University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Public Access Theses and Dissertations from the
College of Education and Human Sciences

Education and Human Sciences, College of (CEHS)

12-2017

Latent Interaction Modeling with Planned Missing Data Designs

Jayden Nord

University of Nebraska-Lincoln, jaydennord@gmail.com

Follow this and additional works at: <http://digitalcommons.unl.edu/cehsdiss>

Nord, Jayden, "Latent Interaction Modeling with Planned Missing Data Designs" (2017). *Public Access Theses and Dissertations from the College of Education and Human Sciences*. 300.

<http://digitalcommons.unl.edu/cehsdiss/300>

This Article is brought to you for free and open access by the Education and Human Sciences, College of (CEHS) at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Public Access Theses and Dissertations from the College of Education and Human Sciences by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

LATENT INTERACTION MODELING WITH PLANNED MISSING DATA
DESIGNS

by

Jayden Nord

A THESIS

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Master of Arts

Major: Educational Psychology

Under the supervision of Professor James A. Bovaird

Lincoln, Nebraska

December 2017

LATENT INTERACTION MODELING WITH PLANNED MISSING DATA DESIGNS

Jayden Nord, M.A.

University of Nebraska, 2017

Advisor: James A. Bovaird

Planned missing data (PMD) designs allow researchers to collect additional data under time constraints, which likely occur in educational research settings. The imposed missing data patterns, however, can hamper the efficiency of statistical models that are of interest to substantive researchers. Typically, PMD designs result in a modest power deficiency; however, this tenet has not been extended to latent interaction models. Such models are of increasing importance as researchers investigate moderated relationships involving continuous latent variables. Monte Carlo simulations were used to assess the efficacy of various latent interaction estimation methods under PMD designs.

Table of Contents

List of Figures	iii
List of Tables	iv
Introduction.....	1
Latent Interaction Background	6
Latent variables	6
Latent interactions	6
Product indicator approaches	9
Constrained approaches	9
Unconstrained approaches	10
Distribution analytic approaches	13
LMS	13
Standard error corrections	15
Planned Missingness Background	16
Missing data mechanisms.....	16
Estimation methods to use with missing data	16
FIML.....	17
Planned missing data designs	18
Present Study	22
Methods.....	24
Population model across conditions.....	24
Data generation conditions.....	26
Interaction effect.....	26
Distribution shape.....	27
Completeness.....	29
Sample size	29
Data analysis conditions.....	30
Analysis method	30
Standard error calculation.....	31
Outcomes.....	31
Relative bias of parameters estimates and standard errors	31
Type I error, power, and coverage.....	32

Relative efficiency	33
Confidence intervals for outcomes.....	34
Results.....	36
Convergence.....	36
Parameter bias	36
Standard error bias	39
Type I error and power.....	40
Coverage.....	43
Relative Efficiency	44
Discussion	47
Limitations and future directions	51
Conclusions and implications for applied research.....	53
References.....	55
Appendix A. Example Mplus syntax	63
Appendix B. Tables of results.....	64

List of Figures

<i>Figure 1.1.</i> Parameter bias for the null interaction condition across all other study conditions.	37
<i>Figure 1.2.</i> Relative parameter bias for the non-zero interaction conditions across all other study conditions.	38
<i>Figure 2.</i> Relative standard error bias.....	40
<i>Figure 3.</i> Type I error and power.....	41
<i>Figure 4.</i> Coverage rates for the population interaction value.	44
<i>Figure 5.</i> Relative efficiency.	45

List of Tables

Table 1. <i>Population coefficients for simulation study by distribution and interaction effect size</i>	28
Table 2. <i>Indicator variables for each PMD set partition</i>	29
Table B.1. <i>Parameter bias and relative parameter bias of the interaction effect.</i>	64
Table B.2. <i>Relative standard error bias</i>	66
Table B.3. <i>Type I error and power</i>	68
Table B.4. <i>Relative efficiency</i>	70

Introduction

In cross-sectional educational and psychological research, it is not always possible to collect all desired questionnaire data due to the length of the questionnaire. The questionnaire may be longer than what the participant is willing to complete (Graham, Taylor, Olchowski, & Cumsille, 2006), resulting in incomplete response data (Smits & Vorst, 2007) or the participant withdrawing from the study (Harel, Stratton, & Aseltine, 2015). Alternatively, the length of the questionnaire may result in fatigue or boredom which threaten the validity of responses near the end of the measure (Thorndike & Thorndike-Christ, 2010). There may also be limited time in which data may be collected. For example, in educational research, data collection may occur during school hours; understandably, school administrators can only offer limited time away from classes for data collection (e.g. Bachman, Johnston, O'Malley, & Schulenber, 2011). If the questionnaire is too long, complete or valid data cannot be collected in the allotted time. There are two possible solutions to issues resulting from long measures (Smits & Vorst, 2007).

One solution is to shorten or reduce the questionnaire to a more manageable length. This solution is only feasible if the researcher knows or can investigate the psychometric properties of the measure for the target population (Smits & Vorst, 2007). The length, however, may prohibit higher quality data from being collected that would provide the psychometric properties needed to shorten the measure. The second, alternative solution is to use a multiform planned missing data (PMD) design. In a multiform PMD design, participants are randomly assigned to one of several subsets of

the complete questionnaire (Graham et al., 2006; Rhemtulla & Hancock, 2016). One common multiform design is the *3-form design* in which a questionnaire is partitioned into four sets X, A, B, and C (Graham et al., 2006). All participants receive the X set and are randomly assigned to receive two of the remaining sets A, B, and C and have the third set missing. Thus, individual participants complete one of three possible reduced questionnaires while the researcher gains information on all four item sets across all participants. Modern missing data handling methods such as full information maximum likelihood (FIML) can be used to estimate model parameters in spite of the missing data at some cost to statistical power (Graham et al., 2006). Thus, it is important to understand how planned missingness impacts the efficiency of parameter recovery in various statistical models of interest to substantive researchers.

Some statistical models of interest may use questionnaire items as reflective indicators of latent constructs (e.g. Maslowsky, Jager, & Hemken, 2014), especially in psychological research (Bollen, 2002). The latent constructs may be involved in confirmatory factor analyses (CFAs) or in latent regression models. Rhemtulla, Savalei, and Little (2016) showed that when the 3-form design is applied to CFA and latent regression models, model parameters are recovered with an acceptable loss of efficiency (i.e. greater variability in the estimate). More advanced models, however, have not been investigated. For instance, the effect of PMD designs on latent regression models in which latent predictors interact or moderate each other's relationship with the outcome, hereafter referred to as latent interaction models, has not been explored. The purpose of the present study is to add to PMD research by exploring the effects of PMD designs on

latent interaction models and inform applied researchers of the best study design and analytical practices for latent interaction studies that require PMD designs.

There is a need for exploration of the effect of PMD designs on latent interaction models because there has been growing interest in interaction/moderation models from research programs and funding agencies. For example, funding agencies have recently called for research on potential moderators of relationships between education and social or health factors (Institute of Education Sciences, 2016; National Institute of Health, 2017). In some cases, these moderators are construed as latent continuous variables measured indirectly via observable indicators such as questionnaire items (Hernández et al., 2016; Kwok, Im, Hughes, Wehrly, & West, 2016; Masland & Lease, 2016). If, however, the questionnaire used to indirectly measure the latent variable is too long such that data validity is threatened or there are time constraints imposed by the research setting, a PMD design may be needed to collect data for the latent interaction model. Summative literature on PMD designs has also broadened access to knowledge of PMD designs (Rhemtulla & Hancock, 2016), encouraging their future use in educational research.

PMD designs reduce the amount of available information; yet, interaction effects, on average, have small effect sizes that require a large amount of information in order to be adequately detected (Cohen, Cohen, West, & Aiken, 2003). Although PMD designs result in acceptable efficiency losses for the latent regression model, the efficiency loss is likely to be compounded for an interaction term. Further complicating matters is that there are multiple methods available for estimating latent interaction effects. Prevalent,

modern methods include product indicator approaches and distribution analytic approaches. Product indicator approaches involve multiplying the indicators from one latent predictor with the indicators of another latent predictor (Algina & Moulder, 2001; Jöreskog & Yang, 1996; Kenny & Judd, 1984; Lin, Wen, Marsh, & Lin, 2010; Marsh, Wen, & Hau, 2004). These “product indicators” serve as indicators for another latent term that stands in for the product of the latent variables. Alternatively, distribution analytic approaches such as latent moderated structural equations (LMS) estimate model parameters, including the interaction effect, by analyzing the multivariate density of the observed indicators (Klein & Moosbrugger, 2000; Klein & Muthén, 2007).

The performance of these methods changes as functions of certain data characteristics. For instance, when data are normally distributed, LMS outperforms some product indicators approaches in bias and efficiency (Cham, West, Ma, & Aiken, 2012). When data are non-normally distributed, LMS suffers compared to product indicator approaches in bias and coverage (Cham et al., 2012). Whether or not these different estimation methods also perform differently under planned missingness has not been explored (Cham, Reshetnyak, Rosenfeld, & Breitbart, 2017; Enders, Baraldi, & Cham, 2014).

Latent interaction models are of growing interest to applied researchers (Hernández et al., 2016; Kwok et al., 2016; Masland & Lease, 2016) and the context of the research may necessitate a PMD design (Bachman et al., 2011). Applied researchers need to understand the potential limitations that planned missingness imposes upon latent interaction estimation and understand under what conditions these limitations are

minimized. Such understanding should be based on supportive methodological research. Monte Carlo simulations will be used to assess the impact of several factors on power, parameter estimates, and standard errors of a latent interaction effect as estimated by product indicator and distribution analytic approaches. The following sections provide the background information on latent interactions and PMD designs that will guide the selection of factors for the simulation study.

Latent Interaction Background

Latent variables

Latent variables are those that have no sample realization for at least some cases of a given sample (Bollen, 2002). That is, if sample values are unavailable for some variable of interest, perhaps due to unavailable or limited measurement tools, that variable of interest is considered latent. Put another way, if the location of an individual on the continuum of values of some variable is uncertain, that variable is considered latent (Borsboom, 2008). Because the true value of the variable is not observed, the latent variable is said to be unobserved. Conversely, a variable is an observed or manifest variable if there is a sample realization of its values (Bollen, 2002) or we can place an individual on the continuum of values for that variable with certainty (Borsboom, 2008). Because latent variables are unobserved, researchers are reliant on observed variables to serve as indicators for or indirect measures of latent variables (Kline, 2011). Latent variables and their indicators can be incorporated into statistical models via structural equation modeling (SEM).

Latent interactions

Two continuous variables are said to interact when the value of one variable influences the effect of another variable on some outcome. Interactions can occur between categorical variables, between continuous variables, or between a continuous and categorical variable. The current discussion is focused on interactions between continuous variables. In linear modelling, the interaction effect is represented by the

product of the interacting variables. For example, a simple interaction model is represented by the formula:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + e \quad (1)$$

where Y is some outcome, X and Z are predictors, e is the residual term, and all β are regression coefficients. Using Cohen, Cohen, West, and Aiken's (2003) terminology, variables X and Z are lower-order terms while XZ is the higher-order interaction term. The above formula can be re-expressed in two ways:

$$Y = \beta_0 + (\beta_1 + \beta_3 Z)X + \beta_2 Z + e \quad (2)$$

$$Y = \beta_0 + \beta_1 X + (\beta_2 + \beta_3 X)Z + e \quad (3)$$

As is illustrated by Equation 2, the effect of X on Y , which is $(\beta_1 + \beta_3 Z)$, is conditional on the value of Z . That is, the change in Y for a one unit increase in X depends on what value Z is held. In such a case, variable Z is said to *moderate* the relationship between X and Y . Likewise, X can be said to moderate the relationship between Z and Y as demonstrated by Equation 3. Either interpretation is correct and is largely the decision of the researcher and extant literature which of X or Z is the moderator (Hayes, Glynn, & Hude, 2012). Regardless, the two predictors are interacting and their effects on the outcome are influenced by each other.

