbons, 2006). The concept of MDMs was formalized by Rubin (1976), who described two types of MDMs, data missing completely at random (MCAR), and missing at random (MAR). Collectively, MCAR and MAR assumptions are known as ignorable MDMs. Rubin’s theory has been expanded to include a third MDM, not missing completely at random (NMAR) or nonignorable (Little & Rubin, 2002).

MCAR. If the probability of missing data on a variable, Y , is unrelated to the value of Y itself or to the values of any other variables in the dataset, data are MCAR and are essentially a simple random subsample from the original set of observations (Allison, 2002; Hedecker & Gibbons, 2006; Little & Rubin, 2002). For example, if weight and age are two variables in a dataset, where some subjects are missing values for the weight variable, data are MCAR if (a) data missing for weight are not related to the missing data values for weight, and (b) data missing for weight are not related to the values for age. When MCAR holds, regression using all complete records, means of available cases, nonparametric tests, and moment-based methods such as generalized estimating equations are all valid or unbiased (Heitjan, 1997). However, by employing common MDTs such as listwise and pairwise deletion, the original sample size is decreased resulting in a reduction in power (Little & Rubin, 2002). In other words, MCAR estimates are generalizable, but less precise because of smaller sample size (Pigott, 2001).

MAR. A slightly weaker assumption than MCAR is data missing at random (MAR). MAR holds when the probability of missing data on Y is unrelated to the value of Y after controlling for other variables in the analysis, or the probability of missingness on Y given the values of X and Y is equal to the probability of missingness on Y given the values of X alone (Allison, 2002; Hedecker & Gibbons, 2006). With MAR, missingness depends only on the components Y_{obs} or on the components of Y that are observed and not on the components that are missing. Continuing with the weight and age example above, data would be

MAR if (a) data missing for weight are not related to the missing data values for weight, and (b) data missing for weight are related to the values for age. It is not possible to test whether or not MAR is satisfied because we don't know the values of the missing data so cannot compare the values of those with and without missing data to see if they differ systematically on that variable (Allison, 2002).

MAR is defined relative to the variables in the dataset. If a variable X is related to both the missingness of other variables and to the values of those variables, and X is removed from the dataset, MAR no longer holds. Because of this, it is wise to include variables in the imputation process that are predictive of missingness (Schafer & Olsen, 1998).

Both MCAR and MAR can be classified as ignorable missing data mechanisms. Ignorability implies that it is not necessary to model the process by which the data are missing (Allison, 2002), which simplifies the model-based methods used for the missing data analysis (Pigott, 2001). Technically, a missing data mechanism is referred to as ignorable if (a) data are MAR, and (b) the parameters that govern the missing data are not related to the parameters to be estimated (Hedeker & Gibbons, 2006). Although the second condition may not always be satisfied for MAR, MAR is still classified as ignorable because methods that assume ignorability work well even when the second assumption is violated. When dealing with ignorable missing data, the strategy is to adjust for all observable differences between missing and nonmissing cases and assume that all remaining differences are unsystematic (Allison, 2002). Methods used to treat ignorable missing data include listwise deletion, pairwise deletion, mean imputation, group mean imputation, regression imputation, the expectation maximization (EM) algorithm, and multiple imputation (MI).

NMAR. The ignorability assumption is often unrealistic because uncontrolled missingness typically arises from a mixture of ignorable and nonignorable sources (Schafer & Olsen, 1998). When the

ignorability assumption is violated, data are considered not missing at random (NMAR) or the missing data mechanism is said to be nonignorable. NMAR implies that the missingness of a variable Y is related to the value of Y . If typical MDTs are applied to data that are NMAR, parameter estimates will be biased (Little & Rubin, 2002). However, the alternative is to propose a formal probability model for nonresponse and to carry out the analysis under that model, which requires a great deal of technical expertise (Schafer & Olsen, 1998). The results are very sensitive to the choice of the model, and there is no empirical way to discriminate the results of one nonignorable model from another (Allison, 2002). For these reasons, Schafer & Olsen (1998) recommend the cautious use of ignorable methods with an awareness of their limitations, which is the focus of this paper.

Missing Data Treatments

There are several methods that are commonly used to help alleviate the problems caused by missing data. These methods are collectively referred to as missing data treatments (MDTs). MDTs can be classified as (a) deletion procedures, (b) imputation procedures, and (c) model-based methods. The discussion of MDTs will be limited to the methods applied in the current study, (a) listwise deletion, (b) mean substitution, (c) group mean substitution, (d) the expectation maximization (EM) algorithm, and (e) multiple imputation (MI).

Listwise deletion. Listwise deletion is by far the most commonly used MDT (Davey, Schanahan, & Schafer, 2001). With listwise deletion, individuals or cases are completely dropped from an analysis if they have missing values on any variable. Listwise deletion is easy to implement as it is the default in most statistical packages. Listwise deletion produces a complete data set, so standard statistical analyses can be applied without modifications. The use of a consistent sample base results in the computation of comparable univariate statistics (Little & Rubin, 2002). Despite its advantages, a major drawback to

listwise deletion is that it can dramatically reduce sample size, resulting in a loss of statistical power (Baraldi & Enders, 2010). When the percentage of missing cases is low, loss of precision and bias in the estimation of parameters will generally be low (Little & Rubin, 2002). This statement is highly contingent on the MDM underlying the data as the use of listwise deletion assumes that data are MCAR. When MCAR does not hold, as it generally does not in practical settings, listwise deletion will produce biased results (Baraldi & Enders, 2010).

When data are MCAR, assuming the original sample would produce unbiased parameter estimates, listwise deletion will produce unbiased estimates, as the reduced sample is essentially a random subsample of the original sample (Allison, 2002). Unbiased parameter estimation generally holds only for central tendency estimates such as the mean (Davey, Shanahan, & Schafer, 2001) but due to the reduction of n in the denominator for variance calculations, variance estimates will be biased upward even if MCAR holds. Stated differently, variance estimates will be inflated due to the loss of information (Allison, 2002; Little & Rubin, 2002), resulting in the construction of inefficient, less precise confidence intervals (Davey, Shanahan, & Schafer, 2001). Loss of precision corresponds directly to a loss of power or the ability to detect a statistically significant effect. A caveat to the above occurs if the variable with missing information, Y_1 , is uncorrelated to other variables in the data set. Because the other variables provide no information for the prediction of Y_1 , the variance estimate will be fully efficient (Little & Rubin, 2002).

If data are MAR, listwise can yield biased estimates if the probability of an independent variable being missing is related to the dependent variable being evaluated (Allison, 2002). Mean estimates can be biased upward or downward, depending on the variable values of the individuals deleted from the analysis. Variance and covariance will be attenuated, or biased downward, because of the restricted range of values used in the analysis (Baraldi & Enders, 2010). However, listwise is the most robust method under

violations of MAR among independent variables so long as the probability of missing values on the independent variables is not dependent on the values of the dependent variable (Allison, 2002). A missing data mechanism that depends only on the values of the independent variables is similar to disproportionate stratified sampling of the independent variables, and results in the estimation of unbiased regression coefficients (Allison, 2002). As described above, there are some situations in which listwise deletion may yield unbiased parameter estimates, but the APA Task Force on Statistical Inference warns that listwise deletion is among the worst options available for handling missing values for practical applications (Wilkinson & APA Task Force on Statistical Inference, 1999).

Mean substitution. Imputation procedures, including mean and group mean substitution, effectively reduce the problems of diminished power caused by listwise deletion by filling in missing values so that the complete sample is utilized for each statistical test. Mean substitution simply substitutes the mean of complete cases for a given variable for the missing cases. Imputing the mean produces biased estimates for variances, covariances, and correlation coefficients (Baraldi & Enders, 2010; Haitovsky, 1968). Variance underestimation is a consequence of adding variables to the center of the distribution. When data are MCAR, the sample variance from the imputed data set underestimates the variance by a factor of $(n^{(j)} - 1)/(n - 1)$, where n is total sample size and $n^{(j)}$ is number of available cases (i.e. number of individuals not missing data or having data on a given variable), which can be applied as a correction factor to estimate the true variance. A similar adjustment factor can be applied for computing covariances. Although correction factors will yield unbiased estimates, their use may result in covariance matrices that are not positive-definite (Little & Rubin, 2002). Without adjustment, correlation coefficients will also be attenuated or biased downward (Baraldi & Enders, 2010; Raymond, 1986). Further, confidence intervals created from the attenuated variances and covariances are less likely to cover the true parameter (Little &

Rubin, 1999). Another problem with this method is that it ignores relationships among other variables that may be valuable in predicting the true missing values (von Hippel, 2003). Even under the most stringent assumption of data MCAR, mean imputation will produce biased results, and therefore is not recommended under any circumstances (Pigott, 2001).

Group mean substitution. Group mean substitution goes a step beyond simple mean substitution, computing means by group membership on one or more other variables and imputing those values for missing cases. Also known as conditional mean imputation, group mean substitution classifies non-respondents and respondents into J adjustment classes, based on the observed variables, and imputes the respondent mean for non-respondents in the same class (Little & Rubin, 2002). Group mean substitution should result in variances and covariances that are less biased than that of simple mean substitution because the method takes into account relationships with other variables (von Hippel, 2003). However, group mean substitution ignores random components so variances will still be underestimated (von Hippel, 2003).

Imputation methods are more sophisticated than deletion procedures in that they utilize information from the data set to predict missing data values, thereby preserving sample size and statistical power. However, variances are always underestimated, regardless of the missing data mechanism, leading to attenuation of correlation coefficients with mean imputation and inflated correlations in regression imputation. If data are MAR or NMAR, estimates will be even more severely biased.

Model-based methods. A better approach to missing data imputation is to employ model-based methods such as the expectation maximization (EM) algorithm and multiple imputation (MI), which are based on the notion of a complete data set or the data planned to have been collected originally (Longford et al., 2000; Baraldi & Enders, 2010). Both procedures are iterative processes that employ maximum

likelihood (ML) algorithms where the goal is to choose estimates that if true will maximize the likelihood of observing the data at hand (Allison, 2002). In general, ML methods borrow information from other variables during the estimation of parameters that involve missing values by incorporating information from the conditional distribution of the missing data given the observed data (Enders, 2001). Simulation studies have suggested that ML algorithms may be superior to traditional ad hoc MDTs in many cases (Arbuckle, 1996; Enders & Bandalos, 2001; Muthen et al., 1987, Wothke, 2000). MI generally uses EM estimates as starting points, therefore the EM algorithm will be described first.

EM algorithm. The EM algorithm is an indirect model-based method that actually imputes values for cases with missing data. Other ML methods such as full information maximum likelihood (FIML) and the multiple-group approach directly estimate parameters without imputation (Enders, 2001). Unlike conventional regression imputation, the EM algorithm always uses all available variables for predictors (Allison, 2002).

The EM algorithm consists of two steps, the expectation (E) and maximization (M) steps, which are repeated in an iterative process until the model converges to the ML estimates. For application to the multivariate normal distribution, the E step is essentially a regression imputation of the missing values (Allison, 2002). Using listwise or pairwise deletion of the data set, variables with missing values are regressed on all available variables in the data set to obtain regression equations, and these equations are used to generate imputations for the missing values (Allison, 2002). Stated differently, the E step finds the conditional expectation of the missing data given the observed data and current estimated parameters (i.e. mean and covariance matrix estimated using listwise or pairwise deletion (Allison, 2002)), then substitutes these expectations for the missing data (Little & Rubin, 2002).

The M step can be described as ML estimation of a parameter as if there had been no missing data (Little & Rubin, 2002). It calculates new values for the means and covariance matrix using the imputed data from the E step with the original nonmissing data (Allison, 2002). Means are calculated using standard formulas (Allison, 2002). However, because the imputations are deterministic in nature (Schafer & Olsen, 1998), falling directly on a regression line and lacking a random error component, variance and covariance calculations must be modified so as not to underestimate these parameters (Allison, 2002; Enders, 2001). After the new estimates are calculated, the process starts over with the E step. The E and M steps are cycled through until the estimates converge, or there is very little change from one iteration step to the next (Allison, 2002). Convergence may be slow if a large percentage of the data is missing (Little & Rubin, 2002). Programs that incorporate the EM algorithm include SPSS Missing Values, EMCOV (Graham & Hofer, 1993), and NORM (Schafer, 1998), and Amelia II (Honaker, King & Blackwell, 2010) available for free download in R. Limitations with the EM algorithm include 1) limited theory and software for models other than linear and log-linear models, and 2) the reliance on corrected variance and covariance calculations to account for the underestimation of these parameters due to the deterministic nature of EM imputation (Allison, 2002).

MI. MI helps to overcome the limitations of the EM algorithm as it can be used with virtually any type of model and can be done with conventional software. Further, MI produces different estimates each time because random variation is intentionally introduced into the imputation process by taking random draws from the residual distribution of each imputed variables and adding these random numbers to the imputed values (Allison, 2002). MI further corrects the downward bias of standard error estimates by repeating data imputation multiple times such that there are n complete datasets. Having multiple data sets produces variability for the parameters of interest, which adjust the standard errors

upward (Allison, 2002). Variance of parameters in MI will have two components, within and between-imputation variance. Within-imputation variance is simply the mean of the variances of the imputed data sets, and between imputation variance is calculated as the variance between the variances of the imputed data sets (Little & Rubin, 2002). Adequate results can be obtained using as few as five imputed data sets (Schafer, 1997). MI will produce unbiased parameter estimates and realistic standard errors if model assumptions hold and the data are MAR or MCAR. Even if MAR fails, MI will still result in robust estimates as long as the proportion of missing data is small (Raudenbush & Bryk, 2002).

MDTs Applied within the HLM Framework

The effectiveness of MDTs has been evaluated by many studies for univariate analyses such as ANOVA and regression, but few have evaluated their effectiveness for hierarchical linear models (HLM), which are commonly employed in organizational and educational research. HLM models take into consideration data at multiple levels. In a simple two-level model, observations for individuals such as test scores are Level-1 or first-level data, and observations for group-level data such as a school's graduation rate are Level-2 or second-level data. Missing data at the first level of an HLM model is not a large problem because grand and group parameters are estimated using Maximum Likelihood (ML) and Bayesian methods, respectively. Both of these procedures produce unbiased parameter estimates because they take into consideration group sample size and variance in their estimation (Bryk and Raudenbush, 1992; Hox, 2002).

Although data missing on individuals within groups (i.e. Level-1 data) does not pose a problem for HLM, missing data at the second level results in entire groups being dropped from the analysis, which greatly impacts the estimation of both grand and group parameters (Hox, 2002). Data may be missing at the second level if there is no information for a group on a second-level variable or if there are no

individual observations within a group on a first-level variable. In either case, there is no information for a group on a variable being used in the model so the parameter estimates will be biased, regardless of the estimation procedure chosen.

Many recent studies have focused on MDTs and their application to structural equation modeling (SEM) (Savalei & Yuan, 2009; Song & Lee, 2008; Shin, Davison, & Long, 2009; Yuan & Lu, 2008), differential item functioning (DIF) research (Furlow et al., 2007; Garrett, 2009), and planned missingness in longitudinal designs (Baraldi & Enders, 2010; Graham, Hofer, & MacKinnon, 1996; Graham et al., 2006). Although many of these studies evaluate the effectiveness of MDTs to data that are hierarchical in nature, they do not explicitly evaluate the effectiveness or provide guidance for the application of MDTs at the second level of HLM models. Most recently, Shin and Raudenbush (2010) proposed a latent cluster-mean approach to the contextual effects model with missing data, but their approach focuses on the use of Level-1 covariates for imputation of missing Level-2 variables. The focus of this study is the evaluation of MDTs at the second level of HLM models without consideration of the value of the Level-1 covariates.

