3-29-2017

Exploring the Test of Covariate Moderation Effect and the Impact of Model Misspecification in Multilevel MIMIC Models

Chunhua Cao
University of South Florida, chunhuacao@mail.usf.edu

Exploring the Test of Covariate Moderation Effect and the Impact of Model Misspecification in Multilevel MIMIC Models

by

Chunhua Cao

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
in Curriculum and Instruction
with an emphasis in Measurement and Evaluation
Department of Educational and Psychological Studies
College of Education
University of South Florida

Co-Major Professor: Eun Sook Kim, Ph.D.
Co-Major Professor: Yi-Hsin Chen, Ph.D.
John Ferron, Ph.D.
Stephen Stark, Ph.D.

Date of Approval:
March 21, 2017

Keywords: Multilevel, MIMIC, Covariates interaction, Misspecification

Acknowledgments

The completion of this dissertation would not have been possible without the support and caring of many people around me. I appreciate and cherish the support they provided to me. First and foremost, I want to express my gratitude for my co-major professors and mentors, Dr. Eun Sook Kim and Dr. Yi-Hsin Chen. Dr. Kim provided me with guidance and support throughout the design, conduct, and completion of this dissertation. She is always there to inspire and encourage me whenever I have questions or encounter difficulties. Dr. Chen's support and encouragement made me to decide to pursue my doctoral study in University of South Florida, and it is him that mentored me through the end. My thanks also go to other committee members, Dr. John Ferron and Dr. Stephen Stark. Dr. Ferron guided me in some projects and he always knew how to make students grow in their own path. Dr. Stark's intriguing questions and feedback made me think more clearly about the design and the writing of the dissertation.

In addition to my committee, I also want to thank other faculty members in our program of Educational Measurement and Research, Dr. Robert Dedrick, Dr. Jeffrey Kromery, Dr. Liliana Rodriguez-Campos, and Dr. Jennifer Wolgemuth. I benefited a lot from the courses I took from them. I feel very lucky to be in such a caring and supportive program. I also want to thank some of the fellow graduate students who made this journey joyful and fruitful, Paromita De, Vanessa Vernaza-Hernandez, Eunkyeng Baek, Yan Wang, Diep Nguyen, Thanh Pham, Isaac Li. I also want to thank my colleagues for their encouragement in my work place of Polk County School Board.

Lastly, I want to thank my family for their unconditional love and belief in me. Without their support I would not have achieved this goal. I thank my parents, Lingzong Cao and Xiangwen Wang for their love and support. I want to thank my husband, Dr. Guilong Li, for his support in my doctoral study. I also want to express gratitude for my two boys, Jierui Li, and Jayden Li, for motivating me.

Dedication

I want to dedicate this dissertation to my father who loved me and always encouraged me to pursue knowledge although he has been in my life for only twelve years and to my mother for her unconditional love. I also want to dedicate this dissertation to my husband and my two sons for their love and support.

Table of Contents

List of Tables

List of Figures

Abstract

In multilevel MIMIC models, covariates at the between level and at the within level can be modeled simultaneously. Covariates interaction effect occurs when the effect of one covariate on the latent factor varies depending on the level of the other covariate. The two covariates can be both at the between level, both at the within level, and one at the between level and the other one at the within level. And they can create between level covariates interaction, within level covariates interaction, and cross level covariates interaction. Study One purports to examine the performance of multilevel MIMIC models in estimating the covariates interaction described above. Type I error of falsely detecting covariates interaction when there is no covariates interaction effect in the population model, and the power of correctly detecting the covariates interaction effect, bias of the estimate of interaction effect, and RMSE are examined. The design factors include the location of the covariates interaction effect, cluster number, cluster size, intra-class correlation (ICC) level, and magnitude of the interaction effect. The results showed that ML MIMIC performed well in detecting the covariates interaction effect when the covariates interaction effect was at the within level or cross level. However, when the covariates interaction effect was at the between level, the performance of ML MIMIC depended on the magnitude of the interaction effect, ICC, and sample size, especially cluster size. In Study Two, the impact of omitting covariates interaction effect on the estimate of other parameters is investigated when the covariates interaction effect is present in the population model. Parameter estimates of factor loadings, intercepts, main effects of the covariates, and residual variances produced by the

correct model in Study One are compared to those produced by the misspecified model to check the impact. Moreover, the sensitivity of fit indices, such as chi-square, CFI, RMSEA, SRMR-B (between), and SRM-W (within) are also examined. Results indicated that none of the fit indices was sensitive to the omission of the covariates interaction effect. The biased parameter estimates included the two covariates main effect and the between-level factor mean.

Chapter One: Introduction

Over the past few decades, there has been increasing attention to investigating group differences on latent variables in many research studies in social behavioral sciences (Smeden & Hessen, 2013). Latent variables are not directly observed, but measured by other variables, such as attitude, confidence, and satisfaction. Muthén (1989) pointed out that in the application of latent variable analysis, the heterogeneity of subpopulations in parameter values should be taken into account. Multigroup confirmatory factor analysis (MG-CFA) has been a commonly used method for studying group differences in latent variables. However, in MG-CFA the model is fitted to data for each group separately by dividing the samples based on the number of groups (Muthén, 1989). This is problematic when sample size is a concern, especially when there are more than two groups, for example, four ethnic groups. The sample size of each group is much smaller than the total sample size when there are multiple groups. Moreover, MG-CFA usually limits to only one grouping variable (e.g., gender or race), but often there are multiple grouping covariates (e.g. gender and race) to be modeled simultaneously in the research context of social behavioral sciences. When more than one grouping variable is present, the multiple indicators multiple causes (MIMIC) model is a more effective alternative technique to study group differences in latent variables. In the presence of two or more grouping variables (e.g. gender, race, and categorical social economic status), instead of fitting the same CFA model to each of gender × race × social economic status groups as in MG-CFA, the MIMIC model fits only one model by regressing a set of dummy coded grouping variables on the construct of the CFA model.

The typical feature of the MIMIC model is that the latent factor is measured by the observed indicators, which is the measurement model part, and is regressed on observed grouping variables, which is the structural model part (Jöreskog & Goldberger, 1975). The MIMIC model was shown to detect and describe heterogeneity situations that the regular MG-CFA cannot handle (Muthén, 1989); that is, when total sample size is small or when there are multiple grouping variables. For example, when the sample size is small or even when the total sample size is big but with many groups, fitting multiple CFA models for each group renders unstable parameter estimate because of insufficient sample size in each group. In contrast to MG-CFA, the MIMIC model uses the total sample size to get the variance covariance matrix to estimate the parameters in the model without splitting the data into different groups based on the grouping variables. Apart from its advantage of no need to divide the total sample size based on number of groups, the MIMIC model is flexible in modeling multiple covariates, which may be continuous or discrete, observed or latent. The inclusion of a set of covariates provides the MIMIC model with important extra information in addition to the measurement model (Chen, 1981; Cheng, Shao, & Lathrop, 2016; Muthén, 1989). The MIMIC model has also been utilized extensively by applied researchers, especially for testing latent factor mean, or the regression effect from the grouping variable on the latent factor (Condon, 2010; Ogg, McMahan, Dedrick, Mendez, 2013, Thompson & Green, 2006). In the study by Ogg et al. (2013), they used the MIMIC model to compare the latent mean of engaging in academic activities with two groups of students: students with ADHD symptoms and those without ADHD. In this study, the MIMIC model is used in the context of latent factor mean differences.