Equation 1 generalizes to latent variables. One may substitute any of the lower-order terms with a latent variable to obtain a latent interaction model. The simple latent interaction model was initially described by Kenny and Judd (1984):

$$\eta = \gamma_1 \xi_1 + \gamma_2 \xi_2 + \gamma_3 \xi_1 \xi_2 + \zeta \quad (4)$$

In Equation 4, η is some latent criterion regressed on some latent predictors ξ_1 and ξ_2 , and their interaction term $\xi_1\xi_2$. The residual or disturbance of the model is ζ . Regression coefficients are represented by γ . Hereafter, lower-order predictors will refer to the latent predictors ξ_1 and ξ_2 ; lower-order indicators refer to those indicators for the lower-order predictors. The higher-order term or interaction term refers to $\xi_1\xi_2$.

The estimation of simple latent interaction models is faced with two challenges. The first and perhaps most obvious one is that one cannot simply multiply the lower-order terms to make an interaction term as there are no realized values with which to multiply. The second challenge stems from the distribution of the interaction term. Even if lower-order latent predictors are normally distributed, the interaction term as a product of the lower-order terms is often non-normally distributed (Dimitruk, Schermelleh-Engel, Kelava, & Moosbrugger, 2007). The non-normality of the interaction term also increases as the covariance of the lower-order terms increases. The criterion will also be non-normally distributed as its distribution is determined by the predictors which, in an interaction model, include the non-normally distributed interaction term. The non-normality is problematic for model estimation because maximum likelihood, which is typically used to estimate models involving latent variables, assumes multivariate normality (Kline, 2011). The consequence of the normality violation due to the interaction term is underestimation of the interaction effect's standard errors (Moosbrugger, Schermelleh-Engel, Kelava, & Klein, 2006). Underestimated standard errors may lead to inflation of Type I error rates involving the interaction effect. Thus, the second challenge to estimating latent interaction models is obtaining estimates and

standard errors that are robust to the non-normality caused by the interaction term.

Methods for estimating latent interactions must either be robust to the non-normality while using a method assuming multivariate normality or explicitly account for the non-normality of the interaction term (Moosbrugger et al., 2006).

The desire for methods to estimate latent interaction models involving continuous predictors and the need to overcome such models' methodological challenges has resulted in an ongoing, decades-long series of studies dedicated to the development and refinement of latent interaction estimation procedures. Two widely accepted (Marsh, Wen, Nagengast, & Hau, 2012) approaches to estimating models with latent interactions include *product indicator* (Algina & Moulder, 2001; Jöreskog & Yang, 1996; Kenny & Judd, 1984; Lin et al., 2010; Marsh et al., 2004) and *distribution analytic approaches* (Klein & Moosbrugger, 2000; Klein & Muthén, 2007). Each approach is an umbrella for a number of procedures.

Product indicator approaches

Given two latent predictors, one cannot simply multiply the predictors to make an interaction term as there are no realized values with which to multiply. One approach to modeling the latent interaction term is to instead make products from the lower-order indicators and use those products as indicators for another latent construct representing the interaction term. Such product indicator (PI) approaches stem from the work of Kenny and Judd (1984).

Constrained approaches. Kenny and Judd (1984) proposed a latent interaction estimation method in which each of the indicators for one lower-order term are multiplied

with each of the indicators of the other lower-order term; all possible product combinations are generated. Numerous nonlinear constraints are imposed on the variances and loadings of the model. The Kenny-Judd method also omits the mean structure of all involved variables as all observed variables were mean-centered. However, Jöreskog and Yang (1996) showed that it is inappropriate to omit the mean structure from the model because, even when the normally distributed lower-order latent terms are centered at zero, the interaction term has a mean equivalent to the covariance of the lower-order terms. Thus, Jöreskog and Yang included the latent and observed mean structure in their model. In addition to including the mean structure, Jöreskog and Yang found that the model could be identified with one product indicator rather than all possible product combinations. Algina and Moulder (2001) further revised Jöreskog and Yang's (1996) method by centering the lower-order indicators prior to product indicator construction; centering resulted in better convergence rates for the interaction model than the Jöreskog-Yang model.

Unconstrained approaches. Marsh, Wen, and Hau (2004) investigated the consequences of omitting the nonlinear constraints in the latent interaction model. If the omission had few negative consequences, then the potentially complicated constraints would not need to be specified. Marsh and colleagues (2004) also note that the nonlinear constraints are dependent on normally distributed lower-order terms. If the latent predictors were non-normally distributed, it would be inappropriate to invoke the nonlinear constraints. Thus, it is possible that an unconstrained approach would be more appropriate for models in which the lower-order terms violate multivariate normality.

In order to evaluate the unconstrained approach, Marsh and colleagues (2004) mean-centered all indicators as done by Algina and Moulder and made product indicators from the centered set. Unlike Algina and Moulder, the unconstrained approach omitted all nonlinear constraints. With normally distributed lower-order indicators, Marsh and colleagues (2004) found that the unconstrained approach performed as well as the constrained approach. With non-normally distributed lower-order indicators, the unconstrained approach yielded unbiased parameter estimates of the interaction effect. Meanwhile, the constrained approach yielded biased estimates of the interaction effect as well as poorer goodness-of-fit measures. Not only does it seem acceptable to omit the nonlinear constraints, it is also appropriate to omit the constraints when the lower-order indicators are non-normally distributed.

Additionally, Marsh and colleagues (2004) examined the impact of different strategies used to construct product indicators. For example, Kenny and Judd (1984) used all possible combinations. Conversely, Jöreskog and Yang (1996) showed that a constrained model could be identified with only one product indicator. Yang (1998) would later use a matched-pairs strategy in which the first indicator from one latent predictor is multiplied with the first indicator from the other latent predictor, the second with the second, and so on to make matched product indicators. Marsh and colleagues (2004) argued that Kenny and Judd's approach unnecessarily reused information to estimate the latent interaction effect and also argued that Jöreskog and Yang's approach used too little information. Marsh and colleagues speculated that the matched-pairs strategy is most optimal because it used all lower-order indicators without redundancy.

Marsh and colleagues (2004) evaluated the three strategies (single-pair, matched-pairs, all-pairs) with the unconstrained model. It should be noted that the single-pair strategy, which uses only one product indicator, is only possible by constraining the indicator's factor loading (e.g. including the nonlinear constraints). The unconstrained approach would otherwise require at least two product indicators to be identified (Marsh et al., 2004). The single-pair strategy tended to result in more biased interaction effect estimates with larger standard errors than the matched- or all-pairs strategies. Small differences between the matched- and all-pairs strategies were found in the average observed standard deviation and the average estimated standard error of the interaction effect. The authors concluded that the matched-pairs strategy was preferable due to its simplicity over the all-pairs strategy.

A mean structure is still required in Marsh and colleagues' (2004) single mean-centering approach. The use of residual-centering, however, eliminates the need to specify a mean structure (Marsh et al., 2006; Little, Bovaird, & Widaman, 2006). In residual centering, the product indicators are regressed onto the lower-order indicators and the resulting residuals are used as indicators for the product term. The impetus of residual centering is to remove non-essential multicollinearity between the lower-order and interaction terms, thus improving power to detect interaction effects (Little et al., 2006). A criticism of the residual centering approach, however, is that the residual centering approach alters the lower-order effects under non-normality (Lin et al., 2010).

In order to fully remove the mean structure, Lin and colleagues (2010) proposed a double-mean-centering (DMC) procedure. DMC first involves the same steps as the

unconstrained mean-centering approach described by Marsh and colleagues (2004). Next, the product indicators are mean-centered. The centered product indicators are used as indicators for the latent interaction term and the model is estimated as usual. This procedure is essentially equivalent to residual centering under normality while not altering the lower-order effects under non-normality. Thus, the mean structure does not need to be specified. Because of the simplicity afforded by omitting the mean structure, the DMC is considered the current optimal product indicator approach (Lin et al.).

Distribution analytic approaches

Unlike product indicator approaches, distribution analytic approaches such as Latent Moderated Structural Equations (LMS; Klein & Moosbrugger, 2000) and Quasi-Maximum Likelihood (QML; Klein & Muthén, 2007) estimate latent interaction models without product indicators. Instead, distribution analytic approaches take the non-normality of the interaction term into account and approximate a multivariate non-normal density function of the model's observed variables (Moosbrugger et al., 2006). LMS and QML differ in how the non-normal density function is approximated. QML will not be discussed further as there is currently no readily available software that can implement QML with missing data.

LMS. In a simple interaction model where the latent predictors are correlated and normally distributed, the interaction term and criterion are non-normally distributed. As a result, the lower-order indicators are normally distributed while the criterion's indicators are non-normally distributed. The multivariate distribution of all observed variables is non-normally distributed due to the non-normality of the criterion and its indicators.

However, if the observed variables are conditioned on specific values of the latent predictors, the multivariate distribution of the observed variables will be normally distributed; such distributions are conditionally normal. Latent moderated structural equations (LMS) is an estimation method that explicitly accounts for the non-normality of the multivariate observed distribution by treating it as a finite mixture of conditionally normal distributions (Klein & Moosbrugger, 2000).

The software Mplus (Muthén & Muthén, 2012) implements LMS via random slopes and numerical integration. A random slopes representation of the simple interaction model is as follows:

$$\begin{aligned}\eta &= \gamma_1\xi_1 + \gamma_2\xi_2 + s\xi_1 + \zeta \\ s &= 0 + \gamma_3\xi_2 + 0\end{aligned}\tag{5}$$

where s has no intercept or variance. Numerical integration then “approximates [the] integration [of the multivariate density] by using a weighted sum over a set of integration points (quadrature nodes) representing values of the latent variable” (Muthén & Muthén, 2012, p. 471). The numerical integration procedure reflects Klein and Moosbrugger’s (2000) original description of weighted sums of conditionally normal distributions. Expectation-maximization is then used to estimate model parameters based on the approximated density.

LMS rests on the assumption that the latent predictors are normally distributed (Klein & Moosbrugger, 2000). When the lower-order indicators are normally distributed, LMS parameter estimates tend to be unbiased compared to PI approaches (Cham et al., 2012; Klein & Moosbrugger, 2000). However, when the lower-order terms are non-normally distributed, coverage rates for the interaction effect worsen with increasing

sample size. When data are skewed and leptokurtic, unconstrained PI approaches outperform LMS in terms of coverage rates (Cham et al., 2012). Put another way, statistical power of LMS drops relative to PI approaches when used with non-normal data. It appears then, that LMS likely outperforms PI approaches such as DMC with normal data; however, the relative performance between the two methods switches in the presence of non-normal data and the performance difference also becomes dependent on sample size.