Only one study to date (Gibson & Olejnik, 2003) has specifically evaluated the effectiveness of MDTs for data missing at the second level of an HLM analysis. Gibson and Olejnik evaluated the effectiveness of (a) traditional or ad hoc MDTs including listwise deletion, mean substitution, group mean substitution, (b) and modern or model-based MDTs including expected maximization (EM) algorithm, and multiple imputation (MI) over four manipulated factors: (a) sample size for the second-level units, (b) correlation between intercepts and slopes, (c) number of second-level variables, and (d) percentage of missing data. Data were generated to simulate a subset of data from the 1982 High School and Beyond Survey (National Center for Education Statistics, 1982) analyzed by Singer (1998) and discussed in Bryk and Raudenbush (1992). Results were evaluated using multivariate repeated measures analysis to evaluate

the stability of parameter estimates over the various study conditions for each MDT. Findings revealed listwise deletion and the EM algorithm performed well in estimating fixed effects for variables having missing values and variables with no missing values. Listwise deletion was the only MDT that produced unbiased random effects estimates. Unexpectedly, the MI procedure yielded the most distorted estimates of all MDTs evaluated. Overall mean substitution also produced very biased effects. None of the MDTs evaluated produced unbiased random effects estimates when the Level-2 sample size was 30 and 40% of the data were missing.

To further evaluate the effectiveness of MDTs for data missing at the second level of HLM models, additional factors and study conditions should be considered. The current study will expand upon the work performed by Gibson and Olejnik, adding a factor for number of variables with missing data, and adding levels for sample size for the second-level units and percentage of missing data. This study will also take a different approach in the data sampling procedures, drawing random samples directly from the subset of the HSB dataset (sub-HSB) cited above, in lieu of Gibson and Olejnik's method of recreating and sampling from the recreated sub-HSB dataset. Further, in place of standard statistical packages such as SPSS and SAS, the present study will use R, a free software environment for statistical computing and graphics, for data sampling, application of MDTs, and data analysis.

Purpose

Although a great deal of work in recent methodological research has focused on model-based MDTs, a substantial gap still exists between MDT procedures applied in practical research and the model-based methods recommended by that the methodological literature (Bodner, 2006; Peugh & Enders, 2004; Wood, White, & Thompson, 2004; Baraldi & Enders, 2010). Further, there is limited research available on the application of MDTs specifically within the HLM framework. To that end, the primary goal

of this study is to evaluate and compare the performance of traditional versus modern MDTs when data is missing at the second level of HLM models, expanding upon Gibson and Olejnik's findings by including a broader range of study conditions that are likely to occur in practical settings. It is also hoped that this research will provide practical researchers and students an alternative to SPSS, SAS, and other standard statistical packages for applying MDTs to their data.

Method

Dataset Chosen for Analysis

The sub-HSB dataset cited above was chosen for analysis in the current study because of the familiarity and wide range of use of the HSB dataset in the field of educational research (Heck, 2010). More specifically, the HSB dataset has become a classic in the demonstration of HLM (Bryk & Raudenbush, 1992; Singer, 1998), the statistical procedure applied in the current study.

Research has also been utilized in one prior study of missing data treatments (MDTs) for Level-2 variables in HLM (Gibson & Olejnik, 2003) specifically to the sub-HSB dataset. This dataset contains information from 90 public and 70 Catholic high schools with a total sample size of 7,185 students. There are four student or Level-1 variables within the dataset: (a) a measure of mathematics achievement from the student's senior year (MATH), (b) a measure of socioeconomic status (SES), (c) a (0, 1) dummy variable indicating student ethnicity, and (d) a (0, 1) dummy variable indicating student gender (FEMALE). The Level-2 variables are: (a) the mean SES (MEANSES) of students within each school, (b) a (0, 1) dummy variable indicating whether the school is Catholic or public (SECTOR), (c) the number of students enrolled in the school (SIZE), (d) the proportion of students on the academic track (PRACAD), (e) a scale measuring disciplinary climate (DISCLIM), and (6) a (0, 1) dummy variable (HIMNTY) to identify schools with greater than 40% minority enrollment.

MDTs Applied

For comparability to the Gibson and Olejnik study findings, the same 5 MDTs, listwise deletion, mean substitution, group mean substitution, the EM algorithm, and MI were applied to the missing data in the current study. For all MDTs, data was imputed, or in the case of listwise deletion, deleted, at the second level. Listwise, mean substitution, and group mean substitution were performed using standard functions in R. For the model-based imputation procedures, EM and MI, packages developed specifically executing these algorithms were applied in lieu of developing new code. For EM, the Amelia package (Honaker, King, & Blackwell, 2010) which employs an EM algorithm for imputing missing data was applied. For MI, the MICE package (van Buuren & Groothuis-Oudshoorn, 2011) was utilized. Appendix A provides the code developed and executed in R.

For listwise deletion, groups missing either a value on MEANSES or SECTOR were deleted from the analysis. Mean substitution imputed the mean for MEANSES based on the mean value of MEANSES for the groups that were not missing data for MEANSES. Expected values of 0 or 1 were imputed for the dichotomous variable, SECTOR.

For group mean imputation, a simple correlation analysis was first performed to determine which second level variables in the sub-HSB dataset were correlated with MEANSES and SECTOR. PRACAD was the best overall predictor for both variables ($r = .65$ for MEANSES; $r = .67$ for SECTOR). For the group mean imputation, PRACAD was dummy coded, with 0 equating to values of $PRACAD \leq .50$ and 1 equating to values of $PRACAD > .50$ (.50 is the median value for PRACAD). The values for MEANSES and SECTOR were imputed based on the mean MEANSES and SECTOR on their dummy coded PRACAD value.

HLM Model Evaluated

In the present study, the 5 MDTs identified above were applied to a two variable HLM model over varying conditions as described below. MATH was chosen as the dependent variable for each model. The HLM model included one Level-1 variable, CSES, two Level-2 variables, MEANSES and SECTOR, and a cross-level interaction between CSES and SECTOR. CSES (the grand mean centered SES variable) was used in the analysis for ease of result interpretation. Several researchers have employed this model in their analyses (Bryk & Raudenbush, 1992; Singer, 1998), including Gibson and Olejnik's (2003) research investigating the performance of MDTs at the second level of HLM models. This model was appropriate for the current study as it is a partial replication of Gibson and Olejnik's work and one goal is to compare the findings from the current study to their work. The HLM model can be seen as:

$$\begin{aligned} \text{MATH}_{ij} = & \gamma_{00} + \gamma_{01}(\text{MEANSES})_j + \gamma_{02}(\text{SECTOR})_j + \gamma_{10}(\text{CSES}_{ij}) + \gamma_{11}(\text{MEANSES})_j(\text{CSES}_{ij}) + \\ & \gamma_{12}(\text{SECTOR})_j(\text{CSES}_{ij}) + \\ & u_{0j} + u_{1j}(\text{CSES}_{ij}) + r_{ij}. \end{aligned} \tag{1}$$

where MATH is modeled as a function of the grand mean (γ_{00}); the main effects of MEANSES (γ_{01}), SECTOR (γ_{02}), and CSES (γ_{10}); two cross-level interactions of MEANSES with CSES (γ_{11}) and SECTOR with CSES (γ_{12}); and the random error terms, $u_{0j} + u_{1j}(\text{CSES}) + r_{ij}$. Parameter estimates are presented in Table 1.

Table 1
HLM Model Parameter Estimates

Parameter	Model 1	
Fixed Effects	Coefficient	se
Intercept	12.114	0.199
CSES	2.936	0.151
MEANSES	5.343	0.369
SECTOR	1.215	0.306
MEANSES x CSES	1.044	0.291
SECTOR x CSES	-1.642	0.233
Random Effects		
σ^2_e	36.766	0.018
$\sigma^2_{\mu 0}$	2.376	0.000
$\sigma^2_{\mu 1}$	0.000	0.072
$\sigma^2_{\mu 01}$	0.000	
Model Fit		
AIC	46,524.791	
BIC	46,593.580	
-2*LL	46,504.791	

Study Conditions and Variables Evaluated

Each missing data treatment was tested for robustness of parameter estimates across three main study factors. The three factors are: (a) percentage of missing data or % missing, (b) Level-2 sample size, and (c) number of variables with missing data.

Percentage of missing data. Percentage of missing data was set at 3 levels, 10%, 25%, and 40%. Ten and 40% represent low and high levels of missingness, respectively, and are consistent with the values used by Gibson and Olejnik. Twenty-five percent was added to evaluate a moderate level of missingness. Both MEANSES and SECTOR were modeled as having missing values. MEANSES was chosen for comparability to results from Gibson and Olejnik. SECTOR was added to evaluate the effects of missing data across multiple variables, a condition more likely to occur in practice.

Level-2 sample size. Level-2 or group sample size was set at 4 levels, 20, 30, 90, and 160, which represents the total number of groups (i.e. schools) included in the SUB-HSB dataset. Twenty was chosen to evaluate the performance of the ad hoc vs. iterative MDTs under conditions violating the assumptions of the central limit theorem, which states that a sample size of 30 is typically needed to obtain normally distributed parameter estimates (Dielman, 2001). A sample size less of 20 could lead to biased parameter estimates for the ad hoc MDTs, which assume normally distributed data, especially for listwise deletion as a percentage of cases will be deleted from the analysis. Another reason for the use of 20 is simply because small sample sizes such as 20 are commonly encountered in practice. Thirty and 160 were chosen for consistency with Gibson and Olejnik's (2003) Level-2 sample sizes. Ninety was included to further address the effect of Level-2 sample size on power. Past studies employing HLM models have used similar ranges. Snijders and Bosker (1993) suggested that the optimal sample size for number of schools is between 30 and 62, and Level-2 sample size ranges from school effect studies that employed HLM have evaluated

ranges from 21 (Bernstein & Burstein, 1994) to 100 (Hill & Goldstein, 1998). Table 2 presents the different levels for each factor described above. The design is fully crossed, yielding a total of 24 possible study conditions.

Table 2
Factor Levels

Factors	Levels
% Missing	10%, 25% and 40%
Sample Size	20, 30, 90, and 160
Predictors w/Missing Values	1 and 2

Parameter estimates of interest from each model were: (a) fixed-effects estimates for the variable(s) with missing data (MEANSES and SECTOR), (b) fixed-effects estimates for the other predictor variables in the model (including all cross-level interaction effects and SECTOR when not missing), and (c) variance components for each model.

Sampling Procedure

Multiple samples (sample size equal to 20, 30, 90, or 160) for each study condition were drawn without replacement from the sub-HSB dataset by applying an iterative sampling procedure written in R (Appendix A). Before data was deleted from the MEANSES and SECTOR variables for a selected sample, an HLM analysis was first performed on the selected sample (i.e. complete data) to derive parameters for the HLM Model outlined above. The parameter estimates and standard errors for the complete data were later compared to estimates generated with the imputed data for each missing data treatment. Data were simulated as MCAR by applying a simple random sampling procedure, such that all values for both MEANSES and SECTOR had an equal probability of being deleted from the dataset. Missingness was modeled separately for the variables MEANSES and SECTOR, therefore a school may have a missing value for MEANSES, but not for SECTOR. For listwise deletion, this led to a greater proportion of cases deleted

from the analysis than the percent missing study condition of 10% to 40%. One hundred iterations of the following steps were performed for each of the 24 study conditions:

1. Sampled k schools from the sub-HSB dataset (i.e. complete data)
2. Performed HLM analysis and generated parameter estimates for complete data
3. Deleted specified percentage of variable values for predictor(s)
4. Applied MDTs to missing data
5. Performed HLM analysis and generated parameter estimates for MDT treated datasets
6. Saved estimates with fields identifying the study conditions applied

One hundred iterations is consistent with the values used in the Gibson and Olejnik (2003) study and yields a sufficient level of power (>0.80) for detecting a statistically ($p < .05$) and practically (partial η^2 or $\eta_p^2 > .10$) significant effect in the application of a repeated measures design (Cohen, 1992; Cohen, 1988). With the repeated measures design, MDT was modeled as the within subjects factor. In total, 2400 iterations and 14,400 total observations were generated (6 observations for each iteration including 1 for the complete data and 1 for each MDT).

Evaluation of MDT Performance

To determine the accuracy and precision of the estimates for each MDT, the mean estimate and mean standard error for each parameter were computed and compared to estimates produced from the complete data sampled prior to deleting a percentage of variables from MEANSES and SECTOR. To determine if there was a statistically significant effect of the 3 chosen factors on the parameter estimates, 6 repeated measure MANOVA models, one for each set of dependent variables, were used to evaluate which factors contributed to deviations in the estimates. For each MANOVA model, MDT served as the within subjects factor or condition across which each condition was measured, and the remaining study

factors (% missing, Level-2 sample size, and number of predictors with missing data) served as between subjects factors. The variables included for each repeated measures MANOVA model are detailed below.

The independent variables for the MANOVAs were % missing, Level-2 sample size, and number of predictors with missing values. For MANOVA 1, the dependent variable was the fixed-effect for MEANSES (γ_{01}). For MANOVA 2, the dependent variables were the fixed-effects for MEANSES (γ_{01}) and SECTOR (γ_{02}). MANOVA 2 evaluates only conditions with missingness on 2 variables because SECTOR was only missing when number of predictors with missing values was equal to 2. Therefore, the number of factors is reduced to 2 for the MANOVA 2 model (% missing and Level-2 sample size). The dependent variables for MANOVA 3 included fixed effects for intercept (γ_{00}), CSES (γ_{10}), and the cross-level interaction of MEANSES x CSES (γ_{11}), and the fixed effect for SECTOR x CSES (γ_{12}). MANOVA 4 included the MANOVA 3 variables plus SECTOR when number of predictors with missing values was equal to 1 (as on MANOVA 2, number of factors for MANOVA 4 was reduced to 2). MANOVA 5 evaluated the random effects for intercept (u_{0j}), CSES slope (u_{1j}), and the Level-1 residual (r_{ij}), and the covariance (u_{01}).

Statistical significance was evaluated at $\alpha = .05$. Partial eta squared, η_p^2 , was evaluated to determine practical significance ($\eta_p^2 = SS_{effect} / (SS_{effect} + SS_{error})$, where SS_{effect} is the sum of squares for the effect of interest, and SS_{error} is the sum of squares error associated with the effect of interest). Practical significance was set at $\eta_p^2 > .10$ for the MANOVAs because the multiple dependent variables and within subjects measurements (i.e. MDTs) make it relatively easy to detect an effect at the standard $\eta_p^2 > .03$ level, the level chosen by Gibson and Olejnik (2003). For planned contrasts and post-hoc analyses comparing 2 MDTs on one dependent variable, the standard $\eta_p^2 > .03$ as recommended by Kromrey and Hines (1994) was considered practically significant. Planned repeated contrasts for the MDTs compared to the complete data sample on each parameter estimate were first evaluated. For planned contrasts

yielding statistically and practically significant results, post-hoc analyses were then performed. To protect against Type I error rate, Bonferroni correction, maintaining a family-wise $\alpha = .05$, was applied to the planned contrasts and post-hoc analyses.

The practical significance of differences between the complete data values with each MDT was assessed for MEANSES using Cohen's d (Cohen, 1992; Cohen, 1988). For the calculation of Cohen's d , data were evaluated by condition, such that there was one Cohen's d calculated for each MDT on each condition ($24 \times 5 = 120$ Cohen's d values). Following the recommendation of Kromrey and Hines (1994), a Cohen's $d \geq |.3|$ was considered significant. Cohen's d was calculated on MEANSES, the variable imputed across all conditions.