In the modeling of the effect of multiple observed covariates on the latent factor, in addition to the main effects of the covariates, the interactions between the covariates should also be investigated. This interaction effect is also referred to as moderation effect. Moderation occurs when the magnitude of the effect of the independent variable on the dependent variable varies with different levels of the moderating variable (Marsh, Wen, & Hau, 2006). If the effect of one covariate on the latent factor is moderated by another covariate, the regression effect of one covariate on the latent factor depends on the level of the other covariate. In this study, moderation and interaction are treated as synonymous, and exchangeable. Compared to research on the interaction effect in multiple regression with observed independent and dependent variables, there has been less research on the interaction effect in latent factor models. Smeden and Hessen (2013) conducted a study comparing the performance of the log likelihood ratio test and the Wilks's lambda statistic in detecting the significance of the interactions in a two-way multigroup common factor model. In their study, the authors didn't mention the term of MIMIC, but in fact, the two-way interaction effect is the interaction effect between the covariates in the MIMIC model as described above.

Just as the MIMIC model is gaining popularity in educational and psychological research (Finch & French, 2011; Thompson & Green, 2006), so is the presence of hierarchical or nested data in educational and other social sciences; for example, students are nested within classrooms which are nested within schools. Multilevel modeling accounts for the fact that individuals belonging to the same group share more similarities with each other than with individuals from other groups. Statistical analyses that do not account for the nested data structure can yield biased parameter estimates and associated standard errors for a variety of statistical models, including latent variable models (Muthén, 1994). The primary problem of employing single level statistical models using nested data is the violation of the assumption of independence of individuals in the same group (Raudenbush & Bryk, 2002). Previous studies showed that models that analyzed hierarchically structured data without accounting for the multilevel nature yielded negatively biased standard error estimates, which consequently resulted in an inflation of the Type I error rates in the context of multiple regression (Snijders & Bosker, 1999). Under the framework of multilevel structural equation modeling (SEM), Muthén and Satorra (1995) conducted one of the earliest simulations to compare the performance of multilevel CFA and the ordinary CFA that did not account for the multilevel data structure. Their findings suggested that the ordinary CFA yielded negatively biased standard errors for factor loadings. Kim, Kwok, and Yoon (2012) investigated factorial invariance testing in multilevel MG-CFA, and their findings suggested that ignoring nested data structure resulted in inflated type I error rates when the grouping variable was at the between level. Similar problems, that is, inflated type I error rates and underestimated standard errors, were observed in multilevel MIMIC models when the nested data structure was not taken into account and the covariate was at the between level (Finch & French, 2011).

For multiple group analysis under the context of SEM, multilevel MIMIC models have also been gaining popularity in educational and psychological research (Davidov, Dülmer, Schlüter, Schmidt, & Meuleman, 2012; Jak, Oort, & Dolan, 2014; Kim & Cao, 2015; Kim, Yoon, Wen, & Kwok, 2015). Given the prevalence of SEM in social and educational sciences, combined with the frequency of the presence of hierarchical data structure, it is critical for the researchers to employ the right statistical methods for modeling hierarchical data in the context of latent variable models. Compared to multilevel MG-CFA, there have been fewer studies examining the performance of the multilevel MIMIC models. One of these studies was conducted by Finch and French (2011). They compared the performance of multilevel MIMIC models and standard MIMIC models that did not take into account the nested data with one covariate occurring at the between level and the within level, respectively. Their results showed that the multilevel MIMIC models outperformed the standard MIMIC models in terms of controlling type I error rates and estimating standard error of the parameter estimate when the observed covariate was at the between level; however, the performance of multilevel MIMIC models and standard MIMIC models was comparable when the covariate was at the within level. It was shown in this study that the nested structure of data should be taken into account in MIMIC models when data are hierarchical. However, in their study, only one covariate at the between level or at the within level was considered. Although some studies have mentioned the flexibility of the MIMIC model in modeling multiple covariates (Cheng, Shao, & Lathrop, 2016; Muthén, 1989; Thompson & Green, 2006), there have been few studies to date that examined the performance of MIMIC models in estimating multiple covariates effects as well as their interactions on the latent factor in the multilevel SEM context.

The intent of this research is to examine the performance of multilevel MIMIC models in estimating the covariates interaction effect under a set of manipulated conditions. In the situation of MIMIC models with a moderating covariate, the effect of one covariate on the latent variable depends on the value or the level of another covariate as presented in Figure 1 and Figure 2. In Figure 1, the two covariates are at the between level, and the effect of covariate 2 on the latent variable depends on the value of covariate 1. Note that the six indicators are represented by circles instead of squares, because the indicators in the between level are considered as latent random intercepts and they are the cluster mean of the observed indicators at the within level. Research scenarios with two covariates at the between level include research on the impact of some contextual variables on latent factors. One example of this type of research scenario is whether the impact of control or treatment status on a construct is identical for public and private schools. In this example, the school control or treatment status, and school type (public or private) are both between level covariates.

In Figure 2, the two covariates are at the within level, and similarly, the effect of covariate 2 on the within level latent variable depends on the value of covariate 1. Example of research scenarios in Figure 2 includes the main effects and interaction effect of two within level covariates (e.g. gender, free lunch status) on a latent construct. In Figure 3, one of the covariates is at the between level and the other one at the within level. The regression effect of the within level covariate C_W predicting the latent factor F_W is a random slope at the between level and this random slope can be explained by the between level covariate C_B. If the effect of C_B on the random slope is significant, there exists the cross level covariates interaction, indicating that the effect of the within level covariate on the latent factor depends on the value of the between level covariate. One possible example of research scenarios in Figure 3 is that the effect of gender on a latent variable differs for students from different types of school (public or private). To the best of my knowledge, there has been a paucity of research in the performance of multilevel MIMIC in estimating the covariates interaction effect. This study aims to explore the performance of multilevel MIMIC models in detecting a categorical covariate moderation effect and the impact of model misspecification in multilevel MIMIC models. The dissertation will comprise two studies. Study One focuses on the parameter estimate of the covariates interaction in multilevel MIMIC models. In Study One, multilevel MIMIC models have two dichotomous covariates and their interaction in addition to the measurement component of CFA. There are 3 scenarios of the two covariates: (a) they both are at the between level (e.g., public vs. private school, urban vs. rural school); (b) they both are at the within level (e. g., gender, categorical social economic status (SES)); and (c) one is at the between level and the other one at the within level (e.g., public vs. private school, male vs. female). Therefore, the covariates interaction involves the between level covariates interaction, the within level covariates interaction, and the cross level covariates interaction.