Standard error corrections

Under the PI approaches, the use of maximum likelihood to estimate a latent interaction model results in underestimated standard errors of the interaction effect due to the non-normality caused by the interaction term (Moosbrugger et al., 2006). Satorra-Bentler standard errors, typically used in cases of non-normality, offer little noticeable improvement with regard to bias when lower-order terms are normal *or* non-normal (Cham et al., 2012; Marsh et al., 2004). Huber-White standard errors, however, have not been systematically examined when used in latent interaction estimation (Cham et al., 2017, 2012; Enders et al., 2014; Lin et al., 2010). In Mplus, normal theory standard errors are the default for PI approaches whereas Huber-White standard errors are the default for the LMS approach (Muthén & Muthén, 2012). To avoid the confounding influence of default standard error estimators and to investigate the potential utility of Huber-White standard errors under PI approaches, standard error estimator should be another factor in comparison studies of LMS and PI approaches such as DMC.

Planned Missingness Background

Missing data mechanisms

Missing data mechanisms describe the cause of missing values observed in data. The classification of different mechanisms is helpful because different treatments of missingness, planned or unplanned, work better depending on the underlying cause of missingness (Enders & Bandalos, 2001). Originally described by Rubin (1976) and clarified by Enders (2010), missing data mechanisms are divided into three categories: missing not at random (MNAR), missing at random (MAR), and missing completely at random (MCAR). Data in some variable are MNAR when the probability of being missing is dependent on the values of the missing data for that variable. Meanwhile, data in some variable are MAR when the probability of being missing is not dependent on the values of the missing data but instead dependent on another or several other variables. Finally, data in some variable are MCAR when the probability of being missing is unaccounted for by any other variable.

Estimation methods to use with missing data

Traditional missing data handling techniques have involved either deletion of entire cases with missing values or single imputation of missing values; however, deletion methods such as listwise and pairwise deletion require MCAR data in order to be efficient and unbiased (Enders & Bandalos, 2001). That is, if data are MAR or MNAR, listwise or pairwise deletion are inefficient procedures. Meanwhile, single imputation methods are often biased regardless of the missing data mechanism and are generally not recommended (Enders, 2010). The traditional missing data handling methods have been

surpassed by modern methods such as full-information maximum likelihood (FIML; Arbuckle, 1996) and multiple imputation (MI; Rubin, 1987). Whereas deletion methods rely on the strict assumption of the MCAR mechanism (B. O. Muthén, Kaplan, & Hollis, 1987) to be effective, FIML and MI remain effective when data are MAR, a mechanism that's assumed existence is less restrictive (Enders, 2001a). In most cases, FIML and MI yield so similar results that the choice of procedure is trivial (Enders et al., 2014; Graham, Olchowski, & Gilreath, 2007). Because of this and the ease of implementing FIML in structural equation modeling software, only FIML is discussed here.

FIML. Originally conceived by Finkelstein (1979), full-information maximum likelihood (FIML) extends maximum likelihood estimation to account for missing data (Arbuckle, 1996; Enders, 2001a). A likelihood function is used for each i -th participant or case in a given dataset:

$$\log L_i = K_i - \frac{1}{2} \log |\Sigma_i| - \frac{1}{2} (x_i - \mu_i)' \Sigma_i^{-1} (x_i - \mu_i) \quad (6)$$

The case-wise likelihoods are added together for the whole sample likelihood:

$$\log L(\mu, \Sigma) = \sum_{i=1}^N \log L_i \quad (7)$$

In the above functions, x_i is the vector of complete data for case i , μ_i is the vector of estimated means corresponding to variables from x_i , Σ_i is the estimated covariance matrix corresponding to variables from x_i , and K_i is a constant that depends on the number of complete variables for case i . Vector μ and matrix Σ can be expressed as functions of a parameter vector γ . As a result, FIML can be applied to multiple contexts such as SEM by maximizing $\log L(\mu(\gamma), \Sigma(\gamma))$ (Arbuckle, 1996).

The performance of FIML to estimate SEM models with incomplete data has been examined (Enders & Bandalos, 2001; Enders, 2001b). Under multivariate normal conditions, which FIML assumes (Arbuckle, 1996), FIML yielded unbiased and efficient parameter estimates compared to listwise and pairwise deletion methods, even at missingness rates of 25% (Enders & Bandalos, 2001). Model rejection rates were also found to be acceptable at such high rates of missingness. However, with non-normal data under MCAR or MAR mechanisms, standard errors tended to be underestimated and model rejection rates tended to be inflated (Enders, 2001b). As mentioned previously, Satorra-Bentler corrected standard errors for non-normal data cannot be used with incomplete data (Muthén & Muthén, 2012). Robust standard errors, which are compatible with incomplete data, were not examined in the above examinations of FIML's performance (Enders & Bandalos, 2001; Enders, 2001b).

Planned missing data designs

Modern missing data handling techniques such as FIML allow researchers to use planned missing data (PMD) designs. In PMD designs, researchers deliberately and randomly assign patterns of missingness to participants; the bulk of the missing data becomes MCAR by design. The first advantage of PMD designs is that the researcher does not have to make assumptions regarding the missing data mechanisms because they have directly controlled the mechanism. Another advantage of PMD designs is that they allow researchers to collect more information than participants would otherwise be able to provide (Graham et al., 2006).

Various PMD designs exist; one PMD design most relevant to cross-sectional research is the multiform design (Rhemtulla & Little, 2012). In the multiform design, a questionnaire or measure is partitioned into several sets. All participants receive one of the sets which is typically referred to as the X set. Then, participants are randomly assigned to receive some number of the remaining sets. The possible combinations of sets result in several forms, the number of which can be expressed with the following combination formula assuming order does not matter:

$$\frac{n!}{r!(n-r)!} \quad (8)$$

where n is the number of all sets minus one and r is the number of partitions to give minus one. The overall formula is read as “ n choose r ”. One is subtracted from the number of partitions because all participants must receive at least one of the same partitions, the X set. Thus, n is the number of *remaining* partitions that *can* be assigned and r is the number of *remaining* partitions *to* assign. For example, consider that participants are only willing to respond to three items sets but a researcher wants to collect information from four partitions denoted X, A, B, and C. All participants receive one of the partitions, X, so n is set to 3 and r is set to 2. Plugging $n = 3$ and $r = 2$ into Equation 8 results in 3 possible forms. Participants are randomly assigned to receive two of A, B, or C. Thus, the participant may receive XAB, XAC, or XBC.

Simulation studies have shown that when FIML is applied to data using the 3-form design, covariance estimates differed only in the third decimal place. Average root mean square residuals and standard errors across simulated datasets were also acceptable (.028 and .025 respectively; Graham, Hofer, & MacKinnon, 1996). More advanced

models, however, are of interest to substantive researchers, especially psychologists that often use latent variables (Bollen, 2002).

Rhemtulla and colleagues (2016) investigated the effects of a 3-form PMD design on parameter efficiency (empirical variability) in CFA and latent regression models relative to the efficiency of models based on complete data. The CFA model consisted of six covarying latent variables each with four indicators. Each indicator under a given latent variable belonged to separate sets X, A, B, and C. The authors found that latent variances and covariances had relative efficiencies of around .90 under a PMD design. Factor loadings, however, had lower relative efficiencies though their efficiencies were never lower than .60. The authors then explored a latent regression model in which two latent variables together predicted four latent outcomes. Again, each latent variable had four indicators each belonging to one of sets X, A, B, and C. Relative efficiency was high for the latent regression coefficients and disturbance variances and covariances.

Sample size was not investigated in Rhemtulla et al. (2016). Of course, the imposition of planned missingness reduces power to detect effects; there is a potential trade between imposed missingness and the need for a larger sample size. Jia and colleagues (2014) investigated the effect of sample size on cross-sectional confirmatory factor analysis, two-time-point CFA, and three-time-point mediation models using a 3-form PMD design. Performance in terms of convergence, parameter bias, and standard error bias differed depending on the type of model and its inherent complexity. The minimum acceptable sample size for the most complex model, the three-time-point mediation model, was 130 when FIML estimation was used. Rhemtulla and Little (2012),

however, generally recommend 375 cases for most structural equation models using 3-form designs to ensure adequate covariance coverage.

The impact of using any multiform PMD design on latent interaction studies is unknown (Cham et al., 2017; Enders et al., 2014). Although the power and efficiency loss is often minimal (Graham et al., 2006; Rhemtulla et al., 2016), interaction effects have small effect sizes that can require a substantial amount of data to maintain adequate power (Cohen et al., 2003). Introducing planned missingness into a latent interaction study may inadvertently remove the ability to detect the interaction effect of interest. The lack of research to affirm or reject this claim is the primary motivation of the present study.

Present Study

The main purpose of the present study is to determine the effectiveness of different latent interaction estimation methods on data with planned missingness. The following research questions are posed regarding the estimation of the latent interaction effect:

1. How does latent interaction estimation perform between complete and PMD design conditions?
2. Given PMD, how does the performance of LMS and DMC compare across conditions of...
 - a. Distribution shape?
 - b. Sample size?
3. How do normal-theory standard error calculation and robust standard error calculation compare across conditions of...
 - a. Estimation method (LMS or DMC)?
 - b. Distribution shape?
 - c. Sample size?

Based on the above research questions and review of the literature, the following hypotheses are proposed regarding the estimation of the latent interaction effect:

1. Models with PMD will perform worse than models with complete data; however, the differences should be negligible (Graham et al., 1996, 2006).
2. Under PMD conditions...

- a. LMS will outperform DMC when data are normal. However, DMC will outperform LMS when data are non-normal (Cham et al., 2012; Klein & Moosbrugger, 2000).
 - b. Moderate sample sizes of at least 375 will be needed for adequate model estimation per Rhemtulla and Little's (2012) recommendation.
3. Robust standard errors will result in unbiased standard errors across all conditions (Muthén & Muthén, 2012).

Methods

Population model across conditions

Data were generated from a simple latent interaction model used in several latent interaction studies (Jöreskog & Yang, 1996; Kenny & Judd, 1984; Lin et al., 2010; Marsh et al., 2004). The structural model is:

$$\eta = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \zeta \quad (4)$$

where η , ξ_1 , and ξ_2 are latent variables each with a variance of 1 and ζ is the residual disturbance of the model. Each latent variable has four indicators with intercepts of 8 so as to mimic situations in which mean-centering would be advised. Each latent variable also accounts for 70% of the variance in each of the four indicators. The following formula for the reliability of an indicator can be manipulated to find appropriate loadings

$$R_x^2 = \frac{\lambda^2 Var(\xi)}{\lambda^2 Var(\xi) + Var(e)} \quad (9)$$

where R_x^2 is the reliability of some indicator x , λ is the loading for given indicator, ξ is the latent variable associated with the given indicator, e is the measurement error associated with the indicator, and $Var(.)$ is the variance function for the object within the parenthesis. The numerator is the variance accounted for by the latent variable while the denominator is the total variance of the indicator. Solving for λ , the above formula becomes:

$$\lambda = \sqrt{\frac{R_x^2 Var(e)}{Var(\xi)(1-R_x^2)}} \quad (10)$$

If x is standardized, then $Var(e) = 1 - R_x^2 = 1 - .7 = .3$. Thus, the loading for a standardized indicator is $\sqrt{.7} \approx .837$. Meanwhile, the first indicators were desired to

have a loading of 1 while the associated latent variable still accounts for 70% of the first indicator's variance. Thus, the equality $Var(e) = 1 - R_x^2$ no longer holds and $Var(e)$ must be found. The above formula can be rewritten to find the variance of the measurement error for each first indicator which is $3/7 \approx .429$.

Thus, the loading and measurement error variance for the first indicator of each latent variable is 1 and .429 respectively. Meanwhile, the loading and measurement error variance for the remaining indicators of each latent variable is .837 and .300 respectively. These loadings and measurement errors give reliabilities of .7 for all indicators.