Expected Findings

Data was evaluated for statistical and practical significance using the procedures outlined above.

Expected findings included:

1. Near equivalent performance of all MDTs with 10% missing data for fixed effects.
2. Model-based MDTs outperforming the ad hoc procedures on 40% missing data.
3. Poorer performance for all MDTs going from 1 to 2 variables with missing data.

Biased variance components for ad hoc MDTs across all conditions.

Results

Fixed-Effects for Variables with Missing Data

MANOVA 1 MEANSES. For MANOVA 1, with fixed-effect MEANSES as the dependent variable, an interaction between MDT and number of missing variables was detected, $F(5, 2372) = 73.43, p < .05, \eta_p^2 = .134$. To further assess this interaction effect, data were aggregated across the levels of % missing and Level-2 sample size and analyzed separately for the number of missing variable levels (1 and 2). Table 3

compares the mean estimates of the MEANSES fixed-effect for each MDT to the values produced with complete data by number of missing variables. For both 1 and 2 missing variables, mean substitution produced estimates that were significantly different, both statistically and practically, from the complete data, $F(1, 1199) = 103.99, p < .05, \eta_p^2 = .080$, and $F(1, 1199) = 62.24, p < .05, \eta_p^2 = .049$, for 1 and 2 missing variables, respectively. Mean substitution underestimated MEANSES when number of missing variables was equal to 1, overestimated MEANSES when variable missing was 2. The EM algorithm estimates also differed from the complete data for the case of missing 1 variable, $F(1, 1199) = 40.60, p < .05, \eta_p^2 = .033$, underestimating the MEANSES complete data estimate.

Table 3
Mean Estimates of the MEANSES Fixed-Effects by MDT and Number of Missing Variables

Number of Missing Variables	MDT	<i>M</i>	<i>SE</i>	<i>F</i>	<i>p</i>	η_p^2
0	Complete data	5.285	.019	--	--	--
1	Listwise deletion	5.287	.026	0.02	.89	.000
1	Mean substitution	5.078	.027	103.99	<.01	.080
1	Group mean substitution	5.317	.025	3.36	.07	.003
1	EM algorithm	5.135	.030	40.60	<.01	.033
1	Multiple imputation	5.242	.027	4.82	<.05	.004
0	Complete data	5.283	.020	--	--	--
2	Listwise deletion	5.261	.035	0.63	.43	.001
2	Mean substitution	5.420	.026	62.24	<.01	.049
2	Group mean substitution	5.263	.027	1.40	.24	.001
2	EM algorithm	5.175	.032	16.39	<.01	.013
2	Multiple imputation	5.308	.027	1.83	.18	.002

Note: Data were aggregated over % missing and Level-2 sample size; *df* = 1, 1199.

Effect size estimates for MEANSES. To evaluate the practical significance of differences in the MDT estimates compared to complete data, effect size estimates were computed for the MDTs on fixed-effects estimates for MEANSES using Cohen's *d*. For a given study condition, *d* compares the mean parameter estimate from each MDT to the mean of the complete data samples. Tables 4 displays Cohen's *d* effect sizes of the MDTs on the fixed-effects for MEANSES as the only missing variable.

Table 4

Effect Size Estimates (Cohen's d) of MDTs on the Fixed-Effects for MEANSES

% Missing	Sample Size	Missing Variables	Listwise	Mean Sub	Grp Mean Sub	EM	MI
10%	20	1	0.035	0.033	0.092	0.015	0.071
10%	20	2	0.006	0.013	0.015	0.066	0.054
10%	30	1	0.057	0.070	0.062	0.006	0.026
10%	30	2	0.048	0.034	0.059	0.004	0.005
10%	90	1	0.025	0.206	0.051	0.002	0.049
10%	90	2	0.080	0.176	0.006	0.226	0.092
10%	160	1	0.013	0.082	0.007	0.027	0.019
10%	160	2	0.023	0.048	0.015	0.046	0.024
25%	20	1	0.059	0.346	0.098	0.516	0.053
25%	20	2	0.117	0.149	0.033	0.500	0.039
25%	30	1	0.140	0.125	0.165	0.041	0.009
25%	30	2	0.182	0.207	0.056	0.203	0.052
25%	90	1	0.116	0.583	0.110	0.363	0.327
25%	90	2	0.008	0.346	0.139	0.165	0.043
25%	160	1	0.031	0.198	0.050	0.078	0.058
25%	160	2	0.035	0.155	0.013	0.115	0.050
40%	20	1	0.027	0.578	0.084	1.267	0.489
40%	20	2	0.121	0.285	0.066	0.838	0.130
40%	30	1	0.036	0.349	0.021	0.478	0.367
40%	30	2	0.169	0.198	0.099	0.343	0.087
40%	90	1	0.001	0.933	0.340	0.211	0.036
40%	90	2	0.654	0.663	0.279	0.745	0.406
40%	160	1	0.015	0.251	0.095	0.161	0.121
40%	160	2	0.035	0.227	0.034	0.197	0.094
Mean d			0.085	0.261	0.083	0.276	0.113
Standard deviation			0.132	0.227	0.082	0.316	0.136

Note: Boldface indicates practical significance (Cohen's $d \geq .3$)

For the estimation of the MEANSES fixed-effect, all MDT procedures performed well across all conditions when % missing was 10%, an expected outcome. With the exception of listwise deletion and group mean imputation, all the procedures started to break down at the 25% missing point. Mean imputation, and the EM algorithm were the poorest performers, particularly in the cases where 40% of the data was missing. Listwise deletion and group mean substitution demonstrated the best performance of the MDTs, breaking down only under conditions where 40% of the data was missing. Group mean substitution was the most precise estimator of MEANSES, with both the smallest mean Cohen's $|d|$ and least amount of variance in Cohen's $|d|$ across all study conditions.

Fixed Effects for Variables without Missing Data

MANOVA 3. For MANOVA 3, which included the Level-1 and cross-level interaction effects, a statistically and practically significant effect was detected for the interaction between number of missing variables and MDT, $F(20, 2357) = 80.14, p < .05, \eta_p^2 = .405$. Planned repeated contrasts of this interaction effect revealed differences across all effects estimated. Data were aggregated across the non-statistically and practically significant predictors for MANOVA 3 (% missing and Level-2 sample size) and the mean estimates produced by each MDT for each fixed-effect were compared to the complete data by number of missing variables (Tables 5-8).

Table 5
Mean Estimates of MANOVA 3 Intercept by MDT and Number of Missing Variables

Effect	Number of Missing Variables	MDT	M	SE	F	p	η_p^2
Intercept	0	Complete data	12.106	.009	--	--	--
	1	Listwise deletion	12.118	.012	3.16	.08	.003
	1	Mean substitution	11.900	.013	564.10	<.01	.320
	1	Group mean substitution	12.106	.012	0.00	.98	.000
	1	EM algorithm	12.074	.013	14.69	<.01	.012
	1	Multiple imputation	12.070	.012	21.20	<.01	.017
	0	Complete data	12.124	.009	--	--	--
	2	Listwise deletion	12.137	.015	1.26	.26	.001
	2	Mean substitution	12.103	.014	3.72	.05	.003
	2	Group mean substitution	12.157	.012	14.42	<.01	.012
	2	EM algorithm	12.135	.013	1.17	.28	.001
	2	Multiple imputation	12.136	.012	1.85	.17	.002

Note: Data were aggregated across % missing and Level-2 sample size; $df = 1, 1199$

Table 6
Mean Estimates of MANOVA 3 CSES by MDT and Number of Missing Variables

Effect	Number of Missing Variables	MDT	M	SE	F	p	η_p^2
CSES	0	Complete data	2.940	.008	--	--	--
	1	Listwise deletion	2.941	.010	0.04	.84	.000
	1	Mean substitution	2.905	.008	200.74	<.01	.143
	1	Group mean substitution	2.930	.008	14.90	<.01	.012
	1	EM algorithm	2.927	.008	26.32	<.01	.021
	1	Multiple imputation	2.929	.008	23.35	<.01	.019
	0	Complete data	2.945	.007	--	--	--
	2	Listwise deletion	2.934	.012	1.08	.30	.001
	2	Mean substitution	2.735	.010	838.90	<.01	.412
	2	Group mean substitution	2.918	.008	42.88	<.01	.035
	2	EM algorithm	2.860	.009	255.11	<.01	.175
	2	Multiple imputation	2.878	.008	185.25	<.01	.134

Note: Data were aggregated across % missing and Level-2 sample size; $df = 1, 1199$

Table 7

Mean Estimates of MANOVA 3 MEANSES x CSES by MDT and Number of Missing Variables

Effect	Number of Missing Variables	MDT	M	SE	F	p	η_p^2
MEANSESxCSES	0	Complete data	0.968	.018	--	--	--
	1	Listwise deletion	0.949	.022	2.11	<.01	.000
	1	Mean substitution	0.919	.022	11.77	<.01	.010
	1	Group mean substitution	0.890	.021	41.74	<.01	.034
	1	EM algorithm	0.892	.019	45.17	<.01	.036
	1	Multiple imputation	0.924	.020	15.40	<.01	.013
	0	Complete data	1.000	.018	--	--	--
	2	Listwise deletion	0.925	.030	9.77	<.01	.008
	2	Mean substitution	0.632	.024	449.74	<.01	.273
	2	Group mean substitution	0.955	.022	10.39	<.01	.009
	2	EM algorithm	0.802	.021	200.43	<.01	.143
	2	Multiple imputation	0.873	.021	90.79	<.01	.070

Note: Data were aggregated across % missing and Level-2 sample size; $df = 1, 1199$

Table 8

Mean Estimates of MANOVA 3 SECTOR x CSES by MDT and Number of Missing Variables

Effect	Number of Missing Variables	MDT	M	SE	F	p	η_p^2
SECTORxCSES	0	Complete data	(1.640)	.013	--	--	--
	1	Listwise deletion	(1.639)	.015	0.02	.89	.000
	1	Mean substitution	(1.553)	.013	314.49	<.01	.208
	1	Group mean substitution	(1.613)	.013	30.57	<.01	.025
	1	EM algorithm	(1.613)	.013	33.58	<.01	.027
	1	Multiple imputation	(1.614)	.013	33.46	<.01	.027
	0	Complete data	(1.643)	.012	--	--	--
	2	Listwise deletion	(1.604)	.020	5.23	<.05	.004
	2	Mean substitution	(1.208)	.015	1,140.07	<.01	.487
	2	Group mean substitution	(1.550)	.014	121.10	<.01	.092
	2	EM algorithm	(1.458)	.014	299.45	<.01	.200
	2	Multiple imputation	(1.503)	.014	217.34	<.01	.153

Note: Data were aggregated across % missing and Level-2 sample size; $df = 1, 1199$

With the exception of mean substitution, all the MDTs performed well in the estimation of the intercept parameter across conditions with both 1 and 2 missing variables. As hypothesized, the MDTs typically performed better in the case of 1 missing variable for the Level-1 and cross-level interaction effects. The major exception was for mean substitution, which performed better when 2 variables were missing, $F(1, 1199) = 564.10$, $p < .05$, $\eta_p^2 = .320$, and $F(1, 1199) = 3.72$, $p < .05$, $\eta_p^2 = .003$, for 1 and 2 missing variables, respectively. Overall, mean substitution performed the poorest across the estimation of the other fixed-effects parameters, consistently underestimating the parameters across both 1 and 2 missing variables. The other MDTs were more balanced in the over and underestimation of the parameters. MI and EM yielded different Level-1 and cross-level parameter estimates than the complete

data when number of variables missing was 2. Group mean substitution performed well in the estimation of 3 of the 4 parameters with both 1 and 2 missing variables, but did yield a sizeable difference from the complete data in the estimation of SECTORxCSES. Listwise deletion was the only MDT not to yield a practically significant difference from the complete data on any of the post-hocs evaluated.

MANOVA 4. With the addition of the fixed-effect for SECTOR in MANOVA 4, an interaction between % missing and MDT was detected, $F(50, 2328) = 18.80, p < .05, \eta_p^2 = .288$. Planned repeated contrasts of this interaction revealed differences across the effects for Intercept, SECTOR, and SECTORxCSES. Post-hoc analyses were conducted across these fixed-effects by % missing to further investigate the differences.

Results of the post-hoc analyses are shown in Tables 9-11. For every level of % missing on every parameter evaluated, mean substitution yielded a significantly, both statistically and practically, different estimate from that of the complete data, with a larger variance detected from the complete data as % missing increased. Variance from the complete data with mean substitution was most noteworthy on SECTOR. Mean substitution underestimated both main effects components (SECTOR and CSES) and underestimated the SECTORxCSES cross-level interaction across all levels. The other MDTs were more balanced in their effects and performed well under 10% and 25% missingness, although MI detected a small variance from the complete data under 25% missingness. Collectively, MI and the EM algorithm deviated from the complete data under 40% missingness on all 3 parameters estimated. Group mean substitution performed well across all conditions except 40% missing on the SECTORxCSES cross-level interaction, where it showed a slight variance from the complete data. Listwise outperformed all MDTs on all conditions evaluated, and did not detect a statistically or practically significant difference from the complete data on any of the parameter estimates.

Table 9

Mean Estimates of MANOVA 4 Intercept by MDT and % Missing

Effect	% Missing	MDT	M	SE	F	p	η_p^2
Intercept	0	Complete data	12.110	.016	--	--	--
	10	Listwise deletion	12.123	.017	3.71	.05	.009
	10	Mean substitution	12.024	.018	127.22	<.01	.242
	10	Group mean substitution	12.113	.017	0.17	.68	.000
	10	EM algorithm	12.102	.017	1.19	.28	.003
	10	Multiple imputation	12.097	.017	4.09	<.05	.010
	0	Complete data	12.115	.015	--	--	--
	25	Listwise deletion	12.131	.019	2.02	.16	.005
	25	Mean substitution	11.932	.022	162.35	<.01	.289
	25	Group mean substitution	12.134	.021	2.00	.16	.005
	25	EM algorithm	12.088	.022	4.13	<.05	.010
	25	Multiple imputation	12.102	.022	0.89	.35	.002
	0	Complete data	12.092	.017	--	--	--
	40	Listwise deletion	12.100	.024	0.24	.62	.001
	40	Mean substitution	11.745	.026	373.77	<.01	.484
	40	Group mean substitution	12.070	.023	1.95	.16	.005
	40	EM algorithm	12.033	.026	9.43	<.01	.023
	40	Multiple imputation	12.011	.025	21.09	<.01	.050

Note: $df = 1, 399$; Number of variables with missing data = 1 for every condition evaluated.

Table 10

Mean Estimates of MANOVA 4 SECTOR by MDT and % Missing

Effect	% Missing	MDT	M	SE	F	p	η_p^2
SECTOR	0	Complete data	(1.641)	.023	--	--	--
	10	Listwise deletion	(1.654)	.025	2.07	.15	.005
	10	Mean substitution	(1.600)	.023	166.07	<.01	.294
	10	Group mean substitution	(1.628)	.023	1.59	.21	.004
	10	EM algorithm	(1.632)	.023	0.06	.81	.000
	10	Multiple imputation	(1.628)	.023	1.18	.28	.003
	0	Complete data	(1.648)	.021	--	--	--
	25	Listwise deletion	(1.631)	.025	0.07	.79	.000
	25	Mean substitution	(1.560)	.021	361.11	<.01	.475
	25	Group mean substitution	(1.617)	.022	1.80	.18	.004
	25	EM algorithm	(1.621)	.021	0.81	.37	.002
	25	Multiple imputation	(1.615)	.021	4.01	.05	.010
	0	Complete data	(1.632)	.023	--	--	--
	40	Listwise deletion	(1.632)	.030	2.59	.11	.006
	40	Mean substitution	(1.500)	.022	735.04	<.01	.648
	40	Group mean substitution	(1.594)	.024	1.48	.22	.004
	40	EM algorithm	(1.587)	.022	23.84	<.01	.056
	40	Multiple imputation	(1.597)	.022	38.51	<.01	.088

Note: $df = 1, 399$; Number of variables with missing data = 1 for every condition evaluated.