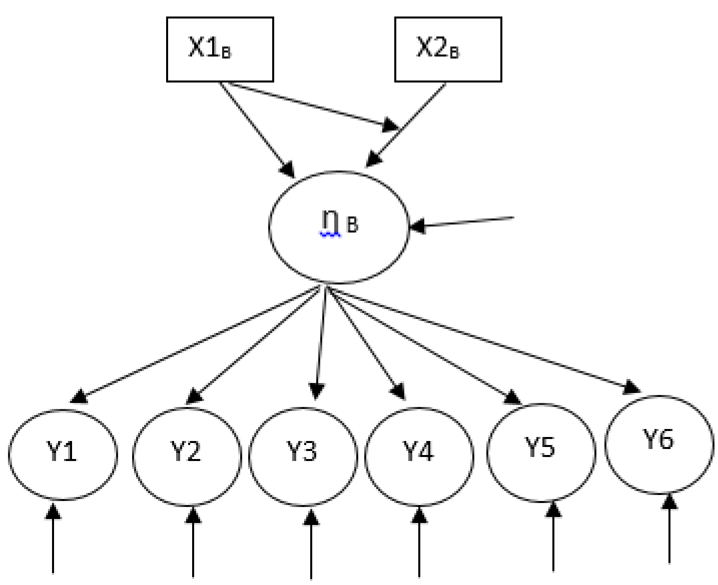

Figure 1. The between level MIMIC model with two categorical covariates and the moderation effect

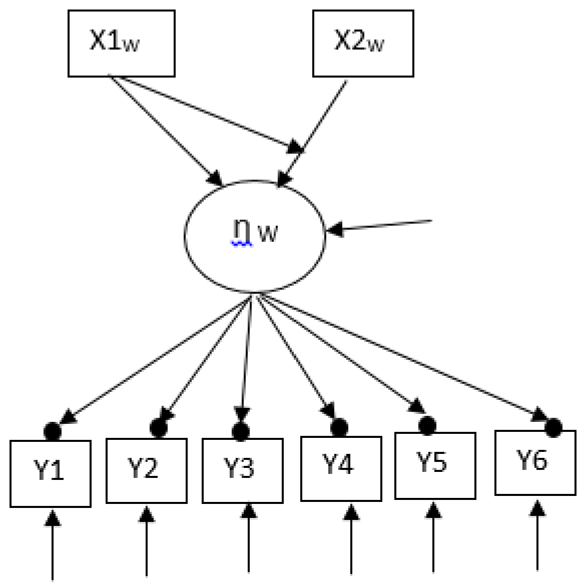

Figure 2. The within level MIMIC model with two categorical covariates and the moderation effect

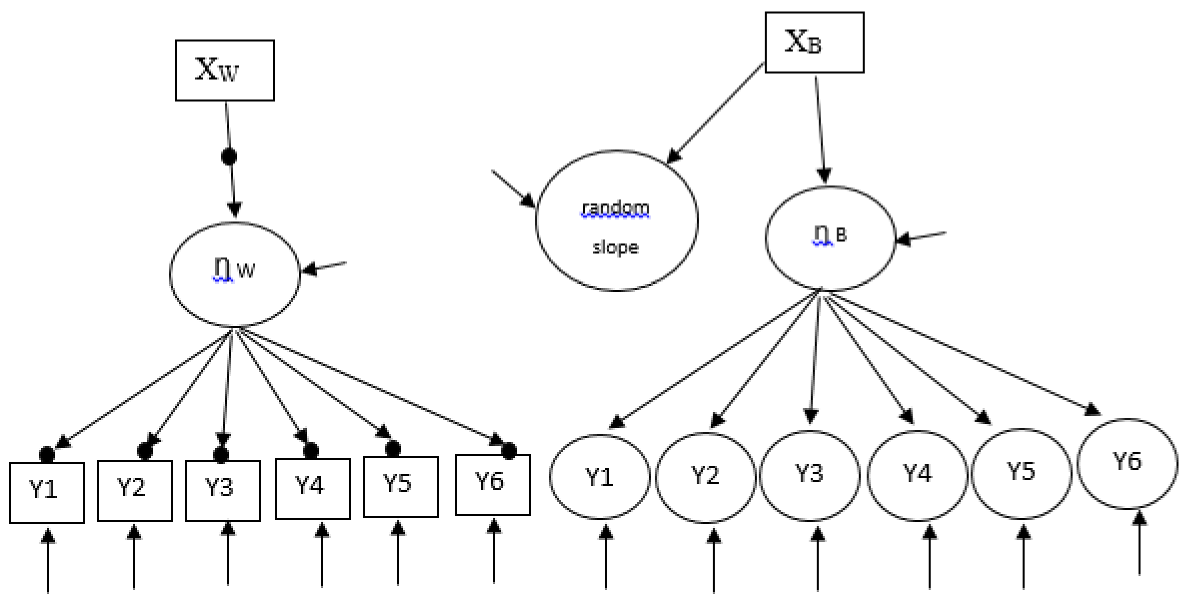

Figure 3. The cross level covariates interaction in the MIMIC model with one categorical covariate at the between level and one categorical covariate at the within level

Study One

Study One will examine the performance of multilevel MIMIC models in estimating the covariates interaction parameter when the two covariates are both at the between level, both at the within level, and one covariate at the between and the other one at the within level, respectively, assuming strict measurement invariance across levels. The presence of two categorical covariates with two levels each creates four groups. For example, there are two dichotomous covariates at the within level: gender (male, female), and SES (high, low).Then, four groups can be formed using the two covariates, i.e., male high SES group, female high SES group, male low SES group, and female low SES group. If there is an interaction effect present, the factor mean difference between male and female varies depending on the SES high or low status. Similarly, the factor mean difference between SES high and low groups hinges on gender. When one of the covariates is at the between level and the other one is at the within level, the cross level interaction exists when the effect of the within level covariate on the factor differs significantly for the two levels of the between level covariate. For example, the effect of gender, which is a within level covariate, on the latent factor is different for school type (e.g., public vs. private), which is a between level covariate. The research questions include:

1. What is the performance of multilevel MIMIC models in testing the covariates interaction effect (between level covariates interaction, within level covariates interaction, and cross level covariates interaction)? The performance will be examined through Type I error control, statistical power, bias and relative bias

of parameter estimate of covariates interaction effect, and root mean square error (RMSE).

2. What impact do the designed factors (i.e., intraclass correlations (ICCs), cluster number, and cluster size) have on the simulation outcomes of testing and estimating the covariates interaction effect?