The covariance between the lower-order latent predictors was set such that correlations between any indicator from one lower-order predictor and an indicator from the other lower-order predictor were equal to .3, a moderate correlation using Cohen's (1988) terminology. With standardized latent variables and indicators, the covariance/correlation between the latent predictors is found with the following formula:

$$Cor(x, z) = \lambda_x Cor(\xi_1, \xi_2) \lambda_z \quad (11)$$

where $Cor(x, z)$ is the correlation between indicators x and z which are respectively indicators of ξ_1 and ξ_2 , and λ_x and λ_z are the loadings of x and z respectively. Thus, the correlation between the latent predictors was set to $3/7 \approx .429$ to maintain a correlation of .3 among the lower-order indicators of different latent variables.

The exact coefficients of the structural model differ across distribution shape conditions in order to maintain the same effect sizes across conditions. The lower-order coefficients were arbitrarily set such that the lower-order predictors account for 30% of

the variance of the outcome. The interaction effect size varies as a function of the interaction effect condition described below.

Data generation conditions

Using the above described population model, data generation varied over the following factors: interaction effect, distribution shape, number of indicators, completeness, and sample size. There are three levels for interaction effect (null, medium, very large), two levels for distribution shape (normal and non-normal), two levels for data completeness (complete and PMD), and three levels for sample size (200, 375, 550). Fully crossing the factor levels results in $3 * 2 * 2 * 3 = 36$ data generation conditions. R (R Core Team, 2016) was used to generate 10,000 datasets or as many datasets needed to obtain 10,000 successful convergences per study condition. During data generation, mean-centered indicators and product indicators were also calculated and saved with the datasets to be used in the analysis stage of the simulation. The generation factors are discussed in further detail below.

Interaction effect. The first design factor was interaction effect size expressed as the ΔR^2 resulting from the inclusion of the interaction term. Interaction effect size was manipulated to be null ($\Delta R^2 = 0$), medium ($\Delta R^2 = .035$), and very large ($\Delta R^2 = .100$). The null condition was chosen to assess Type I error rates. The medium condition is a compromise between what are considered to be a small (.01) and large (.05) interaction effect size (Enders et al., 2014). The very large interaction effect size was chosen arbitrarily so potential trends involving an increasing interaction effect could be examined.

Distribution shape. The second design factor was the distribution of the data which was manipulated to be normal or non-normal. In the non-normal condition, skew and excess kurtosis values of 2 and 7 were imposed upon the latent predictors via decomposition with intermediate correlations (Vale & Maurelli, 1983) and Fleishman's transformation procedure (Fleishman, 1978). These skew and kurtosis values have been described as moderately non-normal (Curran, West, & Finch, 1996) and are similar to those values used in similar research (Enders et al., 2014; Marsh et al., 2004).

The structural model coefficients differ between conditions in order to maintain desired effect sizes. Under multivariate normality, the following expression is used to relate the lower-order terms to the variance accounted for by the lower-order terms:

$$R^2 = \gamma_1^2 Var(\xi_1) + \gamma_2^2 Var(\xi_2) + 2\gamma_1\gamma_2 Cov(\xi_1\xi_2) \quad (12)$$

where R^2 is the variance accounted for by the lower-order terms. Holding $\gamma_1 = \gamma_2$ and desiring $R^2 = .300$, the above formula yields $\gamma_1 = \gamma_2 \approx .324$. Meanwhile, the following expression is used to relate the interaction term to the additional variance accounted for by the interaction effect:

$$\Delta R^2 = \gamma_3^2 [Var(\xi_1) * Var(\xi_2) + Cov(\xi_1\xi_2)^2] \quad (13)$$

where ΔR^2 is the additional variance accounted for by the interaction effect. Table 1, at the end of this section summarizes the different calculated interaction coefficients for the normal condition.

Under non-normality, the above expressions cannot be used to derive coefficients to maintain the desired effect sizes. This is because the interaction term is no longer statistically independent from the lower-order terms and its variance and covariances

with the lower-order terms are complex functions of skewness (Enders et al., 2014). Lower-order and interaction coefficients were iterated across different values and combinations. Within each iteration, 500 datasets of $N = 1,000,000$ were generated. The lower-order terms had skew and excess kurtosis of 2 and 7, respectively, so as to reflect the non-normal condition. The criterion y was formed as a linear combination of the interaction of the generated variables x and z given the iterated coefficients. The criterion y was then regressed on x and z without an interaction term. ΔR^2 values were obtained by taking the difference in the total variance of y and the variance accounted for in the regression (here referred to as R^2) of y on x and z .

An interaction coefficient with a desired ΔR^2 can then be obtained by using the regression of ΔR^2 values on the collection of iterated interaction coefficients. That is, the resulting regression formula is solved for the interaction coefficient given a desired ΔR^2 . The R^2 values or variances accounted for by the lower-order terms, meanwhile, appeared as a function of both lower-order and interaction coefficients. Thus, lower-order coefficients with a desired R^2 were obtained by using the regression of R^2 values regressed on the collection of iterated lower-order coefficients *and* interaction coefficients. That is, the resulting regression formula was solved for the lower-order coefficient given a desired R^2 and interaction coefficient that was obtained from the previous regression solution. Table 1, below, summarizes the coefficients for the normal and non-normal conditions.

Table 1. *Population coefficients for simulation study by distribution and interaction effect size*

ΔR^2	$\gamma_1 = \gamma_2$	γ_3
Normal		
0	.324	0
.035	.324	.172
.100	.324	.291
Non-normal		
0	.324	0
.035	.262	.132
.100	.219	.223

Completeness. The third design factor was completeness, which varies between complete and PMD. Under the PMD condition, observed variables are partitioned such that each partition has one indicator from each latent variable as recommended by the extant literature (Graham et al., 2006; Rhemtulla et al., 2016). Table 2 summarizes the grouping of indicators into partitions. Under the PMD condition, all cases have complete data for indicators associated with partition X. Then, each case is randomly assigned to have complete data in two of the remaining and available partitions.

Table 2. *Indicator variables for each PMD set partition*

Partition	Variable
X	x_1, z_1, y_1
A	x_2, z_2, y_2
B	x_3, z_3, y_3
C	x_4, z_4, y_4

Sample size. The fourth design factor was sample size which varied among 200, 375, and 550. The first sample size of 200 was chosen as it is often recommended as the minimum sample size needed for structural equation models (Kline, 2011). The second sample size of 375 was chosen as it is recommended by Rhemtulla and Little (2012) as the minimum sample size needed for a structural equation model using the three-form PMD. The third

sample size of 550 was chosen arbitrarily so potential trends in sample size across other conditions could be examined.

Data analysis conditions

Each of the datasets from the 16 generation conditions were analyzed over the following factors: analysis method and standard error calculation. There are two levels for analysis method (DMC or LMS) and two levels for standard error calculation (normal-theory and robust). Fully crossing the factor levels results in $2 * 2 = 4$ data analysis conditions. Crossing the data generation conditions and the data analysis conditions results in $36 * 4 = 144$ study conditions. All datasets were analyzed with Mplus version 7.0 (Muthén & Muthén, 2012) and example syntax for both methods can be found in Appendix A. The analysis factors are discussed in further detail below.

Analysis method. The first data analysis factor is analysis method which was manipulated to either be DMC or LMS. For DMC (Lin et al., 2010), mean-centered lower-order indicators were used to compute matched-pairs product indicators that were also mean-centered. Mean-centered indicators were reflective of the latent predictors and criterion while the mean-centered product indicators were reflective of a latent predictor representing the interaction term. The second analysis method was LMS (Klein & Moosbrugger, 2000). Raw, un-centered indicators were reflective of the latent predictors and criterion. The interaction term was specified using the XWITH operator in Mplus. The numerical integration required by LMS is then invoked with the TYPE=RANDOM and ALGORITHM=INTEGRATION options under the ANALYSIS command.

Standard error calculation. The second data analysis factor is standard error calculation which was manipulated to either be normal-theory or robust. In Mplus, under the ANALYSIS command, normal-theory standard errors are requested via the ESTIMATOR=ML option while robust standard errors are requested via the ESTIMATOR=MLR option (Muthén & Muthén, 2012).

Outcomes

Relative bias of parameters estimates and standard errors. Bias is a metric with which to evaluate the accuracy of model estimates and standard errors. Biases are calculated for each of the 10,000 replications per 144 study conditions. Raw parameter bias for a replication is the difference between that replication's parameter estimate and the population parameter within the replication's condition. Relative parameter bias is the raw parameter bias divided by the population parameter for the given condition. In formulaic terms,

$$Bias_{\text{relative}}(\hat{\theta}_{ij}) = \frac{\hat{\theta}_{ij} - \theta_j}{\theta_j} \quad (14)$$

where subscript j refers to one of the 144 study conditions, subscript i refers to one of the 10,000 replications within condition j , and θ is the parameter of interest. θ_j is the population parameter in condition j while $\hat{\theta}_{ij}$ is the estimated population parameter for a replication. The average of $Bias_{\text{relative}}(\hat{\theta}_{ij})$ within condition j provides the relative parameter bias for that condition. Relative parameter bias cannot be calculated for the null interaction condition as the calculation would require dividing by zero. Thus, only raw parameter bias will be examined for the null condition.

Raw standard error bias is the difference between a replication's standard error of a parameter estimate and the empirical standard deviation of parameter estimates in a given condition. Relative standard error bias is the raw standard error bias divided by the empirical standard deviation of parameter estimates in the same condition. In formulaic terms,

$$Bias_{\text{relative}}[SE(\hat{\theta}_{ij})] = \frac{SE(\hat{\theta}_{ij}) - SD(\hat{\theta}_j)}{SD(\hat{\theta}_j)} \quad (15)$$

where i, j , and θ are equivalent to terms in Equation 14. $SE(\hat{\theta}_{ij})$ refers to the estimated standard error of a replication whereas $SD(\hat{\theta}_j)$ refers to the empirical standard deviation of parameter estimates across all replications in the same condition. Parameters and standard errors were considered biased if they exceeded plus or minus .05 (Hoogland & Boomsma, 1998).

Type I error, power, and coverage. Type I error is the rejection of the null hypothesis when the null hypothesis is true. Type II error is the retention of the null hypothesis when the null hypothesis is false. Power, or statistical power, is one minus the probability of a Type II error. Power can be conceptualized as the probability of rejecting the null hypothesis when the null hypothesis is indeed false. In the present study, the null hypothesis of interest is that the latent interaction coefficient is zero indicating that there is no interaction effect. Hypothesis rejection decisions regarding coefficients in SEM are made with a normally distributed Z statistic which is computed as

$$Z_{\theta} = \frac{\hat{\theta}}{SE_{\hat{\theta}}} \quad (16)$$

where θ is the parameter of interest and $\hat{\theta}$ is the predicted parameter value (Muthén & Muthén, 2012). In the present study, the parameter of interest is the latent interaction coefficient. Probability values (p -values) can be calculated for the latent interaction coefficient in each analysis condition on each of the study replications. If the p -value falls below a pre-specified cutoff, the null hypothesis is rejected. For the present study, a p -value cutoff of .05 was chosen. The probability of a Type I error is calculated as the percentage of null hypothesis rejections in the condition with no interaction effect. Power is calculated as the percentage of null hypothesis rejections in the condition with an interaction effect. Coverage is the percentage of interaction effect confidence intervals that contain the true population value.