Table 11

Mean Estimates of MANOVA 4 SECTOR x CSES by MDT and % Missing

Effect	% Missing	MDT	M	SE	F	p	η^2_p
SECTORxCSES	0	Complete data	1.287	.029	--	--	--
	10	Listwise deletion	1.271	.029	1.99	.16	.005
	10	Mean substitution	1.463	.031	54.31	<.01	.120
	10	Group mean substitution	1.272	.031	5.41	<.05	.013
	10	EM algorithm	1.284	.030	3.49	.06	.009
	10	Multiple imputation	1.301	.030	6.89	<.05	.017
	0	Complete data	1.205	.028	--	--	--
	25	Listwise deletion	1.199	.036	1.22	.27	.003
	25	Mean substitution	1.662	.037	101.86	<.01	.203
	25	Group mean substitution	1.175	.036	10.64	<.01	.026
	25	EM algorithm	1.226	.037	10.09	<.01	.025
	25	Multiple imputation	1.248	.037	12.94	<.01	.031
	0	Complete data	1.260	.029	--	--	--
	40	Listwise deletion	1.210	.042	0.00	1.00	.000
	40	Mean substitution	2.018	.041	175.37	<.01	.305
	40	Group mean substitution	1.291	.040	14.91	<.01	.036
	40	EM algorithm	1.430	.046	20.60	<.01	.049
	40	Multiple imputation	1.451	.044	14.22	<.01	.034

Note: $df = 1, 399$; Number of variables with missing data = 1 for every condition evaluated.

Variance Components

For the random effects in MANOVA 5, interaction effects were detected between all the between subjects factors and MDT. Planned contrasts were inspected to see which variables were driving these effects. All parameters varied across the number of missing variables and MDT interaction. Intercept variance and covariance differed across the % missing x MDT interaction. Only intercept variance estimation was effected by the Level-2 sample size x MDT interaction effect. Post-hoc analyses were performed to determine the exact pair-wise effects for each interaction and parameter. Results are detailed in Tables 12-15 below.

Tables 12 and 13 display the intercept variance and covariance estimates by MDT and % missing. With the exception of listwise deletion, intercept variance and covariance estimates were inflated for all MDTs, increasing with the level of % missing. Mean and group mean substitution were the poorest performers, particularly in the estimation of intercept variance with 40% missing data. MI and EM

outperformed mean and group mean substitution, but again, listwise deletion produced the least biased estimates across all levels of % missing.

Table 12
Mean Estimates of Intercept Variance by MDT and % Missing

Effect	% Missing	MDT	M	SE	F	p	η_p^2
Intercept Variance	0	Complete data	2.443	.029	--	--	--
	10	Listwise deletion	2.440	.031	0.06	.81	.000
	10	Mean substitution	3.013	.037	672.49	<.01	.457
	10	Group mean substitution	2.802	.034	356.73	<.01	.309
	10	EM algorithm	2.622	.032	99.93	<.01	.111
	10	Multiple imputation	2.644	.032	140.33	<.01	.149
	0	Complete data	2.424	.028	--	--	--
	25	Listwise deletion	2.432	.043	0.07	.79	.000
	25	Mean substitution	3.815	.044	1,777.77	<.01	.690
	25	Group mean substitution	3.292	.040	1,006.72	<.01	.558
	25	EM algorithm	2.997	.043	319.56	<.01	.286
	25	Multiple imputation	2.968	.040	406.74	<.01	.337
0	Complete data	2.378	.030	--	--	--	
40	Listwise deletion	2.384	.047	0.02	.87	.000	
40	Mean substitution	4.552	.050	3,084.68	<.01	.794	
40	Group mean substitution	3.772	.044	1,917.53	<.01	.706	
40	EM algorithm	3.326	.050	512.39	<.01	.391	
40	Multiple imputation	3.344	.045	751.41	<.01	.485	

Note: Data aggregated across number of missing variables and Level-2 sample size; *df* = 1, 799

Table 13
Mean Estimates of Covariance by MDT and % Missing

Effect	% Missing	MDT	M	SE	F	p	η_p^2
Covariance	0	Complete data	0.137	.013	--	--	--
	10	Listwise deletion	0.163	.014	14.05	<.01	.017
	10	Mean substitution	0.227	.015	99.63	<.01	.111
	10	Group mean substitution	0.247	.015	169.33	<.01	.175
	10	EM algorithm	0.204	.014	82.89	<.01	.094
	10	Multiple imputation	0.208	.014	96.84	<.01	.108
	0	Complete data	0.128	.011	--	--	--
	25	Listwise deletion	0.169	.016	11.41	<.01	.014
	25	Mean substitution	0.279	.017	115.96	<.01	.127
	25	Group mean substitution	0.347	.016	319.23	<.01	.285
	25	EM algorithm	0.249	.015	111.11	<.01	.122
	25	Multiple imputation	0.257	.014	143.87	<.01	.153
	0	Complete data	0.130	.011	--	--	--
	40	Listwise deletion	0.192	.019	12.34	<.01	.015
	40	Mean substitution	0.317	.020	121.12	<.01	.132
	40	Group mean substitution	0.462	.017	620.40	<.01	.437
	40	EM algorithm	0.330	.018	205.33	<.01	.204
	40	Multiple imputation	0.306	.016	183.64	<.01	.187

Note: Data aggregated across number of missing variables and Level-2 sample size; *df* = 1, 799

Table 14

Mean Estimates of Slope Variance by MDT and Number of Missing Variables

Effect	Number of Missing Variables	MDT	M	SE	F	p	η_p^2
Slope Variance	0	Complete data	0.200	.009	--	--	--
	1	Listwise deletion	0.283	.012	125.46	<.01	.095
	1	Mean substitution	0.293	.009	414.28	<.01	.257
	1	Group mean substitution	0.293	.009	479.82	<.01	.286
	1	EM algorithm	0.276	.009	307.10	<.01	.204
	1	Multiple imputation	0.267	.009	279.88	<.01	.189
	0	Complete data	0.218	.010	--	--	--
	2	Listwise deletion	0.345	.015	113.32	<.01	.086
	2	Mean substitution	0.476	.012	994.13	<.01	.453
	2	Group mean substitution	0.351	.011	523.59	<.01	.304
	2	EM algorithm	0.351	.011	473.27	<.01	.283
	2	Multiple imputation	0.340	.011	479.24	<.01	.286

Note: Data aggregated across % missing and Level-2 sample size; $df = 1, 1199$

Table 15

Mean Estimates of Intercept Variance by MDT and Number of Missing Variables

Effect	Number of Missing Variables	MDT	M	SE	F	p	η_p^2
Intercept Variance	0	Complete data	2.435	.024	--	--	--
	1	Listwise deletion	2.446	.030	0.34	.56	.000
	1	Mean substitution	3.654	.038	1,728.66	<.01	.590
	1	Group mean substitution	3.285	.034	1,168.41	<.01	.494
	1	EM algorithm	2.952	.034	412.23	<.01	.256
	1	Multiple imputation	2.990	.033	516.96	<.01	.301
	0	Complete data	2.395	.024	--	--	--
	2	Listwise deletion	2.391	.036	0.02	.89	.000
	2	Mean substitution	3.932	.042	1,982.95	<.01	.623
	2	Group mean substitution	3.292	.034	1,294.66	<.01	.519
	2	EM algorithm	3.011	.037	426.70	<.01	.262
	2	Multiple imputation	2.980	.033	601.97	<.01	.334

Note: Data aggregated across % missing and Level-2 sample size; $df = 1, 1199$

Tables 14 and 15 show the pair-wise comparisons by number of missing variables MDT for the slope and intercept variance components. With the exception of listwise deletion, all MDTs performed poorly in the estimation of intercept variance. As the number of missing variables increased from 1 to 2, estimation of intercept variance was consistently, but not markedly, poorer for each MDT (with the exception of listwise). All of the MDTs differed from the complete data in the estimation of slope variance, again with variance from the complete data increasing from 1 to 2 missing variables with the exception of listwise deletion, which was fairly consistent across both levels of missingness. In most cases, mean imputation was the poorest performer and listwise the best.

Table 16

Mean Estimates of Level-1 Variance by MDT and Number of Missing Variables

Effect	Number of Missing Variables	MDT	<i>M</i>	<i>SE</i>	<i>F</i>	<i>p</i>	η^2_p
Variance Level-1	0	Complete data	36.628	.036	--	--	--
	1	Listwise deletion	36.584	.045	2.87	.09	.002
	1	Mean substitution	36.604	.036	339.57	<.01	.221
	1	Group mean substitution	36.605	.036	335.42	<.01	.219
	1	EM algorithm	36.609	.036	233.91	<.01	.163
	1	Multiple imputation	36.610	.036	211.49	<.01	.150
	0	Complete data	36.665	.037	--	--	--
	2	Listwise deletion	36.638	.053	0.48	.49	.000
	2	Mean substitution	36.666	.037	0.49	.48	.000
	2	Group mean substitution	36.652	.037	59.53	<.01	.047
	2	EM algorithm	36.657	.037	21.68	<.01	.018
	2	Multiple imputation	36.655	.037	34.68	<.01	.028

Note: Data aggregated across % missing and Level-2 sample size; *df* = 1, 1199

Tables 16 and 17 display the post-hoc results by number of missing variables and MDT for the Level-1 variance and covariance parameters. For both of these parameters, the performance of the MDTs improved when going from 1 to 2 missing variable values. With the exception of listwise deletion, all the MDTs performed poorly with 1 missing variable. Mean substitution and listwise deletion were both strong performers in the case of 2 missing variables for both Level-1 variance and covariance. Group mean substitution was the poorest overall performer, deviating from the complete data estimate on each missing variable level for both Level-1 variance and covariance. Although listwise appears to have outperformed the other MDTs in the estimation of Level-1 variance, the standard error for this parameter estimate was inflated for listwise compared to the other MDTs. A result of this is the construction of an inefficient, less precise confidence interval (Davey, Shanahan, & Schafer, 1999) that corresponds directly to the ability to detect a statistically significant effect between the complete data and listwise deletion. It should also be noted that the Level-1 variance is obviously more subjective to individual Level-1 data, and because a random sample of the Level-2 predictors was deleted, the Level-1 variance did not vary in a practical sense for any of the MDTs.

Table 17

Mean Estimates of Covariance by MDT and Number of Missing Variables

Effect	Number of Missing Variables	MDT	<i>M</i>	<i>SE</i>	<i>F</i>	<i>p</i>	η_p^2
Covariance	0	Complete data	0.126	.010	--	--	--
	1	Listwise deletion	0.175	.013	32.52	<.01	.026
	1	Mean substitution	0.389	.014	672.54	<.01	.359
	1	Group mean substitution	0.369	.013	664.22	<.01	.356
	1	EM algorithm	0.291	.013	323.54	<.01	.213
	1	Multiple imputation	0.283	.012	344.32	<.01	.223
	0	Complete data	0.138	.010	--	--	--
	2	Listwise deletion	0.175	.014	9.00	<.01	.007
	2	Mean substitution	0.159	.014	3.68	.06	.003
	2	Group mean substitution	0.336	.013	385.65	<.01	.243
	2	EM algorithm	0.232	.013	99.97	<.01	.077
	2	Multiple imputation	0.231	.012	110.61	<.01	.084

Note: Data aggregated across % missing and Level-2 sample size; *df* = 1, 1199

Table 18 shows the interaction effect of MDT and Level-2 sample size in the estimation of intercept variance. With the exception of listwise deletion, all MDTs consistently overestimated intercept variance compared to the complete data. Mean and group mean substitution were the poorest performers, followed by EM and MI. The η_p^2 effect size indicated poorer performance with increased sample size. However, it should be noted that as sample size increases, the variance in the complete data decreases, resulting in a reduction to error variance, part of the denominator in the calculation of the η_p^2 effect size. The variance in the MDT parameter estimates also decreases as sample size increases, but not at the same rate as the complete data, therefore, the variance of the within subjects factor, MDT, becomes progressively larger in relationship to the error variance, resulting in a larger η_p^2 effect size. Stated another way, as sample size increases, there is greater precision in the estimation of a given parameter, making it easier to detect a difference between the complete data estimate and each MDT estimate of that parameter. Caution should be taken when interpreting the results of the increasing η_p^2 effect size without evaluating the mean differences, which in most cases go down as sample size increases.

Table 18

Mean Estimates of Intercept Variance by MDT and Level-2 Sample Size

Level-2 Sample Size	MDT	<i>M</i>	<i>SE</i>	<i>F</i>	<i>p</i>	η_p^2
20	Complete data	2.458	.051	--	--	--
20	Listwise deletion	2.467	.070	0.04	.84	.000
20	Mean substitution	3.863	.077	579.16	<.01	.492
20	Group mean substitution	3.364	.069	359.81	<.01	.375
20	EM algorithm	3.371	.075	224.51	<.01	.273
20	Multiple imputation	3.210	.068	223.88	<.01	.272
30	Complete data	2.433	.041	--	--	--
30	Listwise deletion	2.448	.058	0.15	.70	.000
30	Mean substitution	3.767	.065	688.69	<.01	.535
30	Group mean substitution	3.287	.057	458.49	<.01	.434
30	EM algorithm	3.018	.057	210.69	<.01	.260
30	Multiple imputation	3.064	.056	278.18	<.01	.317
90	Complete data	2.394	.015	--	--	--
90	Listwise deletion	2.378	.024	0.71	.40	.001
90	Mean substitution	3.776	.041	1,417.25	<.01	.703
90	Group mean substitution	3.255	.030	1,166.45	<.01	.661
90	EM algorithm	2.784	.024	451.93	<.01	.430
90	Multiple imputation	2.868	.025	571.43	<.01	.488
160	Complete data	2.376	.000	--	--	--
160	Listwise deletion	2.380	.014	0.09	.76	.000
160	Mean substitution	3.766	.032	1,875.54	<.01	.758
160	Group mean substitution	3.248	.021	1,678.86	<.01	.737
160	EM algorithm	2.753	.013	798.38	<.01	.571
160	Multiple imputation	2.799	.014	880.22	<.01	.595

Note: Data aggregated over number of missing variables and % missing; *df* = 1, 599

Discussion

All of the MDTs performed well in the estimation of MEANSES with 10% missing, as demonstrated by MANOVA 1 (no practical MDT x % missing interaction effect, $F(20, 2358) = 7.40, p < .05, \eta_p^2 = .059$) and with Cohen's $d \leq |.3|$ for all conditions with 10% missing. With the exception of mean substitution, all the MDTs performed well in the estimation of the other fixed-effects with 10% missing data as well. This result was expected given the fact that the estimates generated by each MDT were compared to those generated by the complete data sampled, not the complete data set. Of course in theory, the smaller the sample selected, the less likely statistics generated from that sample will be normally distributed. In this case, any level of missingness could further skew the distribution, especially if missing data is being imputed by a value that does not represent the population from which it was sampled, which is precisely

what is driving the poor performance of mean imputation in this study. The other imputation strategies outperformed mean substitution simply because they used more information from the dataset to impute a value. It is likely that listwise deletion outperformed mean imputation because the data were modeled as MCAR. It has been shown in simpler statistical models (e.g. regression) when data are MCAR, listwise deletion will produce unbiased estimates assuming the original sample would produce unbiased parameter estimates, as the reduced sample is essentially a random subsample of the original sample (Allison, 2002). Findings from the current study suggest that listwise deletion will also produce unbiased estimates when data are MCAR at the second level of HLM.