Study Two

Study Two focuses on the impact of model misspecification on parameter estimates in multilevel MIMIC models. To be more specific, the effect of ignoring covariates interaction on the estimate of other parameters when the interaction effect is present in the population data. The impact of misspecification in SEM on parameter estimates has been studied extensively by researchers in educational and psychological research (Farley & Reddy, 1987; Gallini, 1983; Kaplan, 1988; Sörbom, 1975), and so has been the sensitivity of model fit indices to model misspecification (e.g., Fan, Thompson, &Wang, 1999; Fan & Wang, 1998; Gerbing & Anderson, 1993; Hu & Bentler, 1988, 1989; Marsh, Balla, & Hau, 1996). However, related research on the effect of model misspecification in multilevel models, particularly multilevel SEM models has been much less. In these few previous studies about model misspecification impact on multilevel modeling parameter estimate, bias was observed. For example, in the study about assuming constant level-1 residual variances when they are in fact varying in each subpopulation in growth mixture models (Enders & Tofighi, 2008), all parameter estimates exhibited bias to some degree. In an investigation into the impact of misspecifying the within level error structure in a two level growth models, results demonstrated that the misspecification produced biased estimates of variance parameters but unbiased estimates of fixed effects (Ferron, Dailey, & Yi, 2002). The model misspecification of multilevel MIMIC models has not been studied in literature. Model misspecification can take various forms. In this study, model misspecification will focus on the omission of the covariates interaction effect in multilevel MIMIC models. The impact of omitting covariates interaction effect at the within or between level on the parameter estimates at both levels in multilevel MIMIC models has not been examined in literature. The population model is the one in Study One with covariates interaction effect, but the specified model omitted the covariates interaction effect. The research questions of Study Two include:

1. What is the impact of omitting covariates interaction at the between level, the within level, and the cross level on parameter estimation at the between and the within levels, including factor mean, factor residual variance, indicator error variance parameters and the main effects of both covariates? The performance will be examined via parameter estimation bias and model fit indices (CFI, RMSEA, SRMR-B (between), SRMR-W (within), AIC, BIC, and SaBIC).
2. What are the effects of manipulated factors on the simulation outcomes?

Significance of the Dissertation

On the basis of the previous research in multilevel SEM, studies conducted in this dissertation are expected to make unique contributions to the literature of multilevel MIMIC models and the moderation effect in the multilevel model context. First, although Finch and French (2011) examined the path linking the covariate and the latent factor by comparing the parameter estimation of multilevel MIMIC models and ordinary MIMIC models that didn't account for the hierarchical data structure, there has been no study investigating the estimation of the effect of multiple covariates and their interaction effects. They called for further studies to include covariates at both the within level and the between level to have a more comprehensive understanding of MIMIC models with more complex data structures. Second, to the best of my knowledge, the cross level covariates interaction effect in multilevel SEM has not been examined. The performance of the cross level covariates interaction in multilevel MIMIC models may shed light on examining the effect of some individual and contextual characteristics on latent factors for practitioners. For example, the effect of gender on engagement in academic activities is different for public school and private school students. In this case, gender is a within level covariate, and its effect on the latent factor depends on the school type, which is a between level covariate. If the multilevel MIMIC performs well in detecting the cross level interaction effect, researchers and practitioners can be confident in their conclusion about the inferences about the presence or absence of the effect of a certain covariates interaction on the latent factor. Third, the impact of ignoring the covariates interaction on other parameters estimation in multilevel MIMIC models has not been examined yet. The simulation result can provide some insight into the credibility of parameter estimation in multilevel MIMIC models when the covariates interaction effect is neglected, which happened very often in multilevel SEM.

Chapter Two: Literature Review

The literature review part consists of four sections: introduction of the MIMIC model and its properties, MIMIC models with multilevel data, covariates interaction, and the impact of model misspecification in multilevel SEM.

Multiple-Indicator Multiple-Cause (MIMIC) Models

Group mean comparison is very common practice in social and behavioral science. When the outcome variable is observed, for example, height, and if there are only two groups involved in the comparison, a t-test is frequently used to compare the group mean. In the case of more than two groups to be compared, analysis of variance (ANOVA) is the method that can be used to test if there is a statistically significant difference in the outcome variable between the groups. Sometimes the outcome variable is a latent variable, for example, the construct of engagement in academic activities. It is usually measured by a set of observed items that are reflective of the latent variable. In the past few decades there has been increasing interest in testing latent group mean differences in educational and social sciences. Various SEM models can be used to test latent mean differences, and the most commonly employed models are MG-CFA and the MIMIC model (Muthén, 1989). In MG-CFA, a CFA measurement model is specified for each group using the grouping variable, to test factor mean differences across observed groups (e.g., gender, and ethnicity). In MIMIC models, a single CFA is specified for the total sample by incorporating a dummy coded covariate that has a direct effect on the latent factor to test latent mean differences (Jöreskog & Goldberger, 1975; Muthén, 1989). The advantages of the MIMIC models over MG-CFA have been already elaborated in the introduction part and will not be repeated here.

The MIMIC method is very flexible in modeling multiple covariates (i.e. more than two covariates) and their interactions (Cheng, Shao, & Lathrop, 2016; Fleishman, Spector, & Altman, 2002). It has been employed in various contexts, including as a way employed for testing group mean differences in latent variables (e.g., Allua, Stapleton, & Beretvas, 2008; Finch & French, 2011; Hancock, 2001, Thompson & Green, 2006), and a method for detecting differential item functioning (DIF) in educational testing and psychological measurement (Chun, Stark, Kim, & Chernyshenko, 2016; Finch, 2005; Woods, 2009; Woods & Grimm, 2011;). When the covariate(X) is categorical, the MIMIC model employs it to explain group differences. The MIMIC model describing the linear relation between observed variables and latent factors consists of two parts: a measurement model and a structural regression model:

$$y_i = \nu + \Lambda\eta_i + \varepsilon_i, \qquad (1\text{-}1)$$

$$\eta_i = \Gamma X_i + \zeta_i, \qquad (1\text{-}2)$$

$$\varepsilon \sim N(0, \Theta_\varepsilon), \qquad (1\text{-}3)$$

$$\zeta \sim N(0, \Psi), \qquad (1\text{-}4)$$

where, for individual i, y is a vector of observed indicator variables, ν is a vector of intercepts, Λ is a matrix of factor loadings, η is a vector of common factors, and ε is a vector of residuals; Γ represents a matrix of pattern coefficients estimating the effect of the covariates (X) on latent factors (η) and ζ is a vector of disturbance. Residuals (ε) and disturbances (ζ) are normally distributed with a mean of zero and a variance of Θ_ε and Ψ, respectively. Moreover, residuals (ε) and disturbances (ζ) are uncorrelated. X can be continuous or categorical, observed or even a latent variable. When the covariate X is a dummy-coded dichotomous variable that represents two groups, the coefficient Γ indicates the latent factor mean difference of the two groups. Note that if there are multiple covariates that have regression effects on the latent factor in MIMIC models, these coefficients linking the covariates and the latent variable can be estimated simultaneously. Similar to multiple regression, the interaction effect of two or more covariates can be included in the model by creating a product term of the covariates.

As mentioned above, residuals are normally distributed with a mean of zero and uncorrelated with latent factors: E (ε) = 0 and Cov (η,ε) = 0. The population variance covariance matrix Σ is written as:

$$\Sigma = \Lambda\Phi\Lambda' + \Theta_\varepsilon, \tag{1-5}$$

where Φ and Θ_ε are variance covariance matrices for common factors and residuals, respectively.