Relative efficiency. In planned missingness literature, relative efficiency (RE) is defined as the ratio of squared empirical standard errors (standard deviation) for some θ_j in a PMD condition to the squared empirical standard error of the same θ_j in a corresponding condition with complete data (Rhemtulla, Jia, Wu, & Little, 2014; Rhemtulla et al., 2016). In formulaic terms, RE is defined as:

$$RE(\hat{\theta}_j) = \frac{SD^2(\hat{\theta}_{j,complete})}{SD^2(\hat{\theta}_{j,PMD})} \quad (17)$$

where $SD^2(\hat{\theta}_j)$ refers to the squared empirical standard error of $\hat{\theta}$ for some condition j exclusive of completeness. Possible values of RE range between zero and one where zero indicates poor efficiency in the PMD condition compared to the complete condition (Rhemtulla et al., 2016). RE can be used to determine the complete data sample size that yields the same power from an incomplete dataset of some sample size (Rhemtulla et al., 2014, 2016). In formulaic terms:

$$N_{\text{complete}}^* = RE \times N_{\text{PMD}} \quad (18)$$

where N_{PMD} is the sample size of the incomplete dataset and N_{complete}^* is the calculated sample size of the complete dataset that will match the incomplete dataset in terms of power for the parameter estimate on which RE is based. For example, if there is a PMD dataset with 500 cases and RE of .80, then a complete dataset that matches the PMD dataset in terms of power would only need 400 cases.

Conversely, RE can be used to determine the PMD data sample size that yields the same power from a complete dataset of some sample size. In formulaic terms:

$$N_{\text{PMD}}^* = \frac{1}{RE} N_{\text{complete}} \quad (19)$$

where N_{PMD}^* is the calculated sample size of the PMD dataset that will match the complete dataset with sample size N_{complete} in terms of power. For example, if there is a complete dataset with 500 cases and RE of .80, then a PMD dataset would need 25% more cases to match the power of the complete dataset. That is, the PMD dataset would need 625 cases to match the complete 500 cases in terms of power.

Confidence intervals for outcomes

To account for sampling error due to the finite number of replications, 95% confidence intervals were constructed around each study condition's outcome. Normal theory confidence intervals were used for parameter and relative parameter bias and calculated with:

$$Bias(\hat{\theta}_j) \pm 1.96 \times \frac{SD[Bias(\hat{\theta}_j)]}{\sqrt{10000}} \quad (20)$$

Because Type I error, power, and coverage are bounded between 0 and 1, a normal approximation for binomial confidence intervals was used:

$$p_j \pm 1.96 \times \sqrt{\frac{p_j(1-p_j)}{10000}} \quad (21)$$

where p is the average outcome (type I error, power, or coverage) for some condition j .

Finally, percentile-based bootstrap confidence intervals were used for relative SE bias and RE. Bootstrap confidence intervals were chosen to account for Monte Carlo sampling error in cases that the computation is unknown or too complex to be done practically. SE bias received bootstrap confidence intervals to account for the additional uncertainty of $SD(\hat{\theta}_j)$. Meanwhile, RE does not lend itself to a clear calculation of standard errors to be used in confidence interval construction. One thousand bootstrap samples were used to calculate confidence intervals for relative SE bias and RE.

Results

Results are presented below for each of the study outcomes along with figures. For completeness, numerical results are tabulated in Appendix B.

Convergence

Before individual outcomes were assessed, convergence rates were checked for each study condition. Any given replication was considered to have successfully converged if it produced no warning or error messages. Convergence rates were calculated as the number of successful replications out of 10,000. Across all conditions, convergence rates ranged between 98.8% and 100%. Any dataset that resulted in a convergence failure in any of the four analysis conditions was replaced until there were 10,000 successful convergences within all conditions. Assessment of individual outcomes were performed using only these successfully converged replications.

Parameter bias

For parameter bias, standard error estimation method was ignored as there were no differences in parameter estimation between standard error conditions. Results for parameter bias are presented in Figures 1.1 and 1.2. Under normality, all conditions yielded unbiased parameter estimates. Under DMC, sample sizes of 200 yielded parameter estimates that were slightly more biased than larger sample sizes. Under the DMC method and for the $\Delta R^2 = .035$ condition, sample sizes of 375 were more biased under PMD designs than under the complete data design. Despite these nuanced differences, parameter estimates were unbiased under multivariate normality.

Figure 1.1. Parameter bias for the null interaction condition across all other study conditions. Note that the scaling should not be directly compared to Figure 1.2 because Figure 1.2 is specifically relative parameter bias. The vertical line provides a marker for zero bias.

Under non-normality, the LMS method produced positively biased parameter estimates. The positive bias did, however, improve from the $\Delta R^2 = .035$ to the $\Delta R^2 = .100$ condition. Bias in the LMS-derived parameter estimates was worse for the PMD design than for the complete data design. Meanwhile, the DMC method yielded unbiased

Figure 1.2. Relative parameter bias for the non-zero interaction conditions across all other study conditions.

The vertical lines within each panel provide a reference for zero bias.

parameter estimates under non-normality. Under the DMC method and for the two interaction conditions, sample sizes of 375 were more biased under PMD designs than under the complete data design.

Standard error bias

Results for standard error bias are presented in Figure 2. Under normality, standard errors were, in general, unbiased across all conditions except for sample sizes of 200. Furthermore, standard error bias improved with increasing sample size. The LMS method with ML standard error estimation tended to yield the least biased standard errors across all conditions. DMC-ML, however, produced negatively biased standard errors under the interaction conditions wherein the bias worsened from $\Delta R^2 = .035$ to $\Delta R^2 = .100$. Under normality, there were no noticeable differences between the complete data and PMD design conditions.

Under non-normality, LMS-ML was the only analysis method that tended to produce unbiased standard errors but only for sample sizes of 375 and 550. DMC-ML only produced unbiased standard errors for the complete design, null condition with a sample size of 500. Under the 200 case complete null condition, DMC produced negatively biased standard errors with exceptionally wide confidence intervals. Meanwhile, the negativity and confidence interval for the 200-case null condition is less pronounced under the PMD design. A similar pattern, albeit less exceptional, is found in the $\Delta R^2 = .035$ condition. Within the 200-case DMC condition, relative standard error bias has slightly wider confidence intervals under the complete design than the PMD design. At the largest interaction effect size, differences in standard error bias confidence intervals are not noticeable between the complete and PMD designs. Beyond the 200-

Figure 2. Relative standard error bias. The vertical line provides a reference for zero bias.

case DMC conditions, there were no noticeable differences between complete and PMD design conditions in terms of standard error bias.

Type I error and power

Results for Type I error and power are presented in Figure 3. Because of the positive relationship between Type I error and power, it was important to examine Type I

Figure 3. Type I error and power. All columns represent rejection rates of a nil interaction effect. The first column is interpreted as Type I error while the second and third columns are interpreted as power. The vertical line in the first column is the conventionally accepted optimal Type I error rate while the vertical line in the remaining columns is the conventionally accepted optimal power.

error for possible inflation. Since all null hypotheses were rejected under $\alpha = .05$,

rejection rates of $\gamma_3 = 0$ should ideally be around .05 under the null ($\Delta R^2 = 0$) condition.

Under normality, LMS-MLR had slightly elevated Type I error rates, but were considered

tolerable. All analysis conditions under normality were deemed to have acceptable Type I error rates with no noticeable differences between complete and PMD designs.

As Type I error is uninflated under normality, power may be interpreted without concern. Under normality and the $\Delta R^2 = .035$ condition, power with 200 cases was below the conventionally accepted .80 level. Although power to detect a $\Delta R^2 = .035$ effect was slightly lower in the PMD design than in the complete design, those differences appear to become less apparent as sample size increases. Power to detect a $\Delta R^2 = .035$ effect was around .90 and higher for sample sizes 375 and 550. Power to detect a $\Delta R^2 = .100$ effect was greater than .95 across all other conditions under normality.

Under non-normality, Type I error was found to be inflated for all analysis conditions except DMC-ML. LMS-MLR was shown to have the most error inflation followed by LMS-ML and then DMC-MLR. For LMS, there were no differences between 375 and 550 cases in terms of Type I error rate under the complete design. Meanwhile, differences were noted between 375 and 550 cases under the PMD design. Under LMS, Type I error rate became more inflated with increasing sample size.

Because inflated Type I errors have likely inflated power in some of the analysis conditions, power under non-normality is only examined for the DMC-ML condition. Under non-normality and the $\Delta R^2 = .035$ condition, power with 200 cases was below the conventionally accepted .80 level using the DMC-ML analysis. Power to detect a $\Delta R^2 = .035$ effect was at least .80 for sample sizes 375 and 550. Like the normal case, there are minute differences in power between the complete and PMD designs under non-

normality. Power for DMC-ML to detect a $\Delta R^2 = .100$ effect was greater than .875 across all sample size conditions.

Coverage

Coverage rates are presented in Figure 4. Under normality, only the LMS-ML analysis method tended to have optimal 95% coverage rates across all studied interaction effect sizes. DMC-ML had 95% coverage for the null effect; however, coverage for DMC-ML worsened as the interaction effect increased. For MLR conditions, coverage tends to improve as sample size increases; meanwhile, there were no noticeable differences in coverage among sample sizes for the ML conditions. There were no noticeable differences in coverage between complete and PMD designs.

Under non-normality, only the DMC-ML method under the null condition had the optimal 95% coverage rate. LMS-MLR had worse coverage than LMS-ML across all non-normal conditions. Across all effect size conditions, there were pronounced differences in coverage among sample size for the LMS analyses. For LMS, larger sample sizes had worse coverage than smaller sample sizes. Differences among sample size for DMC became more noticeable as interaction effect size increased. Like the LMS condition, for DMC-ML, larger sample sizes tended to have worse coverage than smaller sample sizes. Conversely, for DMC-MLR, larger sample sizes tended to have better coverage than smaller sample sizes in general. Coverage was noticeably worse between complete and PMD designs for all analyses except DMC-ML.

Figure 4. Coverage rates for the population interaction value. The vertical line in each panel is the conventionally accepted optimal coverage rate .95.

Relative Efficiency

Because relative efficiency (RE) was unaffected by standard error estimation, differences between standard error estimators were not examined for RE. Results for RE are presented in Figure 5. Under normality, there were no noticeable differences among interaction effect size or between estimation methods. RE ranged from .854 to .907 under

Figure 5. Relative efficiency.

normality. These results suggest that complete designs with 85.4% to 90.7% of the sample size of a PMD design are required to achieve equivalent power. Alternatively, a PMD design would need 10.3% to 17.1% additional cases to match the power of a complete design.

Under non-normality, an anomalous result was found under the null condition for DMC estimation with 200 cases. RE in this condition was 1.052, which contradicts Rhemtulla et al.'s (2014) assertion that RE is bounded between 0 and 1. An RE greater than one implies that the PMD design was *more* efficient than the complete design. The confidence interval for the null-DMC-200 case condition, however, is exceptionally wide compared to all other conditions and overlaps the acceptable RE range. Thus, the result

may be an artifact of Monte Carlo error than evidence that PMD designs can be more efficient than complete designs.

Under non-normality, RE confidence intervals were largest for DMC with 200 cases across all interaction effect sizes. It may be argued that, because the confidence intervals for 200 cases tend to overlap the confidence intervals for 375 and 550 cases under DMC, there is no noticeable difference among sample size in terms of RE for the DMC analysis. A more conservative claim would be that there are at least no noticeable differences between the 375- and 550-case conditions under DMC. Under the LMS analysis, there were no noticeable differences among sample size across all interaction effect sizes. LMS appears to also have smaller RE than DMC across all interaction effect sizes.