As the level of % missing increased, mean substitution demonstrated increasingly poor performance in estimating the fixed-effects for variables without missing values (Intercept, CSES, SECTORxCSES), consistently underestimating the main effects and overestimating the absolute value of each interaction effect (SECTORxCSES slope was negative). Mean substitution was by far the poorest performing MDT on all levels of % missing for the case of number of missing variables = 1. The other MDTs performed well in the estimation of the other fixed-effects up to the 40% level of missingness, where group mean substitution, MI, and the EM algorithm all deviated from the complete data on one or more of the parameter estimates. Listwise deletion was the only MDT that demonstrated solid performance across all levels of % missing in the estimation of the fixed-effects, contrary to the hypothesis that EM and MI would outperform all ad hoc MDTs. Listwise deletion outperformed the other MDTs in the estimation of the fixed effects because data were modeled as MCAR. When data are MCAR unbiased parameter estimation generally holds only for central tendency estimates such as the mean (Davey, Shanahan, & Schafer, 2001). Again, when data are MCAR, listwise deletion will result in unbiased fixed-effects estimates because it essentially produces a random subsample of the original sample (Allison, 2002).

Mean substitution, EM, and MI all showed progressively poorer performance for most of the fixed effects when number of predictors with missing values increases from 1 to 2, underestimating the fixed-effects for CSES, and the two cross-level interactions. Listwise and group mean substitution were fairly consistent regardless of the number of missing variables in the estimation of all the fixed-effects. Mean substitution demonstrated inconsistent performance in underestimating the fixed-effect for MEANSES with 1 variable missing and overestimating MEANSES when 2 variables were missing. Because mean imputation ignores relationships among other variables that may be valuable in predicting the true missing values (von Hippel, 2003), its estimates are likely to be biased in either direction, depending on the sample. Even under the most stringent assumption of data MCAR, mean imputation will produce biased results, and therefore is not recommended under any circumstances (Pigott, 2001).

A somewhat surprising result was mean substitution's underestimation of intercept with 1 missing variable, but showing very little variance from the complete data when 2 variables were missing. The expected result would be for greater variance from the complete data with 2 variables missing. Had a stratified sampling procedure for missing data on SECTOR been incorporated, this anomaly may not have occurred. There are a greater percentage of public vs. Catholic schools included in the SUB-HSB dataset, so the imputation of SECTOR based on the PRACAD value may have served to balance out the estimation of the intercept parameter.

Variance components were biased upward for all MDTs across one or more conditions. Intercept and slope variance were relatively the same across number of missing variables for each MDT, but covariance and Level-1 variance both decreased slightly going from 1 to 2 missing variables. This is probably due to the fact that the second variable modeled with missingness, SECTOR, is a dichotomous variable, so values imputed for this variable could actually contribute to lowering the variance in the

model, unlike the deletion of MEANSES, which depending on a given sample could severely skew the variance estimates. The simple random sampling procedure employed on SECTOR in lieu of a stratified random sample may also have contributed to this result.

As the % missing increased, so did the upward bias in the estimate of intercept variance and covariance across all MDTs except listwise. Most notably, when going from 25% to 40% missing on intercept variance, mean substitution and group mean both yielded $\eta_p^2 > .7$. Overall, mean substitution and group mean substitution performed the worst in estimating the variance components across all conditions. This is a logical result because both of these procedures reduce the variance in a sample by imputing the same value or value based on grouping for any case missing a variable. Reducing sampling variance drastically alters the covariances of imputed variables to other variables in a prediction model. Simply stated, imputing the mean produces biased estimates for variances and covariances (Haitovsky, 1968). As covariance is the driving factor in any prediction model, it is evident as to why both mean and group mean imputation performed so poorly in the estimation of the variance components for the chosen HLM model. Applying group mean imputation may be a viable option for estimating fixed-effects as demonstrated by its strong performance across all study conditions in the prediction of MEANSES and the other fixed effects, but should not be applied toward estimating variance components for the reasons stated above. Given mean substitution's poor performance across the majority of the conditions evaluated, and its known problems as identified in previous research, mean imputation should not be used at all.

In summary, of the MDTs evaluated in the current study, listwise deletion performed the best across all conditions, yielding only one practically significant mean difference as evidenced by Cohen's $d = (0.654)$ on the study condition of % missing 40%, sample size 90, and 2 missing variables, and only one

biased estimate (upward) on slope variance as detected by post-hoc pair-wise analyses. The upward biased slope variance estimate was expected because listwise reduces sample size. Due to the reduction of n in the denominator for variance calculations, variance estimates will be biased upward. Stated differently, variance estimates will be inflated due to the loss of information (Allison, 2002; Little & Rubin, 2002). Further, as evident from the larger standard errors computed for the variance components, the use of listwise deletion results in the construction of inefficient, less precise confidence intervals (Davey, Shanahan, & Schafer, 2001). Loss of precision corresponds directly to a loss of power or the ability to detect a statistically significant effect. Therefore, use of listwise deletion is cautioned for use in the HLM framework under conditions where sample size is low and there are a large percentage of missing values for the Level-2 predictors.

Group mean imputation was a solid performer in the estimation of fixed effects, but not for the variance components. Both model-based MDTs outperformed the mean and group mean substitution strategies for the variance component estimates, and in all cases of the fixed-effects estimates, they outperformed mean imputation. Contrary to findings from Gibson and Olejnik (2003), MI consistently outperformed the EM algorithm. This finding is due in part because (a) Gibson and Olejnik utilized different packages for employing MI and the EM algorithm that the current study which use different algorithms in their execution, and (b) the settings utilized by Gibson and Olejnik in the execution of the model-based MDTs may have differed widely from the values applied in the current study. Gibson and Olejnik used NORM (Schafer, 1994) and a modified version of EM_COVAR.SAS (Gregorich, 1997), respectively, for employing MI and the EM algorithm. The current study utilized the R freeware packages MICE (van Buuren & Groothuis-Oudshoorn, 2011) and Amelia (Honaker, King, & Blackwell, 2010) for MI and the EM algorithm, respectively. It is unknown what settings were used in the NORM and

EM_COVAR.SAS packages for execution of the model-based MDTs; the current study employed the default settings within MICE and AMELIA (e.g. no prior assumptions regarding the missing data). Mean substitution was overall the weakest MDT strategy evaluated, yielding parameter estimates that deviated by far the most from the complete data than any other MDT.

Limitations

Data missing completely at random (MCAR) was the only missing data mechanism included in the current study. As previously stated, the strong performance of listwise is expected as data were modeled as MCAR, which is a major limitation of the current study. If data are missing at random (MAR), listwise can yield biased estimates if the probability of an independent variable being missing is related to the dependent variable being evaluated (Allison, 2002). Gibson and Olejnik's (2003) study resulted in similar findings, with poorer performance of the iterative procedures, specifically MI, compared to listwise. They suggested that the performance of MI would have been improved if (a) the number of imputations had been increased and (b) if the data were modeled as MAR as maximum likelihood approaches such as MI and EM, the less stringent requirement of MAR is sufficient (Rubin, 1976). Because data are more likely to be MAR rather than MCAR, it is wise to test the MDTs for robustness under conditions where data are MAR.

The focus of the current study was to evaluate the effectiveness of the MDTs over a broader range of study conditions (% missing, sample size, number of variables with missing data) than shown in prior research. Twenty-four fully crossed study conditions were included in the study design. So as not to add another layer of complexity to the current study, it was decided to save the inclusion of data modeled as MAR for future research. Future studies could incorporate the following method to evaluate the MAR condition for comparison to the results found with MCAR from the current study. For the MAR condition,

model missingness on MEANSES as a function of the variable HIMNTY such that schools with a high minority enrollment are more likely to have missing values for the variable MEANSES. For SECTOR, simulate MAR simulated as a function of the DISCLIM variable such that schools with a lesser disciplinary climate are more likely to have missing values for the variable SECTOR.

Another note of caution in interpreting the strong performance of listwise deletion in the current study is the sampling and data analysis procedures incorporated. With 2 variables being modeled with missing data on 20 groups, listwise deletion will inevitably fail when complex models are implemented because there may not be enough cases or degrees of freedom compared to the number of parameters in the model to carry out the analysis. When this occurs, the lme function in R used to run the HLM analysis within the code for the current study renders a warning that not enough cases are present in the dataset to compute coefficients and the iterative procedure crashes. Even with error-handling code to proceed to the next MDT, the elimination of listwise observations would have been problematic because the repeated measures MANOVA procedure employed for comparison of the MDT performance requires a balanced design, or equal cell sizes for each study condition evaluated. As a work around to these problems, a repeat loop was incorporated in the code such that a new sample was drawn if there were not enough degrees of freedom to perform the HLM. Of course this is not practical in a real-world setting because there is typically only one sample to work with. In the case of too many missing data points, listwise is simply not a viable option if the analysis cannot be run. Refinements to the current study would be to simply count the number of times listwise failed because of insufficient sample size to better portray real-world performance of this MDT. Going a step further, for iterations where listwise or any other MDT failed, rendering unequal cell sizes for the study conditions, the data could be evaluated using a HLM

model. HLM does not require balanced data, so it is not a problem if the number of available measurements (i.e. MDT treatments) is not the same for each iteration (Hox, 2002).

Another enhancement to the current study would be to employ a more sophisticated sampling procedure. Because the current study evaluated a complete data set in which the parameters were already known, simple random sampling was the more powerful approach. Bootstrapping one sample, then sampling with replacement from that sample could become problematic if the original sample drawn was not representative of the population. If nothing was known about the population distribution from which the sample was drawn, bootstrapping would be a more viable option. An idea for future research would be to use the bootstrap to evaluate the effectiveness of MDTs in the estimation of parameters from non-normal distributions.

Conclusion

Although listwise deletion was found to be the best MDT strategy in the current study, listwise is still not considered to be the most viable MDT option because the MCAR condition modeled here is not likely to occur in practice. Future studies should incorporate the MAR condition, where missing data on a given variable is modeled as a function of values of other variables in the dataset. For the sub-HSB dataset used in the current study, MAR could be modeled on both MEANSES and SECTOR, respectively, by having: (a) schools with a high minority enrollment (HIMINTY) more likely to be missing values for the variable MEANSES, and (b) schools with a lesser disciplinary climate (DISCLIM) more likely to be missing values for the variable SECTOR.

Further, listwise deletion will inevitably prevent an HLM analysis from running in situations where there are simply not enough groups to carry out the analysis. The current study created a workaround for this problem, but future studies enhancing the R code used here should at a minimum count how many

times listwise would have failed to carry out the analysis, and preferably incorporate the missing data for a given condition in the estimation of the HLM parameters. In practice, if there is a high probability of listwise failing, it should not be used even if data are MCAR.

Recent methodological research has recommended the application of model-based MDTs to treat missing data; however, a substantial gap still exists between the recommended methods and the actual MDTs applied in practice. To help narrow the gap between methodology and practice, model-based MDTs must be made more user-friendly and accessible to students and research practitioners for use across multiple areas of research and statistical models.

APPENDIX

R CODE

```

#### Missing Data Treatments at HLM 2nd Level
#### SUB-HSB Dataset
#### Simple Random Sampling

#### Load packages for running code
library(foreign)
library(lme4)
library(nlme)
library(car)
library(Amelia)
library(mice)

#### Read in data from SPSS
hsb.data <- read.spss("f:/hsb.data.sav", use.value.labels=TRUE,
max.value.labels=Inf, to.data.frame=TRUE)
hsb.data$SCHOOL <- factor(hsb.data$SCHOOL)

#### Make a dataframe with two columns
  ## Column 1 is the school id
  ## Column 2 is the number of people in the school
hsb.data2<-data.frame(table(hsb.data$SCHOOL))

#### Define levels for each variable
nboot<-100                #number of iterations for each
study condition
MDT<-0                    #MDT: Complete Data=0, LW=1, Mean
Imp=2, Grp Mean=3, EM=4, MI=5
NumMDT<-5                #Number of MDT's; use for looping
through MDT's for each sample
n.var<-4                  #number of study variables
per.miss<-c(.1, .25, .4)  #percent missingness
n.schools<-c(20,30,90,160) #number of schools sampled (level
2 sample size)
n.var.1<-(n.var+1)        #number of study variables + 1 for
evaluation against J or level 2 school units
n2.predictors<-c(2)       #number of level 2 predictors
n2.1<-(n2.predictors+1)  #number of level 2
explanatory variables + 1 for evaluation against J or level 2 school units
n2.miss<-c(1,2)          #number of level 2 predictors
w/missing values

#### Study conditions matrix
study.conditions<-expand.grid(per.miss,n.schools,n2.predictors,n2.miss)
n.study.conditions<-nrow(study.conditions)

study.conditions.current<-matrix(nrow=1,ncol=4)

#### Creating matrix for collection of model parameters
results.out1<-matrix(nrow=0,ncol=25)

for (i in 1:nrow(study.conditions)) {
  iteration1<-i
  n.study.condition<-i

```



```

        study.conditions.current<-study.conditions[i,1:4]      #looping
through study conditions matrix
        print(study.conditions.current)

#### Define number of schools
n.schools<-study.conditions.current[[2]]

#### Define % missingness
per.miss<-study.conditions.current[[1]]

#### Number of level-2 predictors
n2.predictors<-study.conditions.current[[3]]

#### Number of predictors with missing data
n2.miss<-study.conditions.current[[4]]

#### Number of iterations
results.out<-matrix(nrow=(NumMDT+1),ncol=25)
dimnames(results.out)<-list(NULL,
c("MDT","Condition","Iter","N.Schools","Per.Miss","Model","N2.Miss","Int.V
ar",
  "Slp.Var", "Cov", "Var.L1", "Int", "MEANSES", "CSES",
  "MEANSESxCSES", "Int.SE", "MEANSES.SE", "CSES.SE",
  "MEANSESxCSES.SE", "Var.L1.SE",
"Slp.Var.SE","SECTOR","SECTORxCSES","SECTOR.SE","SECTORxCSES.SE"))
results.out2<-matrix(nrow=1,ncol=25)

for (i in 1:nboot)      {

iteration<-i
#### Randomly pick "X" number of schools
result<-try(
  #### try is beginning of error handling code
  {sample.schools<-sample(levels(hsb.data2$Var1), n.schools, replace=F)
  hsb.new<-data.frame()
  for (i in sample.schools){
  hsb.new<-rbind(hsb.new, hsb.data[hsb.data$SCHOOL==i,])
  }
  hsb.new$SCHOOL<-factor(hsb.new$SCHOOL)
  #### Complete data

J<-0

repeat{

#### Model missingness for 2 variables MEANSES and SECTOR independently
#### Randomly select "X" % of the schools for % missingness on MEANSES
samp.schools<-data.frame(unique(hsb.new$SCHOOL)[sample(n.schools,
  round(n.schools * (1-per.miss))),])

#### Randomly select "X" % of the schools for % missingness on SECTOR
samp.schools2<-data.frame(unique(hsb.new$SCHOOL)[sample(n.schools,
  round(n.schools * (1-per.miss*(n2.miss-1))),)])