Multilevel MIMIC Models

Hierarchical or nested data are very common in educational and psychological research as well as in other social and behavioral sciences. Examples include students nested within classrooms, and patients nested within doctors or clinics. Some sampling methods also result in nested data structure, for example, cluster sampling or multistage sampling. In the previous section about MIMIC models, all individuals are assumed to be sampled using simple random sampling, and the observations are independent from each other. As mentioned above, in educational research, it is very common that individuals are nested in a higher unit, or a cluster. Individuals belonging to the same cluster are more similar to each other than to those from a different cluster because of the environmental context. Given the ubiquity of multilevel data in research, and the popularity of MIMIC models under the framework of SEM, it is important for researchers to be familiar with the detailed modeling procedures and techniques in multilevel MIMIC models.

In multilevel MIMIC, a subscript *j* is included to indicate a certain parameter of interest can vary across clusters. Equation 1-1 is decomposed into the between level and within level models and rewritten in Equation 2-1:

$$y_{ij} = \nu_{\mathrm{B}} + \Lambda_{\mathrm{B}}\eta_{\mathrm{B}j} + \Lambda_{\mathrm{W}}\eta_{\mathrm{W}ij} + \varepsilon_{\mathrm{B}j} + \varepsilon_{\mathrm{W}ij}, \tag{2-1}$$

$$\varepsilon_{\mathrm{W}ij} \sim N(0, \Theta_{\varepsilon\mathrm{W}}) \tag{2-2}$$

$$\varepsilon_{\mathrm{B}j} \sim N(0, \Theta_{\varepsilon\mathrm{B}}) \tag{2-3}$$

where y_{ij} is the observed outcome y for individual *i* in cluster *j*; ν_{B} is the between level intercept; Λ_{B} is the between level factor loading; $\eta_{\mathrm{B}j}$ is the between level cluster factor; Λ_{w} is the within level factor loading, and it is not random; $\eta_{\mathrm{w}ij}$ is the within level latent factor; $\varepsilon_{\mathrm{B}j}$ and $\varepsilon_{\mathrm{W}ij}$ are the between level and the within level residuals respectively. Note that the within level intercept (ν_{w}) is fixed at zero because individual scores are simply the sum of the group mean and individuals' deviation from the group mean (Heck & Thomas, 2000).

When there is a covariate only at the between level, a between level grouping covariate (X_{Bj}) is included in the regression model with a direct effect on the latent factor $\eta_{\mathrm{B}j}$ as shown in Equation 2-4:

$$\eta_{\mathrm{B}j} = \Gamma_{\mathrm{B}} X_{\mathrm{B}j} + \zeta_{\mathrm{B}j}, \tag{2-4}$$

$$\zeta_{\mathrm{B}j} \sim N(0, \Psi_{\mathrm{B}}), \tag{2-5}$$

where Γ_{B} represents the between level group differences in latent variable mean if $X_{\mathrm{B}j}$ is a grouping variable. Similarly, the total covariance matrix in Equation 1-5 is divided into the within level and the between level components:

$$\Sigma_{\mathrm{T}} = \Sigma_{\mathrm{B}} + \Sigma_{\mathrm{W}}, \tag{2-6}$$

where subscripts W and B denote the within and the between respectively; Σ_{T} represents the total covariance matrix; Σ_{B} is the variability across the clusters; and Σ_{W} corresponds to the variability of individual deviations from the cluster mean.

When the covariate is at the between level, Σ_{W} and Σ_{B} can be written as follows, respectively:

$$\Sigma_{\mathrm{W}} = \Lambda_{\mathrm{W}}\Phi_{\mathrm{W}}\Lambda_{\mathrm{W}}{}' + \Theta_{\varepsilon\mathrm{W}}, \tag{2-7}$$

$$\Sigma_{\mathrm{B}} = \Lambda_{\mathrm{B}}(\Gamma_{\mathrm{B}}\Phi_{\mathrm{B}}\Gamma_{\mathrm{B}}' + \Psi_{\mathrm{B}})\Lambda_{\mathrm{B}}{}' + \Theta_{\varepsilon\mathrm{B}}, \tag{2-8}$$

On the other hand, when the covariate is only at the within level, the latent variable at the within level ($\eta_{\mathrm{W}ij}$) is explained by the within level grouping covariate:

$$\eta_{\mathrm{Wij}} = \Gamma_{\mathrm{W}}X_{\mathrm{W}ij} + \zeta_{\mathrm{W}ij}, \tag{2-9}$$

$$\Sigma_{\mathrm{W}} = \Lambda_{\mathrm{W}}(\Gamma_{\mathrm{W}}\Phi_{\mathrm{W}}\Gamma_{\mathrm{W}}' + \Psi_{\mathrm{W}})\Lambda_{\mathrm{W}}{}' + \Theta_{\varepsilon\mathrm{W}}, \tag{2-10}$$

$$\Sigma_{\mathrm{B}} = \Lambda_{\mathrm{B}}\Phi_{\mathrm{B}}\Lambda_{\mathrm{B}}{}' + \Theta_{\varepsilon\mathrm{B}} \tag{2-11}$$

$$\zeta_{\mathrm{Wij}} \sim N(0, \Psi_{\mathrm{W}}). \tag{2-12}$$

In this case, Γ_{W} represents the coefficient of the latent mean difference between groups at the within level, and it is a fixed effect.

In the two scenarios described above, the covariate is either at the between level or the within level. However, sometimes covariates of both the between level (e.g., public vs. private school type) and the within level (e.g., gender) can be introduced into the model. In this case, the effect of the between level covariate X_{Bj} on the between level latent factor and the effect of the within level covariate X_{Wij} on the within level latent factor are specified in the model, that is, the Equation 2-4 and Equation 2-9 are in the model, and correspondingly, the between level covariance matrix and the within level covariance matrix are listed in Equation 2-8 and Equation 2-10 respectively. In addition to its flexibility in modeling different types of covariates, for example, categorical, continuous, and even latent covariate, the multilevel MIMIC model has the advantage of incorporating covariates at different levels simultaneously. This feature is not easy with multiple group analysis.

Measurement Invariance Assumption in MIMIC Models

Unlike MG-CFA, which specifies a separate model for each group separately and tests measurement parameter equalities across groups, for example, factor loadings, item intercepts, and item residual variances, the MIMIC model specifies only one model for the total sample from all the groups. The MIMIC model implicitly assumes that the same measurement model holds in all the groups. To be more specific, factor loadings, item intercepts, residual variances, and factor variance are invariant for all the groups; that is, all sources of covariation among observed variables are assumed to be equal across groups (Hancock, Lawrence, & Nevit, 2000; Muthén, 1989). In a recent study by Kim and Cao (2015), multilevel MIMIC models showed robustness to residual variances noninvariance in terms of estimating factor mean difference. The performance of ML MIMIC in testing latent mean differences across groups was comparable to that of multiple group multilevel CFA (MG ML CFA), which does not require stringent measurement invariance assumptions, when the assumption of homogeneous factor disturbance variances or residual variances was violated across the two groups. However, strong measurement invariance, that is, identical factor loadings and intercepts across groups, is a prerequisite to compare latent factor mean across groups (Meredith, 1993).