Discussion

The primary purpose of this study was to determine and compare the effectiveness of LMS and DMC on data from a PMD design. To supplement this aim, results from the PMD design were compared with a complete design to ensure that PMD designs were viable in a latent interaction modeling. Different standard error estimation methods were also of interest as they had not been examined in prior research comparing LMS and other product indicator approaches (Cham et al., 2017; Cham et al., 2012; Kelava et al., 2011; Lin et al., 2010). The discussion follows with respect to the original research questions. Then, limitations, future directions, and implications are presented.

How does latent interaction estimation perform between compete and PMD design conditions? For normally distributed data, no noticeable differences were found between complete and PMD design conditions in terms of parameter bias, SE bias, coverage, and type I error. PMD designs appeared to perform worse than complete designs in terms of power. Consistent with Graham et al. (2006), the differences in power between complete and PMD designs under normality were minimal. Relative efficiency was above 85.4% for normally distributed data and is consistent with relative efficiency values found by Rhemtulla et al. (2014). These results do not suggest that applied researchers using PMD designs must immediately recruit between 10.3% and 17.1% more participants to match the power of a complete design.

Attempting to increase sample size may be more costly than accepting the lower power of the PMD design. For example, consider a medium interaction effect ($\Delta R^2 = .035$) with 375 participants under a PMD design with normally distributed data. Using

DMC with ML standard errors, power was .896. With complete data, power was .928, a difference of .032 from the PMD design. Relative efficiency for this condition was .868, suggesting that the number of participants in the PMD design must be increased by 15%, or approximately 57 participants, to match the power of the complete design with 375 participants. Recruiting 57 additional participants for a theoretical increase in .032 power may not justify the added cost of the additional recruits. Nonetheless, Graham et al.'s (2006) assertion that PMD designs negligibly hamper power is supported by this study's findings for normally distributed data.

Under non-normality, however, differences between complete and PMD designs were noted for nearly all study outcomes. Parameter estimates were more positively biased for LMS under incomplete data than complete data. For the null, DMC, 200-case condition, there was dramatic uncertainty in the negative standard error bias with complete data than incomplete data. Type I error was more inflated for LMS with incomplete data than with complete data. These differences, however, may not be critical given that, across complete and PMD conditions, parameter estimates were already positively biased for LMS, standard errors were already negatively biased for the null, DMC, 200-case condition, and Type I error was already inflated for LMS. That is, any differences between complete and PMD designs with non-normal data are not as critical as the faults of the estimation methods in general. This lack of critical differences between complete and incomplete data is consistent with SEM research on the use of FIML with incomplete, non-normal data (Enders, 2001b).

For normally distributed data, differences between the complete and PMD designs were discussed in terms of power and RE. Such a discussion is ignored for non-normal data because power and RE are ultimately derived from information that is biased under non-normality. That is, discussion of power and RE are confounded by complications in the parameter estimates and standard errors.

Given PMD, how does the performance of LMS and DMC compare across conditions of distribution shape and sample size? It was hypothesized that LMS would outperform DMC when data were normal, but DMC would outperform LMS when data were non-normal. Under normality, LMS produced less biased parameter estimates at smaller sample sizes than DMC. LMS also produced less biased standard errors than DMC across different interaction effect sizes. In general, standard error bias was reduced with increasing sample size. Type I error rates were acceptable for both methods and power approached optimal levels above a sample size of 200. With 375 cases, power to detect a medium interaction effect was close to .90, suggesting that Rhemtulla and Little's (2012) sample size recommendation for a 3-form PMD design may be conservative. The current study's model, however, is fairly simple; more conservative recommendations may be appropriate given possible added complexities during applied research.

LMS appears to have advantages over DMC when data were normally distributed; however, consistent with Cham et al. (2012), LMS produced considerably biased parameter estimates under non-normality. Parameter estimates from DMC, meanwhile, continued to be unbiased. With regard to parameter bias, Cham et al.'s (2012)

recommendation for unconstrained product indicator approaches with non-normal data is maintained. Cham et al. (2012), however, did not examine standard error bias. Under non-normality, DMC consistently produced negatively biased standard errors that tended to be more biased than with LMS. Also under non-normality, DMC produced increasingly biased standard errors as the interaction effect increased. Standard error bias was mitigated with increasing sample size; however, sample sizes much larger than 550 would be required to possibly obtain unbiased standard errors for DMC.

One peculiar finding under non-normality was the reverse relationship between LMS coverage and sample size. As sample size increased, LMS coverage worsened. This is likely due to the inherent positive bias of the parameter estimate and a narrowing standard error as sample size increases. That is, as sample size increases, the confidence bounds move towards the positively biased parameter estimate and further away from the true interaction effect.

How do normal-theory standard error calculation and robust standard error calculation compare across conditions of analysis method, distribution shape, and sample size? Previous research had not compared Huber-White robust standard errors to normal-theory standard errors when estimating latent interaction effects (Cham et al., 2017, 2012; Enders et al., 2014; Klein & Moosbrugger, 2000; Lin et al., 2010). Given the non-normal distribution of the interaction term (Dimitruk et al., 2007) and given Huber-White standard errors' robustness to non-normality (Muthén & Muthén, 2012), it was hypothesized that Huber-White robust standard errors would outperform normal-theory

standard errors in all conditions. Contrary to this hypothesis, robust standard errors did not always offer benefits over normal standard errors.

Under normality, normal standard errors were slightly better than robust standard errors when using the LMS method. Conversely, robust standard errors remained unbiased for DMC while normal standard errors became more biased with increasing interaction effect size. For LMS, coverage was slightly better with normal standard errors than with robust standard errors. This relationship again reverses for DMC where robust standard errors result in higher coverage than normal standard errors. For both analytical methods, robust standard errors resulted in higher Type I error rates than normal standard errors. In general, LMS performed better with normal standard errors while DMC performed better with robust standard errors, a surprising finding given robust standard errors' supposed robustness to non-normality (Muthén & Muthén, 2012). Furthermore, robust standard errors are the defaults when using LMS in Mplus 7.0 while normal standard errors are the default when using DMC (Muthén & Muthén, 2012).

Similar patterns were found under non-normality; robust standard errors tended to outperform normal standard errors when using LMS and vice versa while using DMC. This pattern did not replicate for coverage while using DMC. As the interaction effect increased in size, normal standard error bias became worse than robust standard error bias.

Limitations and future directions

The present study is not without limitations and opportunities for future research. The first limitation was that a simple latent interaction model was used and may not

generalize to more complex models that researchers will hope to investigate. For example, researchers may add quadratic forms of the lower-order predictors to their model. The inclusion of quadratic terms is recommended in interaction modeling because an interaction effect may be detected when in reality one of the lower-order predictors has a curvilinear relationship with the outcome (Ganzach, 1997; Kelava, Moosbrugger, Dimitruk, & Schermelleh-Engel, 2008). Such quadratic terms were omitted from the present study on the assumption that applied researchers are including interactions based on theory in which there are no errant curvilinear relationships.

Another limitation in the present study was that the simulated complete design was ideal. That is, the concerns that encourage the use of a PMD over a complete design (e.g. participant burden, test validity) were not built into the simulation of the complete data. The simulated complete data is still useful as a reference point; readers can use the simulated reference point to gauge how PMD designs differ from the most ideal approximation of data. Nonetheless, future methodological research of PMD should begin considering the cost/benefits of a realistic complete design (including participant burden and test validity) compared to the costs/benefits of a PMD design.

The third limitation was that only the three-form PMD design was considered. The three-form design was studied here because it is one of the most commonly mentioned and studied PMD design in the methodological literature (Graham et al., 2006; Rhemtulla et al., 2016). Other designs such as multi-form or two-method designs are possible and worth consideration in future research with possibly different contexts. For

instance, although a three-form design is sensible in a cohort study, a wave missing design (Rhemtulla & Hancock, 2016) might be more appropriate.

The fourth limitation was that only a single-level interaction model was considered. In educational research, children or students are nested within classrooms and classrooms are nested within schools. This nesting tends to violate the assumption of independent observations but can be accounted for with multilevel modeling (Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). It was important in the current study, however, to first investigate the PMD design with a comparatively simpler interaction model. Future research will extend these findings into multilevel contexts that introduce the possibility of within- and cross-level interactions.

Conclusions and implications for applied research

The present study illustrates that PMD designs can result in fairly minimal loss of power and extends that assertion to latent interaction modeling. Some cautions are advised. First, applied researchers should be aware, however, of software defaults. With normally distributed data, LMS tends to perform better with ML standard errors rather than the default robust standard errors in Mplus 7.0. Meanwhile, DMC tends to perform better with robust standard errors rather than default ML standard errors in Mplus 7.0. Defaults aside, DMC with robust standard errors tended to perform nearly as well as LMS. If applied researchers lack access to Mplus and thus lack access to LMS, DMC, which can be implemented in most if not all SEM packages, may be an acceptable alternative. A challenge with DMC, however, is that it relies on the creation of products

from indicator pairs. If latent variables have an unequal number of indicators, researchers will need to explore alternative strategies such as parceling (Marsh et al., 2004).

Under non-normality, either analysis method had some complication (e.g. biased parameters, biased standard errors). DMC had unbiased parameter estimates but biased standard errors while LMS had biased parameter estimates and tended to have unbiased standard errors. DMC's standard error issues may be mitigated with bootstrapping methods (Enders, 2001b; Kelava et al., 2011). Meanwhile, mixture modeling approaches to account for general lower-order non-normality might help to correct problems with LMS (Kelava, Nagengast, & Brandt, 2014). Future research is needed to explore these “non-normality treatments” in PMD contexts. In the meantime, applied researchers should be aware of their data's distributional characteristics and the potential problems that will result.

References

- Algina, J., & Moulder, B. C. (2001). A note on estimating the Jöreskog-Yang model for latent variable interaction using LISREL 8.3. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(1), 40–52.
- 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).
- Bachman, J. G., Johnston, L. D., O'Malley, P. M., & Schulenberg, J. E. (2011). *The Monitoring the Future Project after thirty-seven years: Design and procedures*. Ann Arbor, MI.
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53, 605–634.
- Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research & Perspective*, 6(1-2), 25–53.
- Cham, H., Reshetnyak, E., Rosenfeld, B., & Breitbart, W. (2017). Full information maximum likelihood estimation for latent variable interactions with incomplete indicators. *Multivariate Behavioral Research*, 52(1), 12–30.
- Cham, H., West, S. G., Ma, Y., & Aiken, L. S. (2012). Estimating latent variable interactions with non-normal observed data: A comparison of four approaches. *Multivariate Behavioral Research*, 47(6), 840–876.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression and correlation for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1(1), 16–29.
- Dimitruk, P., Schermelleh-Engel, K., Kelava, A., & Moosbrugger, H. (2007). Challenges in nonlinear structural equation modeling. *Methodology*, 3(3), 100–114.
- Enders, C. K. (2001a). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(1), 128–141.
- Enders, C. K. (2001b). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, 6(4), 352–70.
- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: The Guilford Press.
- Enders, C. K., & Bandalos, D. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*.
- Enders, C. K., Baraldi, A. N., & Cham, H. (2014). Estimating interaction effects with incomplete predictor variables. *Psychological Methods March 2014*, 19(1), 39–55.
- Finkelstein, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika*, 44(4), 409–420.

- Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43(4), 521–532.
- Ganzach, Y. (1997). Misleading interaction and curvilinear terms. *Psychological Methods*, 2(3), 235–247.
- 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(2), 197–218.
- Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, 8(3), 206–213.
- Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, 11(4), 323–343.
- Harel, O., Stratton, J., & Aseltine, R. (2015). Designed missingness to better estimate efficacy of behavioral studies-application to suicide prevention trials. *Journal of Medical Statistics and Informatics*, 3(1), 2.
- Hayes, A. F., Glynn, C. J., & Huge, M. E. (2012). Cautions regarding the interpretation of regression coefficients and hypothesis tests in linear models with interactions. *Communication Methods and Measures*, 6(1), 1–11.
- Hernández, M. M., Eisenberg, N., Valiente, C., Vanschyndel, S. K., Spinrad, T. L., Silva, K. M., ... Southworth, J. (2016). Emotional expression in school context, social relationships, and academic adjustment in kindergarten. *Emotion*, 16(4), 553–566.

- Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods and Research*, 26, 329–367.
- Institute of Education Sciences. (2016). *Request for applications: Education Research Grants: CFDA Number: 84.305A*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.
- Jöreskog, K. G., & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 57–88).
- Kelava, A., Moosbrugger, H., Dimitruk, P., & Schermelleh-Engel, K. (2008). Multicollinearity and missing constraints: A comparison of three approaches for the analysis of latent nonlinear effects. *Methodology*, 4(2), 51–66.
- Kelava, A., Nagengast, B., & Brandt, H. (2014). A nonlinear structural equation mixture modeling approach for nonnormally distributed latent predictor variables. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3), 468–481.
- Kelava, A., Werner, C. S., Schermelleh-Engel, K., Moosbrugger, H., Zapf, D., Ma, Y., ... West, S. G. (2011). Advanced nonlinear latent variable modeling: Distribution analytic LMS and QML estimators of interaction and quadratic effects. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(3), 465–491.
- Kenny, D. a., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, 96(1), 201–210.
- Klein, A. G., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent

- interaction effects with the LMS method. *Psychometrika*, 65(4), 457–474.
- Klein, A. G., & Muthén, B. O. (2007). Quasi-Maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research*, 42(4), 647–673.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York, NY: The Guilford Press.
- Kwok, O.-M., Im, M., Hughes, J. N., Wehrly, S. E., & West, S. G. (2016). Testing statistical moderation in research on home-school partnerships: Establishing the boundary condition. In S. M. Sheridan & E. M. Kim (Eds.), *Family-School Partnerships in Context*. Springer International Publishing.
- Lin, G.-C., Wen, Z., Marsh, H., & Lin, H.-S. (2010). Structural equation models of latent interactions: Clarification of orthogonalizing and double-mean-centering strategies. *Structural Equation Modeling: A Multidisciplinary Journal*, 17(3), 374–391.
- Little, T. D., Bovaird, J. A., & Widaman, K. F. (2006). On the merits of orthogonalizing powered and product terms: Implications for modeling interactions among latent variables. *Structural Equation Modeling: A Multidisciplinary Journal*.
- Marsh, H. W., Wen, Z., & Hau, K.-T. (2004). Structural equation models of latent interactions: evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, 9(3), 275–300.
- Marsh, H. W., Wen, Z., Nagengast, B., & Hau, K.-T. (2012). Structural equation models of latent interaction. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling*. New York, NY: The Guilford Press.

- Masland, L. C., & Lease, A. M. (2016). Characteristics of academically-influential children: achievement motivation and social status. *Social Psychology of Education, 19*, 195–215.
- Maslowsky, J., Jager, J., & Hemken, D. (2014). Estimating and interpreting latent variable interactions: A tutorial for applying the latent moderated structural equations method. *International Journal of Behavioral Development, 39*(1), 87–96.
- Moosbrugger, H., Schermelleh-Engel, K., Kelava, A., & Klein, A. G. (2006). Testing multiple nonlinear effects in structural equation modeling: A comparison of alternative estimation. *Structural Equation Modeling in Educational Research: Concepts and Applications*, 103–136.
- Muthén, B. O., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 52*(3), 431–462.
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- National Institute of Health. (2017). *Request for applications: Education and Health: New Frontiers: CFDA Number: 93.865*. Washington, DC: U.S. Department of Health and Human Services: National Institute of Health. Retrieved from <https://grants.nih.gov/grants/guide/pa-files/PAR-16-079.html>
- R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and*

- data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Rhemtulla, M., & Hancock, G. R. (2016). Planned missing data designs in educational psychology research. *Educational Psychologist*, 51(3-4), 305–316.
- Rhemtulla, M., Jia, F., Wu, W., & Little, T. D. (2014). Planned missing designs to optimize the efficiency of latent growth parameter estimates. *International Journal of Behavioral Development*.
- Rhemtulla, M., & Little, T. D. (2012). Planned missing data designs for research in cognitive development. *Journal of Cognition and Development*, 13(4), 425–438.
- Rhemtulla, M., Savalei, V., & Little, T. D. (2016). On the asymptotic relative efficiency of planned missingness designs. *Psychometrika*, 81(1), 60–89.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley.
- Smits, N., & Vorst, H. C. M. (2007). Reducing the length of questionnaires through structurally incomplete designs: An illustration. *Learning and Individual Differences*, 17(1), 25–34.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand oaks, CA: Sage Publications.
- Thorndike, R. M., & Thorndike-Christ, T. (2010). *Measurement and evaluation in psychology and education* (8th ed.). Boston, MA: Pearson.
- Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, 48(3), 465–471.

Yang, F. (1998). Modeling interaction and nonlinear effects: A step-by-step LISREL example. In R. E. Schumacker & G. a. Marcoulides (Eds.), *Interaction and nonlinear effects in structural equation modelling* (pp. 17–42). Mahwah, NJ: Erlbaum.

Appendix A. Example Mplus syntax

Below are example Mplus syntax for the different analysis conditions. Note that ESTIMATOR=MLR for the robust standard error condition.

For DMC method:

```
ANALYSIS:
  TYPE = GENERAL;
  ESTIMATOR = ML;
  COVERAGE = .00001;
  PROCESSORS = 3;

MODEL:
  X BY x1c-x4c;
  Z BY z1c-z4c;
  XZ BY x1cz1cc-x4cz4cc;
  Y BY y1c-y4c;

  X WITH Z* XZ*;
  Z WITH XZ*;

  Y ON X Z XZ;

  [x1c-y4c@0 x1cz1cc-x4cz4cc@0];
```

For LMS method:

```
ANALYSIS:
  TYPE = RANDOM;
  ALGORITHM = INTEGRATION;
  ESTIMATOR = ML;
  PROCESSORS = 3;

MODEL:
  X BY x1-x4;
  Z BY z1-z4;
  Y BY y1-y4;
  XZ | X XWITH Z;
  Y ON X Z XZ;
```

Appendix B. Tables of results

Table B.1. *Parameter bias and relative parameter bias of the interaction effect.*

Distribution	ΔR^2	Sample size	Complete				PMD			
			LMS		DMC		LMS		DMC	
			ML	MLR	ML	MLR	ML	MLR	ML	MLR
Normal	0	200	.000 (-.002, .001)	.000 (-.002, .001)	.000 (-.002, .001)	.000 (-.002, .001)	-.001 (-.003, .000)	-.001 (-.003, .000)	-.001 (-.003, .000)	-.001 (-.003, .000)
		375	.001 (.000, .002)	.001 (.000, .002)	.001 (.000, .002)	.001 (.000, .002)	.000 (-.001, .001)	.000 (-.001, .001)	.000 (-.001, .001)	.000 (-.001, .001)
		550	.000 (-.001, .001)	.000 (-.001, .001)	.000 (-.001, .001)	.000 (-.001, .001)	.000 (-.001, .000)	.000 (-.001, .000)	.000 (-.001, .001)	.000 (-.001, .001)
	.035	200	.010 (.002, .017)	.010 (.002, .017)	.026 (.018, .035)	.026 (.018, .035)	.003 (-.005, .011)	.003 (-.005, .011)	.018 (.009, .027)	.018 (.009, .027)
		375	.003 (-.003, .009)	.003 (-.003, .009)	.009 (.003, .015)	.009 (.003, .015)	.008 (.002, .014)	.008 (.002, .014)	.016 (.010, .022)	.016 (.010, .022)
		550	.004 (-.001, .008)	.004 (-.001, .008)	.010 (.006, .015)	.010 (.006, .015)	-.003 (-.008, .002)	-.003 (-.008, .002)	.003 (-.002, .008)	.003 (-.002, .008)
	.100	200	.006 (.001, .010)	.006 (.001, .010)	.022 (.017, .028)	.022 (.017, .028)	.003 (-.002, .008)	.003 (-.002, .008)	.019 (.013, .025)	.019 (.013, .025)
		375	.003 (-.001, .006)	.003 (-.001, .006)	.010 (.006, .013)	.010 (.006, .013)	.002 (-.002, .005)	.002 (-.002, .005)	.008 (.004, .012)	.008 (.004, .012)
		550	.001 (-.002, .004)	.001 (-.002, .004)	.008 (.005, .011)	.008 (.005, .011)	.001 (-.002, .004)	.001 (-.002, .004)	.007 (.004, .010)	.007 (.004, .010)
Non-normal	0	200	.020 (.018, .021)	.020 (.018, .021)	.000 (-.001, .002)	.000 (-.001, .002)	.027 (.026, .028)	.027 (.026, .028)	.001 (-.001, .002)	.001 (-.001, .002)
		375	.020 (.020, .021)	.020 (.020, .021)	.002 (.001, .003)	.002 (.001, .003)	.026 (.025, .027)	.026 (.025, .027)	.001 (.000, .001)	.001 (.000, .001)
		550	.019 (.019, .020)	.019 (.019, .020)	.001 (.000, .001)	.001 (.000, .001)	.025 (.025, .026)	.025 (.025, .026)	.000 (-.001, .000)	.000 (-.001, .000)

.035	200	.156 (.146, .165)	.156 (.146, .165)	.038 (.026, .049)	.038 (.026, .049)	.201 (.191, .212)	.201 (.191, .212)	.027 (.015, .039)	.027 (.015, .039)
	375	.154 (.147, .161)	.154 (.147, .161)	.016 (.009, .023)	.016 (.009, .023)	.212 (.205, .219)	.212 (.205, .219)	.019 (.011, .026)	.019 (.011, .026)
	550	.158 (.153, .163)	.158 (.153, .163)	.015 (.009, .020)	.015 (.009, .020)	.208 (.202, .214)	.208 (.202, .214)	.008 (.002, .014)	.008 (.002, .014)
.100	200	.085 (.080, .091)	.085 (.080, .091)	.030 (.023, .037)	.030 (.023, .037)	.115 (.109, .121)	.115 (.109, .121)	.024 (.016, .032)	.024 (.016, .032)
	375	.093 (.089, .097)	.093 (.089, .097)	.016 (.012, .021)	.016 (.012, .021)	.124 (.120, .129)	.124 (.120, .129)	.017 (.012, .022)	.017 (.012, .022)
	550	.098 (.095, .101)	.098 (.095, .101)	.015 (.011, .018)	.015 (.011, .018)	.127 (.123, .130)	.127 (.123, .130)	.011 (.008, .015)	.011 (.008, .015)

Note: 95% confidence intervals are in parenthesis. Shaded values are parameter bias, unshaded values are relative parameter

bias.