```

```

#### Merge two datasets before using subset to remove N/A else school that
should be deleted may appear? no subset takes care of this on successive
code

#### Subset function removes N/A values from hsb.new
new.without.NA<-subset(hsb.new, SCHOOL%in%samp.schools)
new.without.NA$MEANSES.miss<-new.without.NA$MEANSES

#### Subset function removes N/A values from hsb.new; create new variable
SECTOR.miss so schools not selected on MEANSES sample will have a NA value
when samples are merged, else would only be removing schools from analysis
that had missing values on both SECTOR and MEANSES
new.without.NA2<-subset(hsb.new, SCHOOL%in%samp.schools2)
new.without.NA2$SECTOR.miss<-new.without.NA2$SECTOR

#### Merge two datasets with missingness on MEANSES and SECTOR variables
new.without.NA<-merge(new.without.NA,new.without.NA2, all=T)
new.without.NA<-subset(new.without.NA,new.without.NA$MEANSES.miss!="NA")
new.without.NA<-subset(new.without.NA,new.without.NA$SECTOR.miss!="NA")

#### Number of level 2 or school units in new.without.NA
J.data<-data.frame(table(new.without.NA$SCHOOL))
J<-nrow(data.frame(J.data[,2][J.data[,2]>0]))

#### Check that samp.schools has enough Level 2 units (SCHOOL) to run
model J > p+1 or J>n2.1
if(J>n.var.1) break

}

#### Looping through MDT's for each school sample
MDT<-0                                     ##### Everytime new
sample of schools made, recycle through MDT's for new sample
for (i in 1:(NumMDT+1)) {

iteration2<-i
#### If complete data, just run model on hsb.new dataframe
{if(MDT==0)
  model.2<-
lme(data=hsb.new,MATHACH~CSES*(MEANSES+SECTOR),random=~CSES|SCHOOL,
  control=lmeControl(opt="optim"))

#### Listwise deletion
else if(MDT==1) {

#### Running model 2 for list-wise deletion
model.2<-
lme(data=new.without.NA,MATHACH~CSES*(MEANSES+SECTOR),random=~CSES|SCHOOL,
  control=lmeControl(opt="optim"))
}

#### Mean imputation
else if(MDT==2) {

```

```

#### Subset function removes N/A values from hsb.new
new.without.NA<-subset(hsb.new, SCHOOL%in%samp.schools)
new.without.NA$MEANSES.miss<-new.without.NA$MEANSES      #### add a
variable that hsb.new doesn't contain so when merging the two dataframes,
will have missing values for imputation in MEAN.SES.miss (MEAN.SES is
first variable modeling missingness for)
new.with.NA<-merge(hsb.new,new.without.NA,all=T)          #### merging two
datasets will keep only unique values; contains NA's for running
imputation on

#### Mean imputation for MEANSES
new.with.NA$MEANSES.mean.impute<-recode(new.with.NA$MEANSES.miss,
  "NA=summary(new.with.NA$MEANSES.miss)[[4]];
else=new.with.NA$MEANSES.miss")

#### Subset function removes N/A values from hsb.new; create new variable
SECTOR.miss so schools not selected on MEANSES sample will have a NA value
when samples are merged, else would only be removing schools from analysis
that had missing values on both SECTOR and MEANSES
new.without.NA2<-subset(hsb.new, SCHOOL%in%samp.schools2)
new.without.NA2$SECTOR.miss<-new.without.NA2$SECTOR
new.with.NA2<-merge(hsb.new,new.without.NA2,all=T)

#### Mean imputation for SECTOR
new.with.NA2$SECTOR.mean.impute<-recode(new.with.NA2$SECTOR.miss,
  "NA=round(summary(new.with.NA2$SECTOR.miss),0)[[4]];
else=new.with.NA2$SECTOR.miss")

#### Merge two datasets with missingness on MEANSES and SECTOR variables
new.without.NA3<-merge(new.without.NA,new.without.NA2, all=T)
new.without.NA3<-
subset(new.without.NA3,new.without.NA3$MEANSES.miss!="NA")
new.without.NA3<-subset(new.without.NA3,new.without.NA3$SECTOR.miss!="NA")

new.with.NA3<-merge(new.with.NA,new.with.NA2, all=T)

#### Running model 2 for mean imputation
model.2<-
lme(data=new.with.NA3,MATHACH~CSES*(MEANSES.mean.impute+SECTOR.mean.impute
),random=~CSES|SCHOOL,
  control=lmeControl(opt="optim"))
  }

#### Group mean imputation
else if(MDT==3)      {

#### Creating matrix with MEANSES missing values
new.without.NA<-subset(hsb.new, SCHOOL%in%samp.schools)  #### Subset
function removes N/A values from hsb.new
new.without.NA$MEANSES.miss<-new.without.NA$MEANSES      #### add a
variable that hsb.new doesn't contain so when merging the two dataframes,
will have missing values for imputation in MEANSES.miss (MEANSES is first
variable modeling missingness for)

```

```

new.with.NA<-merge(hsb.new,new.without.NA,all=T)          ##### merging data
such that NA's are now shown

##### Creating matrix with SECTOR missing values
new.without.NA3<-subset(hsb.new, SCHOOL%in%samp.schools2) ##### Subset
function removes N/A values from hsb.new
new.without.NA3$SECTOR.miss<-new.without.NA3$SECTOR      ##### add a
variable that hsb.new doesn't contain so when merging the two dataframes,
will have missing values for imputation in SECTOR.miss (SECTOR is second
variable modeling missingness for)
new.with.NA3<-merge(hsb.new,new.without.NA3,all=T)      ##### merging data
such that NA's are now shown

##### Merging matrices with missing values for MEANSES and SECTOR
new.with.NA.2<-merge(new.with.NA,new.with.NA3,all=T)      #####
merging dataframes for missingness on MEANSES and SECTOR
new.with.NA.2$SCHOOL<-as.numeric(new.with.NA.2$SCHOOL)    ##### must
make SCHOOL variable numeric to run imputation model
new.with.NA.2<-cbind(new.with.NA.2[,1:2],new.with.NA.2[,4:16]) #####
removing "constant" variable CONS from data to run imputation model

##### Aggregating level 2 data so that imputation is run at level 2
L2.with.NA<-
data.frame(cbind(new.with.NA.2[1],new.with.NA.2[6],new.with.NA.2[9:15]))
##### Selecting L2 variables
L2.with.NA2<-
data.frame(aggregate(L2.with.NA,list(L2.with.NA$SCHOOL),mean))[,2:10]
##### Aggregating L2 data by school

##### If 2 variables missing, chance that splitting data on SECTOR to
compute meanses will result in error as SECTOR could also be missing, use
PRACAD as best overall correlation with both MEANSES and SECTOR (>0.6 for
both)
L2.with.NA2.0<-
data.frame(L2.with.NA2$MEANSES.miss[L2.with.NA2$PRACAD<=.5]) #####
Filtering MEANSES.miss values for PRACAD <= 0.5
##### Splitting data on PRACAD to compute means
L2.with.NA2.1<-data.frame(L2.with.NA2$MEANSES.miss[L2.with.NA2$PRACAD>.5])
##### Filtering MEANSES.miss values for PRACAD > 0.5

MEANSES.imp.0<-apply(L2.with.NA2.0,2,mean, na.rm=T)
##### Mean value for MEANSES when PRACAD <=0.5
MEANSES.imp.1<-apply(L2.with.NA2.1,2,mean, na.rm=T)
##### Mean value for MEANSES when PRACAD >0.5

new.with.NA.2$MEANSES.imp<-new.with.NA.2$MEANSES.miss
##### Create a new variable for MEANSES based on PRACAD dummy
variable
new.with.NA.2$MEANSES.imp[new.with.NA.2$PRACAD<=.5]<-MEANSES.imp.0
##### Recode MEANSES as mean value based on MEANSES for PRACAD <=0.5
new.with.NA.2$MEANSES.imp[new.with.NA.2$PRACAD>.5]<-MEANSES.imp.1
##### Recode MEANSES as mean value based on MEANSES for PRACAD >0.5

```

```

L2.with.NA2.S0<-
data.frame(L2.with.NA2$SECTOR.miss[L2.with.NA2$PRACAD<=.5]) #####
Filtering SECTOR.miss values for PRACAD <= 0.5
##### Splitting data on PRACAD to compute means
L2.with.NA2.S1<-data.frame(L2.with.NA2$SECTOR.miss[L2.with.NA2$PRACAD>.5])
##### Filtering SECTOR.miss values for PRACAD > 0.5

SECTOR.imp.0<-round(apply(L2.with.NA2.S0,2,mean, na.rm=T))
##### Rounded mean value for SECTOR when PRACAD <=0.5
SECTOR.imp.1<-round(apply(L2.with.NA2.S1,2,mean, na.rm=T))
##### Rounded mean value for SECTOR when PRACAD >0.5

new.with.NA.2$SECTOR.imp<-new.with.NA.2$SECTOR.miss
##### Create a new variable for SECTOR based on PRACAD dummy variable
new.with.NA.2$SECTOR.imp[new.with.NA.2$PRACAD<=.5]<-SECTOR.imp.0
##### If PRACAD <= 0.5, recode SECTOR as 0
new.with.NA.2$SECTOR.imp[new.with.NA.2$PRACAD>.5]<-SECTOR.imp.1
##### If PRACAD > 0.5, recode SECTOR as 1

new.with.NA.2$MEANSES.imp[!is.na(new.with.NA.2$MEANSES.miss)]<-
##### If MEANSES missing, impute value based on PRACAD, else use
value in dataset
new.with.NA.2$MEANSES.miss[!is.na(new.with.NA.2$MEANSES.miss)]

new.with.NA.2$SECTOR.imp[!is.na(new.with.NA.2$SECTOR.miss)]<-
##### If SECTOR missing, impute value based on PRACAD, else use value
in dataset
new.with.NA.2$SECTOR.miss[!is.na(new.with.NA.2$SECTOR.miss)]

GrpMean.dataset<-new.with.NA.2

##### Running model 2 for Grp Mean Imputation
model.2<-
lme(data=GrpMean.dataset,MATHACH~CSES*(MEANSES.imp+SECTOR.imp),random=~CSE
S|SCHOOL,
control=lmeControl(opt="optim"))
}

##### EM Algorithm
else if(MDT==4) {

##### Creating matrix with MEANSES missing values
new.without.NA<-subset(hsb.new, SCHOOL%in%samp.schools) ##### Subset
function removes N/A values from hsb.new
new.without.NA$MEANSES.miss<-new.without.NA$MEANSES ##### add a
variable that hsb.new doesn't contain so when merging the two dataframes,
will have missing values for imputation in MEANSES.miss (MEANSES is first
variable modeling missingness for)
new.with.NA<-merge(hsb.new,new.without.NA,all=T) ##### merging data
such that NA's are now shown

##### Creating matrix with SECTOR missing values
new.without.NA3<-subset(hsb.new, SCHOOL%in%samp.schools2)##### Subset
function removes N/A values from hsb.new

```

```

new.without.NA3$SECTOR.miss<-new.without.NA3$SECTOR      ##### add a
variable that hsb.new doesn't contain so when merging the two dataframes,
will have missing values for imputation in SECTOR.miss (SECTOR is second
variable modeling missingness for)
new.with.NA3<-merge(hsb.new,new.without.NA3,all=T)        ##### merging data
such that NA's are now shown

##### Merging matrices with missing values for MEANSES and SECTOR
new.with.NA.2<-merge(new.with.NA,new.with.NA3,all=T)      #####
merging dataframes for missingness on MEANSES and SECTOR
new.with.NA.2$SCHOOL<-as.numeric(new.with.NA.2$SCHOOL)    ##### must make
SCHOOL variable numeric to run imputation model
new.with.NA.2<-cbind(new.with.NA.2[,1:2],new.with.NA.2[,4:16]) #####
removing "constant" variable CONS from data to run imputation model

##### Aggregating level 2 data so that imputation is run at level 2
L2.with.NA<-
data.frame(cbind(new.with.NA.2[1],new.with.NA.2[6],new.with.NA.2[9:15]))
##### Selecting L2 variables
L2.with.NA2<-
data.frame(aggregate(L2.with.NA,list(L2.with.NA$SCHOOL),mean))[,2:10]
##### Aggregating L2 data by school
L2.with.NA3<-cbind(L2.with.NA2[1],L2.with.NA2[3],L2.with.NA2[5:9])

##### EM for missing values for MEANSES and SECTOR
EM.list.L2<-
amelia(x=L2.with.NA3,idvars=c("SCHOOL"),tolerance=.00001,noms="SECTOR.miss
",ords="HIMINTY")
EM.list.L2.no.NA<-EM.list.L2$imputations[!is.na(EM.list.L2$imputations)]
##### Extracting imputations yielding NAs

###Creating dataframe of logical imputations, using mean of EM estimated
values for MEANSES imputation for each school
EM.imp.num<-dim(matrix(EM.list.L2.no.NA))[1]
##### Variable for number of logical imputations derived from EM
EM.MEANSES.imp<-matrix(nrow=n.schools,ncol=EM.imp.num)
##### Creating matrix for mean imputations
EM.SECTOR.imp<-matrix(nrow=n.schools,ncol=EM.imp.num)

for (i in 1:EM.imp.num){
##### Loop required for selecting only logical values produced from
EM
  EM.MEANSES.imp[,i]<-data.frame(EM.list.L2.no.NA[i])[,6]
##### Dataframe of mean imputations
  EM.SECTOR.imp[,i]<-data.frame(EM.list.L2.no.NA[i])[,7]
}

EM.MEANSES.imp<-data.frame(rowMeans(EM.MEANSES.imp))
#####Mean of EM imputations for MEANSES
EM.SECTOR.imp<-data.frame(round(rowMeans(EM.SECTOR.imp),0))
#####Mean of EM imputations for SECTOR

EM.imp<-data.frame(cbind(EM.MEANSES.imp,EM.SECTOR.imp))

```

```

EM.imp.L2<-data.frame(cbind(L2.with.NA2,EM.imp))
#####Creating dataframe w/ L2 data and imputed MEANSES and SECTOR
values

###Merge aggregate L2 data with imputed MEANSES and SECTOR values with
dataset to be used in lme
EM.dataset<-merge(EM.imp.L2,new.with.NA.2,all=T)
EM.dataset$MEANSES.imp<-EM.dataset$rowMeans.EM.MEANSES.imp.
EM.dataset$SECTOR.imp<-EM.dataset$round.rowMeans.EM.SECTOR.imp...0.

#### Running model 2 for EM
model.2<-
lme(data=EM.dataset,MATHACH~CSES*(MEANSES.imp+SECTOR.imp),random=~CSES|SCH
OOL,
  control=lmeControl(opt="optim"))
  }

#### Multiple imputation
else if(MDT==5) {

#### Creating matrix with MEANSES missing values
new.without.NA<-subset(hsb.new, SCHOOL%in%samp.schools) ##### Subset
function removes N/A values from hsb.new
new.without.NA$MEANSES.miss<-new.without.NA$MEANSES ##### add a
variable that hsb.new doesn't contain so when merging the two dataframes,
will have missing values for imputation in MEANSES.miss (MEANSES is first
variable modeling missingness for)
new.with.NA<-merge(hsb.new,new.without.NA,all=T) ##### merging data
such that NA's are now shown

#### Creating matrix with SECTOR missing values
new.without.NA3<-subset(hsb.new, SCHOOL%in%samp.schools2)##### Subset
function removes N/A values from hsb.new
new.without.NA3$SECTOR.miss<-new.without.NA3$SECTOR ##### add a
variable that hsb.new doesn't contain so when merging the two dataframes,
will have missing values for imputation in SECTOR.miss (SECTOR is second
variable modeling missingness for)
new.with.NA3<-merge(hsb.new,new.without.NA3,all=T) ##### merging data
such that NA's are now shown

#### Merging matrices with missing values for MEANSES and SECTOR
new.with.NA.2<-merge(new.with.NA,new.with.NA3,all=T) #####
merging dataframes for missingness on MEANSES and SECTOR
new.with.NA.2$SCHOOL<-as.numeric(new.with.NA.2$SCHOOL) ##### must make
SCHOOL variable numeric to run imputation model
new.with.NA.2<-cbind(new.with.NA.2[,1:2],new.with.NA.2[,4:16]) #####
removing "constant" variable CONS from data to run imputation model

#### Aggregating level 2 data so that imputation is run at level 2
L2.with.NA<-
data.frame(cbind(new.with.NA.2[1],new.with.NA.2[6],new.with.NA.2[9:15]))
#### Selecting L2 variables