Covariates Interaction

Instances in which the effect of one predictor on the dependent outcome are hypothesized to depend on the value of another predictor are prevalent in substantive theories within psychology, education and other social sciences (Bauer & Curran, 2005).There are different types of interactions that are of interest to researchers, and the most widely known covariates interaction is in the regular multiple linear regression context. The most common case concerns two continuous observed variables interacting in their influence on a dependent variable (see Aiken & West, 1991). With respect to modeling, the interaction effect is handled by including a product term of the two variables to predict the dependent variable in multiple regression. When the interaction effect of the covariates is statistically significant, main effects should be interpreted with caution. In multiple regression the relationship between the outcome variable y and the two predictors can be written in Equation 3-1:

$$y = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_1 x_2 + \varepsilon, \quad (3\text{-}1)$$

where γ_0 represents the intercept; γ_1 and γ_2 correspond to the main effects, or the first-order effects (Marsh, Wen, & Hau, 2004) of predictors $x_1$and x_2; γ_3 is the interaction effect of the two predictors; the residual ε represents the difference between observed and predicted y value. Apart from being both continuous, one of the two predictors in Equation 3-1 can be continuous and the other one can be categorical. The categorical predictor can be dummy coded in linear regression.

When the two predictors are both categorical with a relatively small number of levels, the interaction effect can be estimated very conveniently with traditional analysis of variance (ANOVA).

Covariates Interaction in Multilevel Data

The multiple regression model in Equation 3-1 belongs to the fixed effect regression models. However, in hierarchical or nested data structure, the same model above can be a random effect regression model, with coefficients that can vary across clusters. An interaction effect can be present in a variety of ways in multilevel multiple regression models. When the two predictors are both at the within level, the interaction occurs at the within level. On the other hand, when they are at the between level, the interaction is modeled at the between level. Moreover, cross level interaction arises if the random effect of a within level predictor is predicted by a between level predictor. For simplicity, the level-1 model with a single predictor can be written as:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \varepsilon_{ij}, \tag{3-2}$$

where y_{ij} is the dependent variable for individual *i* in cluster *j*, x_{1ij} is the predictor, β_{0j} and β_{1j} are the intercept and the regression coefficient of y on x_1 within cluster j, respectively, and ε_{ij} is the residual.

If the intercept and slope coefficients vary randomly across clusters, they are considered as random effects that can be predicted by a level-2 predictor (a between level variable). Again for simplicity, there is only one level-2 predictor and the level-2 model can be written as:

$$\beta_{oj} = \gamma_{00} + \gamma_{01} w_{1j} + u_{0j}, \tag{3-3}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} w_{1j} + u_{1j}, \tag{3-4}$$

where w_{1j} is the level-2 predictor for cluster j, γ_{00} and γ_{10} are the fixed intercepts of β_{oj} and β_{1j} on w_{1j}, $\gamma_{01}$and γ_{11} correspond to the coefficients of β_{oj} and β_{1j} predicted by w_{1j}, and u_{0j} and u_{1j} represent the residual of the level-1intercept and slope coefficient predicted by w_{1j}. The combined equation is as follows when β_{oj} and β_{1j} are substituted using Equations 3-3, and 3-4:

$$y_{ij} = \gamma_{00} + \gamma_{01} w_{1j} + \gamma_{10} x_{1ij} + \gamma_{11} x_{1ij} w_{1j} + u_{oj} + u_{1j} x_{1ij} + u_{1j}. \tag{3-5}$$

In the reduced form equation, γ_{11} reflects the interaction effect of the level-1 predictor x_{1ij} and the level-2 predictor w_{1j} (Raudenbush & Bryk, 2002). The approach for modeling the interaction in linear multiple regression, be it single level or multilevel, has already been established and is simple to implement in applied research. Researchers have emphasized on the importance of estimating interaction effect using theoretical and substantive implications (Marsh, Wen, & Hau, 2004).

Covariates Interaction Involving Latent Variables

The interaction effect discussed in the previous section involves only manifest or observed variables, and the modeling of the interaction is relatively simple and straightforward. However, educational and psychological research is replete with latent constructs that are measured by multiple indicators. The modeling of interaction effect involving latent variables, for example, the interaction between a latent variable and an observed variable, the interaction between latent variables, falls into the framework of SEM. Such models have advantages in that they allow us to estimate linear model coefficients while controlling for measurement error (Kenny & Judd, 1984).

Muthén and Asparouhov (2003) summarized different cases of interaction effect between latent variables and observed variables and their corresponding modeling approaches in an Mplus web note. The first case concerns the interaction between a latent continuous variable and an observed categorical variable, and it can be handled by traditional SEM using multiple group analysis. The categorical variable can be treated as groups, and the regression coefficient of the latent continuous variable predicting the dependent variable can vary across different groups. One example of this case can be the regression parameter of one latent factor predicting another latent factor varies for control and treatment groups. The second case is the interaction of a latent categorical variable known as latent classes and an observed variable. The mixture modeling under the framework of SEM can handle this type of interaction. One example of this type of interaction is the regression parameter of the observed variable on the dependent variable varies depending on the latent classes. The third case involves the interaction of a latent continuous variable and an observed continuous variable, which cannot be handled by traditional SEM. The last case is about the interaction between two latent continuous variables, and a common example of this case is the interaction of two exogenous latent constructs that are measured by multiple indicators. Interaction between two latent constructs is very common in applied psychological and other social sciences; however, many applied researchers didn't use the right modeling approach to estimate the interaction effect because of the complexity of specifying and estimating the model (Marsh, Wen, & Hau, 2006).

Because the last case is the most complicated situation in modeling interaction effects in SEM, more detail is provided here. Kenny and Judd (1984) started the research on the interaction between two latent continuous variables, and they hypothesized the interaction between two exogenous latent variables and used cross-products to represent the latent interaction effect. The specification of this SEM model in which two exogenous variables predict the endogenous latent variables and the two latent predictors interact with one another can be written as:

$$\eta = \alpha + \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \zeta, \quad (3\text{-}6)$$

where η is a latent endogenous variable, ξ_1 and ξ_2 refer to the two latent exogenous variables, $\xi_1\xi_2$ represents the latent interaction term, α is the regression intercept, γ_1 and γ_2 correspond to the two first-order regression coefficients, γ_3 is the interaction effect of the two latent continuous predictors, and ζ denotes the latent regression residual. Equation 3-6 looks very similar to multiple regression, but it involves two latent predictors that are measured by multiple indicators. As to how to construct the cross-product interaction using the measured indicators, and how to deal with the nonlinearity and non-normality using different specification approaches have evolved during the past few decades. For more specifics, the research done by Marsh, Wen, and Hau (2004), and Harring, Weiss, and Li (2015) provided very elaborative and useful summary.