Table B.2. *Relative standard error bias*

Distribution	ΔR^2	Sample size	Complete				PMD			
			LMS		DMC		LMS		DMC	
			ML	MLR	ML	MLR	ML	MLR	ML	MLR
Normal	0	200	-.017 (-.032, -.003)	-.045 (-.059, -.031)	-.028 (-.044, -.013)	-.041 (-.058, -.027)	-.022 (-.036, -.009)	-.040 (-.053, -.025)	-.037 (-.051, -.022)	-.033 (-.046, -.018)
		375	-.015 (-.027, -.001)	-.031 (-.044, -.017)	-.022 (-.036, -.009)	-.030 (-.043, -.016)	-.014 (-.028, .001)	-.027 (-.041, -.012)	-.021 (-.035, -.005)	-.020 (-.034, -.004)
		550	-.006 (-.020, .008)	-.018 (-.033, -.004)	-.014 (-.028, .001)	-.020 (-.034, -.005)	-.016 (-.031, -.002)	-.026 (-.040, -.012)	-.025 (-.040, -.011)	-.024 (-.038, -.010)
	.035	200	-.017 (-.031, -.003)	-.042 (-.057, -.027)	-.063 (-.077, -.048)	-.049 (-.065, -.033)	-.020 (-.033, -.006)	-.038 (-.052, -.024)	-.068 (-.083, -.054)	-.041 (-.056, -.026)
		375	-.016 (-.029, -.002)	-.032 (-.045, -.018)	-.052 (-.066, -.038)	-.030 (-.044, -.016)	-.023 (-.036, -.008)	-.037 (-.049, -.022)	-.058 (-.071, -.044)	-.030 (-.043, -.016)
		550	-.002 (-.016, .012)	-.015 (-.029, -.001)	-.039 (-.052, -.025)	-.013 (-.027, .001)	-.019 (-.032, -.005)	-.027 (-.040, -.013)	-.059 (-.071, -.045)	-.028 (-.040, -.013)
	.100	200	-.013 (-.025, .001)	-.035 (-.048, -.021)	-.103 (-.115, -.090)	-.049 (-.062, -.036)	-.021 (-.034, -.006)	-.037 (-.051, -.022)	-.110 (-.122, -.096)	-.046 (-.059, -.031)
		375	-.016 (-.031, -.003)	-.031 (-.045, -.019)	-.106 (-.120, -.093)	-.040 (-.055, -.025)	-.014 (-.027, .000)	-.023 (-.036, -.010)	-.106 (-.118, -.093)	-.035 (-.048, -.020)
		550	-.006 (-.020, .008)	-.018 (-.031, -.003)	-.096 (-.108, -.083)	-.022 (-.036, -.008)	-.010 (-.025, .003)	-.019 (-.034, -.006)	-.095 (-.108, -.081)	-.019 (-.034, -.004)
Non-normal	0	200	-.056 (-.070, -.038)	-.129 (-.142, -.112)	-.155 (-.246, -.086)	-.188 (-.258, -.132)	-.061 (-.076, -.045)	-.122 (-.136, -.108)	-.093 (-.113, -.073)	-.115 (-.134, -.096)
		375	-.043 (-.056, -.029)	-.107 (-.119, -.093)	-.055 (-.069, -.040)	-.102 (-.115, -.087)	-.038 (-.053, -.023)	-.097 (-.110, -.082)	-.061 (-.076, -.045)	-.098 (-.113, -.083)
		550	-.020 (-.034, -.005)	-.076 (-.090, -.062)	-.031 (-.046, -.014)	-.075 (-.089, -.059)	-.033 (-.048, -.019)	-.083 (-.097, -.068)	-.046 (-.061, -.030)	-.081 (-.096, -.064)
	.035	200	-.047 (-.063, -.031)	-.116 (-.130, -.101)	-.140 (-.174, -.113)	-.144 (-.167, -.123)	-.054 (-.069, -.039)	-.110 (-.124, -.096)	-.142 (-.160, -.126)	-.128 (-.145, -.112)

.100	375	-.035 (-.049, -.020)	-.093 (-.106, -.079)	-.127 (-.142, -.113)	-.122 (-.136, -.107)	-.037 (-.052, -.021)	-.087 (-.102, -.072)	-.124 (-.138, -.108)	-.114 (-.128, -.097)
	550	-.020 (-.034, -.005)	-.071 (-.086, -.058)	-.113 (-.127, -.098)	-.098 (-.113, -.083)	-.026 (-.040, -.011)	-.073 (-.087, -.059)	-.117 (-.131, -.102)	-.099 (-.114, -.083)
	200	-.047 (-.061, -.033)	-.109 (-.122, -.096)	-.217 (-.239, -.199)	-.175 (-.197, -.156)	-.047 (-.062, -.032)	-.099 (-.113, -.085)	-.229 (-.253, -.206)	-.172 (-.192, -.150)
	375	-.021 (-.036, -.007)	-.074 (-.088, -.060)	-.209 (-.222, -.196)	-.140 (-.154, -.126)	-.052 (-.067, -.039)	-.096 (-.110, -.083)	-.215 (-.230, -.201)	-.143 (-.158, -.128)
	550	-.022 (-.036, -.007)	-.066 (-.079, -.051)	-.218 (-.230, -.204)	-.125 (-.139, -.109)	-.032 (-.047, -.019)	-.068 (-.083, -.055)	-.214 (-.227, -.201)	-.127 (-.140, -.112)

Note: 95% confidence intervals are in parenthesis.

Table B.3. *Type I error and power*

Distribution	dR2	Sample size	Complete				PMD			
			LMS		DMC		LMS		DMC	
			ML	MLR	ML	MLR	ML	MLR	ML	MLR
Normal	0	200	.047 (.043, .051)	.063 (.058, .068)	.049 (.044, .053)	.055 (.051, .059)	.050 (.045, .054)	.063 (.058, .068)	.048 (.044, .052)	.054 (.049, .058)
		375	.050 (.046, .055)	.062 (.057, .066)	.049 (.045, .054)	.055 (.051, .059)	.050 (.046, .054)	.060 (.055, .065)	.052 (.048, .056)	.054 (.050, .059)
		550	.051 (.047, .056)	.058 (.053, .062)	.052 (.048, .057)	.057 (.052, .061)	.053 (.049, .058)	.062 (.058, .067)	.054 (.049, .058)	.056 (.051, .060)
		.035								
		200	.744 (.735, .752)	.752 (.744, .760)	.724 (.716, .733)	.713 (.704, .722)	.698 (.689, .707)	.706 (.697, .715)	.663 (.654, .672)	.642 (.632, .651)
		375	.948 (.944, .952)	.948 (.943, .952)	.939 (.934, .944)	.933 (.928, .938)	.928 (.923, .933)	.924 (.919, .929)	.912 (.906, .917)	.902 (.896, .908)
		550	.993 (.991, .994)	.992 (.990, .994)	.990 (.988, .992)	.989 (.987, .991)	.985 (.982, .987)	.984 (.981, .986)	.979 (.976, .982)	.977 (.974, .980)
	.100	200	.992 (.990, .994)	.990 (.988, .992)	.988 (.985, .990)	.983 (.980, .985)	.982 (.980, .985)	.979 (.977, .982)	.975 (.972, .978)	.962 (.959, .966)
		375	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (.999, 1.000)
		550	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)
Non-normal	0	200	.061 (.056, .065)	.104 (.098, .110)	.048 (.044, .052)	.078 (.073, .083)	.066 (.061, .070)	.106 (.100, .112)	.049 (.045, .054)	.069 (.064, .074)
		375	.081 (.075, .086)	.119 (.113, .125)	.053 (.049, .057)	.077 (.072, .083)	.089 (.083, .094)	.130 (.123, .136)	.052 (.048, .056)	.074 (.069, .079)
		550	.083 (.077, .088)	.120 (.113, .126)	.049 (.045, .054)	.073 (.068, .078)	.112 (.106, .118)	.146 (.139, .153)	.051 (.047, .055)	.070 (.065, .075)

.035	200	.700 (.691, .709)	.734 (.725, .742)	.609 (.600, .619)	.615 (.606, .625)	.683 (.673, .692)	.711 (.702, .720)	.562 (.552, .571)	.561 (.551, .571)
	375	.925 (.920, .930)	.927 (.922, .932)	.860 (.854, .867)	.851 (.844, .858)	.921 (.915, .926)	.924 (.919, .929)	.822 (.815, .830)	.814 (.806, .821)
	550	.983 (.981, .986)	.985 (.983, .987)	.953 (.949, .958)	.946 (.942, .950)	.983 (.980, .985)	.982 (.979, .985)	.937 (.932, .941)	.929 (.924, .934)
.100	200	.944 (.940, .949)	.946 (.942, .951)	.912 (.907, .918)	.893 (.887, .899)	.942 (.937, .946)	.939 (.934, .944)	.888 (.882, .894)	.855 (.848, .862)
	375	.998 (.997, .999)	.998 (.997, .998)	.995 (.994, .997)	.992 (.990, .994)	.997 (.996, .998)	.997 (.996, .998)	.991 (.989, .992)	.984 (.981, .986)
	550	1.000 (.999, 1.000)	1.000 (1.000, 1.000)	1.000 (.999, 1.000)	.999 (.999, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (.999, 1.000)	.999 (.999, 1.000)

Note: 95% confidence intervals are in parenthesis.

Table B.4. *Relative efficiency*

Distribution	dR2	Sample size	LMS		DMC	
			ML	MLR	ML	MLR
Normal	0	200	.901 (.864, .939)	.901 (.864, .939)	.876 (.837, .916)	.876 (.837, .916)
		375	.907 (.869, .944)	.907 (.869, .944)	.893 (.856, .933)	.893 (.856, .933)
		550	.883 (.846, .922)	.883 (.846, .922)	.865 (.829, .902)	.865 (.829, .902)
	.035	200	.898 (.862, .936)	.898 (.862, .936)	.870 (.830, .913)	.870 (.830, .913)
		375	.888 (.855, .926)	.888 (.855, .926)	.868 (.831, .907)	.868 (.831, .907)
		550	.878 (.844, .913)	.878 (.844, .913)	.856 (.822, .893)	.856 (.822, .893)
	.100	200	.880 (.845, .915)	.880 (.845, .915)	.854 (.816, .894)	.854 (.816, .894)
		375	.905 (.871, .943)	.905 (.871, .943)	.880 (.844, .918)	.880 (.844, .918)
		550	.891 (.855, .925)	.891 (.855, .925)	.882 (.847, .920)	.882 (.847, .920)
	Non-normal	200	.901 (.858, .946)	.901 (.858, .946)	1.052 (.873, 1.359)	1.052 (.873, 1.359)
		375	.902 (.861, .941)	.902 (.861, .941)	.879 (.835, .923)	.879 (.835, .923)
		550	.871 (.832, .909)	.871 (.832, .909)	.861 (.818, .904)	.861 (.818, .904)
	.035	200	.882 (.839, .926)	.882 (.839, .926)	.885 (.809, .986)	.885 (.809, .986)
		375	.876 (.841, .917)	.876 (.841, .917)	.887 (.843, .936)	.887 (.843, .936)
		550	.887 (.849, .926)	.887 (.849, .926)	.904 (.861, .947)	.904 (.861, .947)
	.100	200	.908 (.870, .953)	.908 (.870, .953)	.873 (.796, .958)	.873 (.796, .958)
		375	.825 (.789, .862)	.825 (.789, .862)	.865 (.820, .912)	.865 (.820, .912)
		550	.865 (.827, .900)	.865 (.827, .900)	.900 (.857, .944)	.900 (.857, .944)

Note: 95% confidence intervals are in parenthesis.