```

```

L2.with.NA2<-
data.frame(aggregate(L2.with.NA,list(L2.with.NA$SCHOOL),mean))[,2:10]
#### Aggregating L2 data by school
L2.with.NA3<-cbind(L2.with.NA2[1],L2.with.NA2[3],L2.with.NA2[5:9])
      #### removing MEANSES and SECTOR so they aren't used in imputation
model

#### Multiple imputation for missing values for MEANSES and SECTOR
MI.imp<-mice(L2.with.NA3,maxit=5)                                     #####
Running multiple imputation (MI) on level 2 data
MI.imp.MEANSES<-data.frame(complete(MI.imp,"repeated")[,26:30])
      #### Generating imputed data for MEANSES
MI.imp.MEANSES<-data.frame(apply(MI.imp.MEANSES,1,mean,na.rm=T))
      #### Taking average of imputations
MI.imp.SECTOR<-data.frame(complete(MI.imp,"repeated")[,31:35])    #####
Generating imputed data for SECTOR
MI.imp.SECTOR<-data.frame(round(apply(MI.imp.SECTOR,1,mean,na.rm=T)))
      #### Taking average of imputations
MI.imp<-cbind(MI.imp.MEANSES,MI.imp.SECTOR)                         #####
Combining imputed SECTOR and MEANSES values
colnames(MI.imp)<-c("MEANSES.imp","SECTOR.imp")
      #### Defining column names for imputed dataset

MI.imp<-data.frame(cbind(L2.with.NA2[1:7],MI.imp[1:2]))            #####
Combine MI.imp with L2 data, removing common fields

###Merge aggregate L2 data with imputed MEANSES and SECTOR values with
dataset to be used in lme
MI.dataset<-merge(MI.imp,new.with.NA.2,all=T)                      ##### Merge
imputed L2 data with full dataset for lme

#### Running model 2 for EM
model.2<-
lme(data=MI.dataset,MATHACH~CSES*(MEANSES.imp+SECTOR.imp),random=~CSES|SCH
OOL,
  control=lmeControl(opt="optim"))
  }
}                                                                 ##### End of
if else statements block

#### Data collected for parameter matrix
intercept.variance<-as.double(VarCorr(model.2)[1])
slope.variance<-as.double(VarCorr(model.2)[2])
correlation<-as.double(VarCorr(model.2)[8])
covariance<-correlation*sqrt(intercept.variance)*sqrt(slope.variance)
variance.level.1<-as.double(VarCorr(model.2)[3])
intercept<-as.double(summary(model.2)$tTable[[1]])
mean.ses<-as.double(summary(model.2)$tTable[[3]])
cses<-as.double(summary(model.2)$tTable[[2]])
mean.sesXcses<-as.double(summary(model.2)$tTable[[5]])
intercept.se<-as.double(summary(model.2)$tTable[[7]])
meanses.se<-as.double(summary(model.2)$tTable[[9]])
cses.se<-as.double(summary(model.2)$tTable[[8]])

```



```

mean.sesXcses.se<-as.double(summary(model.2)$tTable[[11]])
variance.level.1.se<-
as.double(VarCorr(model.2)[6])/sqrt(as.double(model.2$dim[1]))
slope.variance.se<-
as.double(VarCorr(model.2)[5])/sqrt(as.double(model.2$dim[1]))
sector<-as.double(summary(model.2)$tTable[[4]])
sectorXcses<-as.double(summary(model.2)$tTable[[6]])
sector.se<-as.double(summary(model.2)$tTable[[10]])
sectorXcses.se<-as.double(summary(model.2)$tTable[[12]])
results.out2<-
cbind(MDT,n.study.condition,iteration,n.schools,per.miss,n2.predictors,n2.
miss,intercept.variance,slope.variance,covariance,
      variance.level.1,intercept,mean.ses,cses,mean.sesXcses,intercept.se,
      meanses.se,cses.se,mean.sesXcses.se,variance.level.1.se,slope.varian
ce.se,sector,sectorXcses,sector.se,sectorXcses.se)
results.out[iteration2,]<-results.out2          #####
Combining data for all MDT's for a given study condition
MDT<-MDT+1
}          ##### Ends
loop through MDT's for each sample

##### Combining data for each study condition

results.out1<-rbind(results.out,results.out1)})
if(inherits(result,"try-error"))next          ##### If
error, code shouldn't crash, should take another school sample
}          ##### Ends
loop for a sample, new sample of schools drawn
#####creating dataframe with all study conditions for all MDT's on Model 2
results.out.dataframe<-data.frame(results.out1)
results.out.L22_MISS1_2<-results.out.dataframe

#####Creating subj factor for unique iterations by condition (1000
iterations for each condition are akin to 1000 subjects for each condition
and must be unique to each condition)
results.out.L22_MISS1_2$subj<-(results.out.L22_MISS1_2$Condition-
1)*nboot+results.out.L22_MISS1_2$Iter
}

##### Need to examine results to determine if some iterations did not yield
results (skipped to next iteration/sample if error)
##### If missing, determine which conditions and iterations are missing
using MissingIterations code (will have to vary based on specifics of data
run)
diag.table<-table(results.out.L22_MISS1_2$N.Schools,
                  results.out.L22_MISS1_2$Per.Miss,
                  results.out.L22_MISS1_2$N2.Miss,
                  dnn = list("N.Schools","Per.Miss","N2.Miss"))

(diag.table<-diag.table/(nboot*6))          ##### Each
cell should equal nboot, if not, missing iteration for that condition

```

```

####Combining data for each subject on one row to handle doubly MANOVA
analysis in R(row for each subj contains lme model parameters calculated
for each MDT)
a2<-
split(results.out.L22_MISS1_2_temp3,results.out.L22_MISS1_2_temp3$subj)
      #####Split data by subject

####Fix row numbers to capture number of iterations
L22.data<-matrix(nrow=(100*24),ncol=114)
      #####Create matrix for L22 data by subject
dimnames(L22.data)<-list(NULL,
c("Condition", "Iter", "N.Schools", "Per.Miss", "Model", "N2.Miss",
  "Int.Var.0", "Int.Var.1", "Int.Var.2", "Int.Var.3", "Int.Var.4", "Int.Var
.5",
  "Slp.Var.0", "Slp.Var.1", "Slp.Var.2", "Slp.Var.3", "Slp.Var.4", "Slp.Var
.5",
  "Cov.0", "Cov.1", "Cov.2", "Cov.3", "Cov.4", "Cov.5",
  "Var.L1.0", "Var.L1.1", "Var.L1.2", "Var.L1.3", "Var.L1.4", "Var.L1.5",
  "Int.0", "Int.1", "Int.2", "Int.3", "Int.4", "Int.5",
  "MEANSES.0", "MEANSES.1", "MEANSES.2", "MEANSES.3", "MEANSES.4", "MEANSES
.5",
  "CSES.0", "CSES.1", "CSES.2", "CSES.3", "CSES.4", "CSES.5",
  "MEANSESxCSES.0", "MEANSESxCSES.1", "MEANSESxCSES.2", "MEANSESxCSES.3",
"MEANSESxCSES.4", "MEANSESxCSES.5",
  "Int.SE.0", "Int.SE.1", "Int.SE.2", "Int.SE.3", "Int.SE.4", "Int.SE.5",
  "MEANSES.SE.0", "MEANSES.SE.1", "MEANSES.SE.2", "MEANSES.SE.3", "MEANSES
.SE.4", "MEANSES.SE.5",
  "CSES.SE.0", "CSES.SE.1", "CSES.SE.2", "CSES.SE.3", "CSES.SE.4", "CSES.SE
.5",
  "MEANSESxCSES.SE.0", "MEANSESxCSES.SE.1", "MEANSESxCSES.SE.2", "MEANSES
xCSES.SE.3", "MEANSESxCSES.SE.4", "MEANSESxCSES.SE.5",
  "Var.L1.SE.0", "Var.L1.SE.1", "Var.L1.SE.2", "Var.L1.SE.3", "Var.L1.SE.4
", "Var.L1.SE.5",
  "Slp.Var.SE.0", "Slp.Var.SE.1", "Slp.Var.SE.2", "Slp.Var.SE.3", "Slp.Var
.SE.4", "Slp.Var.SE.5",
  "SECTOR.0", "SECTOR.1", "SECTOR.2", "SECTOR.3", "SECTOR.4", "SECTOR.5",
  "SECTORxCSES.0", "SECTORxCSES.1", "SECTORxCSES.2", "SECTORxCSES.3", "SEC
TORxCSES.4", "SECTORxCSES.5",
  "SECTOR.SE.0", "SECTOR.SE.1", "SECTOR.SE.2", "SECTOR.SE.3", "SECTOR.SE.4
", "SECTOR.SE.5",
  "SECTORxCSES.SE.0", "SECTORxCSES.SE.1", "SECTORxCSES.SE.2", "SECTORxCSE
S.SE.3", "SECTORxCSES.SE.4", "SECTORxCSES.SE.5"))

L22.data.temp<-matrix(nrow=1,ncol=108)
      #####Temp matrix for L21 data
par.subj.lmdt<-matrix(nrow=1,ncol=6)
      #####Matrix for parameter data for indiv subj on one MDT
par.subj.allmdt<-matrix(nrow=1,ncol=0)
      #####Combined parameter data matrix across all MDT's for indiv subj
i2<-1
      #####Counter for subj in parameter data collection

####Main loop to capture between subj variables, then loop through
parameter estimates for each subject on each MDT

```

```

for (i in 1:(100*24)){
    #####Loops = number of
    subj/iterations = iterations x # conditions
    i3<-i
    #####Counter for subj
    for betw subj variable collection
        L22.data.temp<-
    cbind(a2[[i3]][1,2],a2[[i3]][1,3],a2[[i3]][1,4],a2[[i3]][1,5],a2[[i3]][1,6
    ],a2[[i3]][1,7])
        #####Between subj variables

    #####Loop for parameter estimates for each subj
    for (i in 8:25){
        #####Columns 8:25 are
        lme parameter estimates
        i4<-i
        #####Counter for column
        of data (parameter) to be collected
        par.subj.lmdt<-
    cbind(a2[[i2]][1,i4],a2[[i2]][2,i4],a2[[i2]][3,i4],a2[[i2]][4,i4],a2[[i2]]
    [5,i4],a2[[i2]][6,i4])
        #####Parameter data for indiv subj on
        one MDT
        par.subj.allmdt<-cbind(par.subj.allmdt,par.subj.lmdt)
        #####Combining
        parameters for indiv subj across all MDT's
    }

    L22.data.temp<-cbind(L22.data.temp,par.subj.allmdt)
    #####Combinining betw subj data with parameter estimates for indiv
    subj
    L22.data[i3,]<-L22.data.temp
    #####Adding line of data for indiv subj to L21.data matrix
    i2<-i2+1
    #####Move to next subj for parameter data collectin
    par.subj.allmdt<-matrix(nrow=1,ncol=0)
    #####Reset combined parameter data matrix so it is empty
}

L22.data<-data.frame(L22.data)
    #####Changing L21.data to dataframe
attach(L22.data)
    #####Attach L21.data for MANOVA's
L22.data$Per.Miss<-factor(L22.data$Per.Miss)
    #####Converting numeric betw subj variables to factors for use in
    MANOVA
L22.data$N.Schools<-factor(L22.data$N.Schools)

##### Dependent variables for MANOVAs, use cbind for ease of reading
Model.2.Y1<-
cbind(MEANSES.0,MEANSES.1,MEANSES.2,MEANSES.3,MEANSES.4,MEANSES.5)

Model.2.Y1.2<-
cbind(MEANSES.0,MEANSES.1,MEANSES.2,MEANSES.3,MEANSES.4,MEANSES.5,
    SECTOR.0,SECTOR.1,SECTOR.2,SECTOR.3,SECTOR.4,SECTOR.5)[N2.Miss==2,]

```

```

Model.2.Y2<-
cbind(Int.0,Int.1,Int.2,Int.3,Int.4,Int.5,CSES.0,CSES.1,CSES.2,CSES.3,CSES
.4,CSES.5,
      MEANSESxCSES.0,MEANSESxCSES.1,MEANSESxCSES.2,MEANSESxCSES.3,MEANSESx
CSES.4,MEANSESxCSES.5,
      SECTORxCSES.0,SECTORxCSES.1,SECTORxCSES.2,SECTORxCSES.3,SECTORxCSES.
4,SECTORxCSES.5)

Model.2.Y2.2<-
cbind(Int.0,Int.1,Int.2,Int.3,Int.4,Int.5,CSES.0,CSES.1,CSES.2,CSES.3,CSES
.4,CSES.5,
      MEANSESxCSES.0,MEANSESxCSES.1,MEANSESxCSES.2,MEANSESxCSES.3,MEANSESx
CSES.4,MEANSESxCSES.5,
      SECTOR.0,SECTOR.1,SECTOR.2,SECTOR.3,SECTOR.4,SECTOR.5,
      SECTORxCSES.0,SECTORxCSES.1,SECTORxCSES.2,SECTORxCSES.3,SECTORxCSES.
4,SECTORxCSES.5) [N2.Miss==1,]

Model.2.Y3<-
cbind(Int.Var.0,Int.Var.1,Int.Var.2,Int.Var.3,Int.Var.4,Int.Var.5,
      Slp.Var.0,Slp.Var.1,Slp.Var.2,Slp.Var.3,Slp.Var.4,Slp.Var.5,
      Var.L1.0,Var.L1.1,Var.L1.2,Var.L1.3,Var.L1.4,Var.L1.5,
      Cov.0,Cov.1,Cov.2,Cov.3,Cov.4,Cov.5)
edit(L22.data)
#### MANOVA 4 code (MANOVA4 MEANSES main effect w/N2.Miss as factor;
MANOVA4.1 SECTOR added as main effect but only across N2.Miss=2)

#### Creating within subj matrix; Uses repeated contrasts compare means
across MDT's (within subj factor)
WI.Matrix1<-matrix(c(
  -1, 0, 0, 0, 0,          ##### Complete data
   1,-1, 0, 0, 0,          ##### LW
   0, 1,-1, 0, 0,          ##### Mean Imp
   0, 0, 1,-1, 0,          ##### Grp Mean Imp
   0, 0, 0, 1,-1,          ##### EM
   0, 0, 0, 0, 1),6,5,byrow=T) ##### MI

colnames(WI.Matrix1)<-
cbind("MEANSES.C1","MEANSES.C2","MEANSES.C3","MEANSES.C4","MEANSES.C5")

#### Multivariate test for all contrasts, shows if any contrasts
statistically sig
MANOVA4.mod<-lm(Model.2.Y1~Per.Miss*N.Schools*N2.Miss,data=L22.data)
##### Multivariate regression
(MANOVA4<-Anova(MANOVA4.mod,imatrix=list(MDT=WI.Matrix1),test="Wilks"))
##### Uses imatrix command to identify within
subjects effects

##### Using imatrix with Anova function in place of
contrasts below as it allows for labeling of within subj line on output

#### Univariate test for each contrast(labeled as response in summary
output), shows which contrast is statistically sig, remember to use
Bonferroni correction to adj for Type I error rate