Covariates Interactions in Multilevel MIMIC Models

In this study, the two covariates in MIMIC models are observed categorical variables, as in multiple regression; however, the dependent variable is a latent factor instead of a manifest variable. When the two covariates are at the same level (the within level or the between level), the interaction of the two covariates is similar to but still different from multiple regression, because this model is under the framework of SEM which models the hypothesized pattern of linear relations between a set of observed variables and a latent variable. When there are two grouping covariates ($X1_{Bj}$ and $X2_{Bj}$) at the between level, and there exists interaction effect of the two covariates, the regression model in multilevel MIMIC models with paths from the covariates to the latent factor in Equation 2-4 can be written as:

$$\eta_{\mathrm{B}j} = \gamma_{1\mathrm{B}} X1_{\mathrm{B}j} + \gamma_{2\mathrm{B}} X2_{\mathrm{B}j} + \gamma_{3\mathrm{B}} X1_{\mathrm{B}j} * X2_{\mathrm{B}j} + \zeta_{\mathrm{B}j}, \qquad (4\text{-}1)$$

where $\gamma_{1\mathrm{B}}$ is the main effect of covariate $X1_{\mathrm{B}j}$, $\gamma_{2\mathrm{B}}$ is the main effect of covariate $X2_{\mathrm{B}j}$, and $\gamma_{3\mathrm{B}}$ is the interaction effect of the two covariates above. This model is displayed in Figure 1. The parameter of interest is the covariates interaction effect, $\gamma_{3\mathrm{B}}$. Similarly, when there are two grouping covariates ($X1_{\mathrm{Wij}}$ and $X2_{\mathrm{Wij}}$) at the within level, the latent variable at the within level ($\eta_{\mathrm{W}ij}$) will be explained by the within level grouping covariates and their interaction. Equation 2-9 can be written as:

$$\eta_{\mathrm{W}ij} = \gamma_{1\mathrm{w}} X1_{\mathrm{W}ij} + \gamma_{2\mathrm{w}} X2_{\mathrm{Wij}} + \gamma_{3\mathrm{W}} X1_{\mathrm{W}ij} * X2_{\mathrm{W}ij} + \zeta_{\mathrm{W}ij} \qquad (4\text{-}2)$$

In this case, $\gamma_{1\mathrm{W}}$ is the main effect of the covariate $X1_{\mathrm{W}ij}$, $\gamma_{2\mathrm{W}}$ is the main effect of covariate $X2_{\mathrm{W}j}$, and $\gamma_{3\mathrm{W}}$ is the interaction effect of the two covariates above.

On the other hand, when two covariates are at different levels (i.e., one is at the within level and the other one at the between level), the interaction effect of the two covariates falls into the first case described above. To be more specific, the interaction is constructed by a latent continuous variable and an observed categorical variable. This is because the cross level interaction indicates that the random regression coefficient of the within level covariate on the within level latent variable (this regression slope is a latent continuous variable) is moderated by the between level categorical variable as shown in Figure 3. An empirical example of cross level covariates interaction in the educational context is that the effect of gender (male vs. female) on a construct is moderated by school type (public vs. private). In this example, gender is the within level covariate, and school type is the between level covariate. The regression slope of gender predicting the construct is a random slope effect at the between level, and it is explained by the between level covariate, school type. As shown in Figure 3, the within level latent variable is regressed on the within level covariate X_{W}, and this regression slope is a random effect that varies across the clusters at the between level. The random slope effect is designated by a solid dot on the regression slope and it is written as:

$$\eta_{\mathrm{W}ij} = \gamma_{\mathrm{Wj}} X_{\mathrm{W}ij} + \zeta_{\mathrm{W}ij} \,, \tag{4-3}$$

where γ_{Wj} is the regression slope of covariate $X_{\mathrm{W}ij}$ predicting the within level latent variable, $\eta_{\mathrm{W}ij}$; $\zeta_{\mathrm{W}ij}$ refers to the residual of the latent factor, $\eta_{\mathrm{W}ij}$. At the between level, the between level latent factor is regressed on the between level covariate, X_{B}. X_{B} also explains the random slope effect at the between level. The equations of the between level latent variable and the random slope regressed on the between level covariate, X_{B} can be written as:

$$\eta_{\mathrm{B}j} = \gamma_{\mathrm{B}} X_{\mathrm{B}j} + \zeta_{\mathrm{B}j} \,, \tag{4-4}$$

$$\gamma_{\mathrm{Wj}} = \gamma_{10} + \gamma_{\mathrm{C}} X_{\mathrm{B}j} + \zeta_{\gamma \mathrm{W}j} \,, \tag{4-5}$$

where γ_{B} indicates the between level regression slope of the between level latent variable $\eta_{\mathrm{B}j}$ on the between level covariate, $X_{\mathrm{B}j}$; $\zeta_{\mathrm{B}j}$ refers to the residual of $\eta_{\mathrm{B}j}$ that cannot be explained by $X_{\mathrm{B}j}$. γ_{10} and γ_{C} represents the regression intercept and slope of random slope, γ_{Wj} on $X_{\mathrm{B}j}$, respectively. $\zeta_{\gamma \mathrm{W}j}$ is the residual variance, the proportion of variance in the random slope that cannot be explained by the between level covariate, $X_{\mathrm{B}j}$. The combined model for the within level latent variable, $\eta_{\mathrm{W}ij}$, in Equation 4-3, can be written as follows by substituting γ_{Wj} in Equation 4-5:

$$\eta_{\mathrm{W}ij} = \gamma_{10} X_{\mathrm{W}ij} + \gamma_{\mathrm{C}} X_{\mathrm{W}ij} * X_{\mathrm{B}j} + \zeta_{\gamma \mathrm{W}j} * X_{\mathrm{W}ij} + \zeta_{\mathrm{W}ij} \,, \tag{4-6}$$

where γ_{C} is the cross level covariates interaction effect, which is the focus of the study when the two covariates in MIMIC models are at the within level and the between level.

Misspecification in Multilevel Structural Equation Modeling

SEM has been a very attractive methodology for educational and psychological researches to investigate the linear relationship between a set of observed variables and latent factors, and it takes into account measurement errors. However, researchers should be cautious about model misspecification in SEM, using empirical theory to specify models.

Misspecification in Single-level SEM

The impact of misspecification in SEM on parameter estimates has been studied extensively by educational and psychological researchers (e.g., Gallini, 1983; Kaplan, 1988; Yuan, Marshall, & Bentler, 2003). Gallini (1983) utilized path analysis and CFA models to show that omitted variables cause severe parameter bias if the omitted variables are strongly related to the exogenous variable. Kaplan (1988) used CFA and structural equation models between exogenous and endogenous latent factors to illustrate the impact of misspecification on parameter estimates. The author considered misspecification of the measurement model, the specification error of the structural component (e.g., the path between exogenous and endogenous factors, and covariance among themselves), as well as misspecification of both measurement part and structural components. In this study it was found that the combination of measurement and structural misspecification produced the most biased parameter estimates (Kaplan, 1988). Bentler and Chou (1993) proposed to examine the mean absolute parameter change for all the free parameters and their findings implied that not all parameters in the model were affected when a variable was omitted in the model, and only parameters that were closely related to the misspecification would be affected. Yuan, Marshall, and Bentler (2003) also studied whether all parameters were affected by the misspecification, and they used three procedures to evaluate the impact - that is, analyzing the path, using a functional relationship, and using a statistical test. Interested readers can refer to this research for details about the three procedures. Most studies about misspecification in SEM employed CFA and path models to evaluate the impact of model misspecification on parameter bias, and no studies used MIMIC models, especially the path linking the observed covariate and the latent factor. The misspecification of the path linking the observed covariates and the latent factor is the focus of Study Two.