```

```

ANOVA4.mod<-aov(Model.2.Y1%*%WI.Matrix1~Per.Miss*N.Schools*N2.Miss,
data=L22.data)
(ANOVA4<-summary(ANOVA4.mod,intercept=T))

MANOVA4etasq<-etasq(MANOVA4,anova=TRUE,partial=TRUE)

WI.Matrix4<-kronecker(diag(2),WI.Matrix1)
#### kronecker function expands matrix for multiple DV's

colnames(WI.Matrix4)<-
cbind("MEANSES.C1","MEANSES.C2","MEANSES.C3","MEANSES.C4","MEANSES.C5",
"SECTOR.C1","SECTOR.C2","SECTOR.C3","SECTOR.C4","SECTOR.C5")

L22.data.N2.2<-data.frame(L22.data[L22.data$N2.Miss==2,])
#### Creating new dataframe for L22 data split by N2.Miss

#### Multivariate test for all contrasts, shows if any contrasts
statistically sig
MANOVA4.1mod<-
lm(Model.2.Y1.2~L22.data.N2.2$Per.Miss*L22.data.N2.2$N.Schools,data=L22.da
ta.N2.2)
#### Multivariate regression
(MANOVA4.1<-Anova(MANOVA4.1mod,imatrix=list(MDT=WI.Matrix4),test="Wilks"))
#### Uses imatrix command to identify within
subjects effects

#### Using imatrix with Anova function in place of
contrasts below as it allows for labeling of within subj line on output

#### Univariate test for each contrast(labeled as response in summary
output), shows which contrast is statistically sig, remember to use
Bonferroni correction to adj for Type I error rate
ANOVA4.1mod<-
aov(Model.2.Y1.2%*%WI.Matrix4~L22.data.N2.2$Per.Miss*L22.data.N2.2$N.Schoo
ls, data=L22.data.N2.2)
(ANOVA4.1<-summary(ANOVA4.1mod,intercept=T))

MANOVA4.1etasq<-etasq(MANOVA4.1,anova=TRUE,partial=TRUE)

#### MANOVA 5 code (MANOVA5 fixed effects without missing data over all
conditions; MANOVA5.1 adds SECTOR but only with N2.Miss==1)

WI.Matrix5<-kronecker(diag(4),WI.Matrix1)
#### kronecker function expands matrix for multiple DV's

colnames(WI.Matrix5)<-cbind("Int.C1","Int.C2","Int.C3","Int.C4","Int.C5",
"CSES.C1","CSES.C2","CSES.C3","CSES.C4","CSES.C5",
"MEANSESxCSES.C1","MEANSESxCSES.C2","MEANSESxCSES.C3","MEANSESxCSES.
C4","MEANSESxCSES.C5",
"SECTORxCSES.C1","SECTORxCSES.C2","SECTORxCSES.C3","SECTORxCSES.C4",
"SECTORxCSES.C5")

#### Multivariate test for all contrasts, shows if any contrasts
statistically sig

```

```

MANOVA5.mod<-lm(Model.2.Y2~Per.Miss*N.Schools*N2.Miss,data=L22.data)
##### Multivariate regression
(MANOVA5<-Anova(MANOVA5.mod,imatrix=list(MDT=WI.Matrix5),test="Wilks"))
##### Uses imatrix command to identify within
subjects effects

##### Using imatrix with Anova function in place of
contrasts below as it allows for labeling of within subj line on output

##### Univariate test for each contrast(labeled as response in summary
output), shows which contrast is statistically sig, remember to use
Bonferroni correction to adj for Type I error rate
ANOVA5.mod<-aov(Model.2.Y2%*%WI.Matrix5~Per.Miss*N.Schools*N2.Miss,
data=L22.data)
(ANOVA5<-summary(ANOVA5.mod,intercept=T))

MANOVA5etasq<-etasq(MANOVA5,anova=TRUE,partial=TRUE)

WI.Matrix5.1<-kronecker(diag(5),WI.Matrix1)

colnames(WI.Matrix5.1)<-
cbind("Int.C1","Int.C2","Int.C3","Int.C4","Int.C5",
      "CSES.C1","CSES.C2","CSES.C3","CSES.C4","CSES.C5",
      "MEANSESxCSES.C1","MEANSESxCSES.C2","MEANSESxCSES.C3","MEANSESxCSES.
C4","MEANSESxCSES.C5",
      "SECTOR.C1","SECTOR.C2","SECTOR.C3","SECTOR.C4","SECTOR.C5",
      "SECTORxCSES.C1","SECTORxCSES.C2","SECTORxCSES.C3","SECTORxCSES.C4",
"SECTORxCSES.C5")

L22.data.N2.1<-data.frame(L22.data[L22.data$N2.Miss==1,])
##### Creating new dataframe for L22 data split by
N2.Miss

##### Multivariate test for all contrasts, shows if any contrasts
statistically sig
MANOVA5.1mod<-
lm(Model.2.Y2.2~L22.data.N2.1$Per.Miss*L22.data.N2.1$N.Schools,data=L22.da
ta.N2.1)
##### Multivariate regression
(MANOVA5.1<-
Anova(MANOVA5.1mod,imatrix=list(MDT=WI.Matrix5.1),test="Wilks"))
##### Uses imatrix command to identify within subjects
effects

##### Using imatrix with Anova function in place of
contrasts below as it allows for labeling of within subj line on output

##### Univariate test for each contrast(labeled as response in summary
output), shows which contrast is statistically sig, remember to use
Bonferroni correction to adj for Type I error rate
ANOVA5.1mod<-
aov(Model.2.Y2.2%*%WI.Matrix5.1~L22.data.N2.1$Per.Miss*L22.data.N2.1$N.Sch
ools, data=L22.data.N2.1)
(ANOVA5.1<-summary(ANOVA5.1mod,intercept=T))

```

```

MANOVA5.1etasq<-etasq(MANOVA5.1,anova=TRUE,partial=TRUE)

#### MANOVA 6 code on random effects as DV's
WI.Matrix6<-kronecker(diag(4),WI.Matrix1)
##### kronecker function expands matrix for multiple
DV's

colnames(WI.Matrix6)<-
cbind("Int.Var.1","Int.Var.2","Int.Var.3","Int.Var.4","Int.Var.5",
      "Slp.Var.1","Slp.Var.2","Slp.Var.3","Slp.Var.4","Slp.Var.5",
      "Var.L1.1","Var.L1.2","Var.L1.3","Var.L1.4","Var.L1.5",
      "Cov.1","Cov.2","Cov.3","Cov.4","Cov.5")

#### Multivariate test for all contrasts, shows if any contrasts
statistically sig
MANOVA6.mod<-lm(Model.2.Y3~Per.Miss*N.Schools*N2.Miss,data=L22.data)
##### Multivariate regression
(MANOVA6<-Anova(MANOVA6.mod,imatrix=list(MDT=WI.Matrix6),test="Wilks"))
##### Uses imatrix command to identify within
subjects effects

##### Using imatrix with Anova function in place of
contrasts below as it allows for labeling of within subj line on output

#### Univariate test for each contrast(labeled as response in summary
output), shows which contrast is statistically sig, remember to use
Bonferroni correction to adj for Type I error rate
ANOVA6.mod<-aov(Model.2.Y3%*%WI.Matrix6~Per.Miss*N.Schools*N2.Miss,
data=L22.data)
(ANOVA6<-summary(ANOVA6.mod,intercept=T))

MANOVA6etasq<-etasq(MANOVA6,anova=TRUE,partial=TRUE)

```

REFERENCES

- Allison, P. D. (2002). *Missing data*. Thousand Oaks, CA: Sage Publications.
- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G.A. Marcoulides & R.E. Schumacker (Eds.) *Advanced structural equation modeling: Issues and techniques* (pp. 243-277). Hillsdale, NJ: Erlbaum.
- Baraldi, A.N., & Enders, C.K. (2010). An introduction to modern missing data analysis. *Journal of School Psychology, 48*, 5-37.
- Bernstein, L., & Burstein, N. (1994). *Bias vs. precision: Combining estimates in multisite evaluation research*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.
- Bodner, T. E. (2006). Missing data: Prevalence and reporting practices. *Psychological Reports, 99*, 675-680.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. (2nd ed.) Hillsdale, NJ: Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.
- Davey, A., Shanahan, M. J., & Schafer, J. L. (2001). Correcting for selective nonresponse in the national longitudinal survey of youth using multiple imputation. *Journal of Human Resources, 36*, 500-519.
- Delucchi, K. L. (1994). Methods for the analysis of binary outcome results in the presence of missing data. *Journal of Consulting and Clinical Psychology, 62*, 569-575.
- Dielman, T. (2001). *Applied regression analysis for business and economics*. (3rd ed.). Pacific Grove, CA: Duxbury Press.
- Enders, C. (2001). A primer on maximum likelihood algorithms available for use with missing data. teacher's corner. *Structural Equation Modeling, 8*, 128-141.
- Enders, C., & Bandalos, D. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*, 430-457.
- Fichman, M., & Cummings, J. N. (2003). Multiple imputation for missing data: Making the most of what you know. *Organizational Research Methods, 6*, 282.

- Furlow, C. F., Fouladi, R., & Whittaker, T. (August, 2003). *The impact of missing data, and procedures for handling missing data using an item response theory framework*. Paper presented at the Joint Statistical Meetings, San Francisco, CA.
- Garrett, P. (2009). *A Monte Carlo study investigating missing data, differential item functioning, and effect size* (Doctoral dissertation). Available from ProQuest. (Document ID 1996754181)
- Gibson, N. M., & Olejnik, S. (2003). Treatment of missing data at the second level of hierarchical linear models. *Educational & Psychological Measurement, 63*, 204.
- Gleason, T. C., & Staelin, R. (1975). A proposal for handling missing data. *Psychometrika, 40*, 229-252.
- Graham, J. W., & Hofer, S. M. (1993). EMCOV reference manual [Computer software]. Los Angeles: University of Southern California, Institute for Prevention Research.
- Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research, 31*, 197-218.
- Graham, J. W., Taylor, B. J., Olchowski, A. E., Cumsille, P. E., (2006). Planned missing data designs in psychological research. *Psychological Methods, 11*, 323-343.
- Gregorich, S. (1997). EM_COVAR.sas: A SAS macro for estimating a covariance matrix and mean vector via the EM algorithm [Computer program]. Retrieved from <http://mywebpage.netscape.com/segregorich/missing.html>
- Haitovsky, Y. (1968). Missing data in regression analysis. *Journal of the Royal Statistical Society, Series B, 30*, 67-82.
- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2010). *Multilevel modeling and longitudinal modeling with IBM SPSS*. New York: Routledge.
- Hedeker, D. & Gibbons, R.D. (2006): *Longitudinal data analysis*. Hoboken, NJ: John Wiley and Sons.
- Heitjan, D. F. (1997). Annotation: What can be done about missing data? Approaches to imputation. *American Journal of Public Health, 87*, 548-548.
- Hill, P., & Goldstein, H. (1998). Multilevel modeling of educational data with cross- classification and missing identification for units. *Journal of Educational and Behavioral Statistics, 23*, 117-128.
- Honaker, J., King, G., Blackwell, M. (2010). Amelia II: A Program for Missing Data (Version 1.2-18) [Computer software package]. Retrieved from <http://www.r-project.org/>
- Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

- Kromrey, J. D., & Hines, C. V. (1994). Nonrandomly missing data in multiple regression: An empirical comparison of common missing-data treatments. *Educational and Psychological Measurement, 54*, 573-593.
- Little, R. J. A., & Rubin, D. B. (1999). Comment. *Journal of the American Statistical Association, 94*, 1130-1132.
- Little, R. J. A., Rubin, D. B. (2002): *Statistical analysis with missing data* (2nd ed.). New York: John Wiley and Sons.
- Longford, N. T., Ely, M., Hardy, R., & Wadsworth, M. E. J. (2000). Handling missing data in diaries of alcohol consumption. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 163*, 381.
- Muthen, B., Kaplan, D., & Hollis, M. (1987). On the structural equation modeling with data that are not missing completely at random. *Psychometrika, 51*, 431-462.
- National Center for Education Statistics (1982). *High school and beyond* [NCES 95361]. Available via restricted-use data license at <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=95361>
- Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practice and suggestions for improvement. *Review of Educational Research, 74*, 525-556.
- Pigott, T. D. (2001). A review of methods for missing data. *Educational Research and Evaluation: An International Journal on Theory and Practice, 7*, 353-383.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*, (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Raymond, M. R. (1986). Missing data in evaluation research. *Evaluation and the Health Professions, 9*, 395-420.
- Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology, 47*, 537-560.
- Roth, P. L., & Switzer, F. S. (1995). A Monte Carlo analysis of missing data techniques in a HRM setting. *Journal of Management, 21*, 1003-1023.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581-592.
- Savalei, V., & Yuan, K.-H. (2009). On the model-based bootstrap with missing data: obtaining a p-value for a test of exact fit. *Multivariate Behavioral Research, 44*, 741-763.

- Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: a data analyst's perspective. *Multivariate Behavioral Research, 33*, 545-571.
- Schafer, J.L. (1997). *Analysis of incomplete multivariate data*. London: Chapman and Hall / CRC Press.
- Schafer, J. L. (1994). NORM: Multiple imputation of incomplete multivariate data under a normal model, version 2 [Computer software]. Software for Windows 95/98/NT. Available: <http://www.stat.psu.edu/~jls/misoftwa.html>
- Shin, T., Davison, M. L., & Long, J. D. (2009). Effects of missing data methods in structural equation modeling with nonnormal longitudinal data. *Structural Equation Modeling, 16*, 70-98.
- Shin, Y., & Raudenbush, S. W. (2010). A latent cluster-mean approach to the contextual effects model with missing data. *Journal of Educational and Behavioral Statistics, 35*, 26-53.
- Singer, J. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23*, 323-355.
- Song, X. Y., & Lee, S. Y. (2008). A Bayesian approach for analyzing hierarchical data with missing outcomes through structural equation models. *Structural Equation Modeling, 15*, 272-300.
- Snijders, T., & Bosker, R. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics, 18*, 237-259.
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Multiple imputation by chained equations (MICE) (Version 2.5) [Computer software package]. Retrieved from <http://www.r-project.org/>
- von Hippel, P. T. (2003). Data analysis with missing values. Retrieved from <http://www.sociology.ohio-state.edu/people/ptv/fag/missing/missing.ppt>
- Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.
- Wood, A. M., White, I. R., Thompson, S. G. (2004). Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials Review, 1*, 368-376.
- Wothke, W. (2000). Longitudinal and multi-group modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multiple group data: Practical issues, applied approaches and specific examples*. Mahwah, NJ: Lawrence Erlbaum Associates Inc., 219-240.
- Xiao-Li Meng. (2000). Missing data: Dial M for ??? *Journal of the American Statistical Association, 95*, 1325-1330.

Yuan, K. H., & Lu, L. (2008). SEM with missing data and unknown population distributions using two-stage ML: Theory and its application. *Multivariate Behavioral Research*, 43, 621-652.