Sensitivity of Model Fit Indices in SEM

Closely related to the research in model misspecification in SEM is the research in the sensitivity of model fit indices to model misspecification in SEM. To evaluate how well the hypothesized models fit the data, several fit indices have been commonly used by empirical researchers in the SEM framework. Chi-square (χ^2) test examines the exact model fit, that is, whether the sample covariance matrix is identical to the model-implied covariance matrix (Bollen & Long, 1993). A significant χ^2 test indicates that the sample covariance matrix is not equal to the model implied covariance matrix, but it cannot show the magnitude of the misfit. Furthermore, χ^2 is a function of sample size, implying that a large sample size leads to a big χ^2 value (rejecting the null hypothesis) even when there is a minor difference between the model implied covariance matrix and the sample covariance matrix. Therefore, it tends to reject the null hypothesis, and the exact overall fit χ^2 test is not of major interest. Some other commonly used fit indices, for example, root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker-Lewis Index (TLI), are derived from the χ^2 statistic.

RMSEA is based on the non-centrality parameter, and its formula is:

$$\text{RMSEA}=\sqrt{\max(0,\frac{\chi 2-\text{df}}{df(N-1)})}, \tag{4-7}$$

where χ^2 is the chi-square test statistic, df represents degrees of freedom, and N denotes the total sample size. RMSEA measures the average discrepancy between observed sample covariance matrix and model implied covariance matrix per degree of freedom after taking model complexity into account (Browne & Cudeck, 1993; Steiger, 2007). A smaller value of RMSEA is associated with a better model fit, as RMSEA reflects the degree of misfit or badness of the tested model.

CFI measures the relative goodness of fit of a particular tested model compared with a baseline model in which there is no covariance between any pair of variables (Bentler, 1990, Tanaka, 1993). Its computation formula is:

$$\text{CFI}=1-\frac{max(\chi 2_{1-df1},\ 0)}{\max(\chi 2_{0-df0},\ \ \chi 2_{1-df1},\ 0)}, \tag{4-8}$$

where $\chi^2{}_1$ and df_1 represent the chi-square test statistic and degrees of freedom of the tested model, while $\chi^2{}_0$ and df_0 denote chi-square test statistic and degrees of freedom of the baseline model. CFI ranges between zero and 1 (Bentler, 1990, Tanaka, 1993), and a larger CFI indicates a better overall model fit of the tested model.

TLI penalizes for model complexity (Tucker & Lewis, 1973), and its computation formula is as follows:

$$\text{TLI}=(\frac{\chi 2_0}{df0}-\frac{\chi 2_1}{df1})/(\frac{\chi 2_0}{df0}-1), \tag{4-9}$$

Where $\chi^2{}_0$ and df0 denote chi-square test statistic and degrees of freedom of the baseline model, and $\chi^2{}_1$ and df1 represent chi-square test statistic and degrees of freedom of the tested model. Like CFI, a larger TLI suggests a better overall model fit.

Standardized root mean square residual (SRMR) is defined as the standardized discrepancy between the observed sample covariance matrix and model implied covariance matrix. It is computed as follows:

$$\text{SRMR}=\sqrt{\left\{2\sum_{k=1}^{p}\ \sum_{j=1}^{k}\ \left[\frac{skj-\sigma kj}{s_{kk}s_{jj}}\right]^2\right\}/p(p+1)}, \tag{4-10}$$

where skj is the sample covariance between any two variables k and j; σkj represents the model implied covariance between variables k and j; s_{kk} and s_{jj} denote the observed standard deviations for variables k and j; and p is the total number of variables in the model. A smaller SRMR means a better fit. In multilevel SEM, the covariance matrix can be separated into the between-level covariance matrix and the within-level covariance matrix, and SRMR can be obtained using the within-level covariance matrix (SRMR-W), and the between-level covariance matrix (SRMR-B), separately.

In the past few decades, substantial research has been done in the sensitivity of model fit indices in detecting model misspecification in SEM (Fan & Sivo, 2007; Hu & Bentler, 1998; Hsu, Kwok, Lin, & Acosta, 2015; Ryu & West, 2009) and in establishing cutoff criteria for fit indices produced in SEM models, and the most widely known research of this type was conducted by Hu and Bentler (1988, 1989). The cutoff golden rule of model fit indices recommended by Hu and Bentler (1989) (RMSEA≤0.06; CFI ≥ 0.95; TLI ≥ 0.95; SRMR ≤ 0.08) was referred to very frequently by applied researchers in the area of SEM in their reporting of model results. However, some researchers started to ask questions about the generalizability of the criterion, for example, whether or not the cutoff criterion applies to any type of models in SEM (simple models with fewer variables, or complicated models with more variables), and whether or not the fit indices are equally sensitive to different degree of misspecification (e.g., Kenny & McCoach, 2003; Fan & Sivo, 2007). Fan and Sivo (2007) used a systematic design to study the sensitivity of fit indices to model misspecification by controlling for the magnitude of misspecification and model types based on complexity as represented by number of indicators in CFA models. The results showed that absolute indices, such as Gamma Hat (GAMMA), root-mean-square error of approximation (RMSEA), and McDonald's centrality index (Mc), outperformed incremental fit indices, such as the comparative fit index (CFI) and Tucker-Lewis index (TLI); GAMMA performed better than RMSEA and Mc when the model was simple with fewer indicators. Therefore, researchers are supposed to report different types of fit indices and keep in mind that the cutoff criterion is not generally useful for any type of models.

Misspecification in Multilevel SEM

Methodology for analyzing hierarchical or multilevel data has witnessed great development, such as approach for hierarchical linear models (Raudenbush & Bryk, 2002), and multilevel structural equation models (Heck & Thomas, 2015; Mehta & Neale, 2005; Muthén, 1994). The popularity of multilevel SEM among educational and psychological researchers is due to the fact that these models allow researchers to analyze multilevel data by simultaneously examining the separate within level and between level models while accounting for measurement errors in the models (Heck & Thomas, 2015). However, compared to the research on the impact of model misspecification in traditional single level SEM, there has been a paucity of research in model misspecification in multilevel models, and even less in the multilevel SEM context. Among the infrequent research about model misspecification in multilevel data, Enders and Tofighi (2008) utilized growth mixture modeling to investigate the impact of specifying level-1 residual variance constant across latent classes when they are in fact varying in each subpopulation. Their results demonstrated that misspecification resulted in bias at the within class growth slopes and variance components, and parameter estimates were substantially less precise than those produced by a correctly specified growth mixture modeling. In an investigation into the impact of misspecifying the first level error structure in a two-level growth models, results demonstrated that the misspecification produced biased estimates of variance parameters but unbiased estimates of fixed effects (Ferron, Dailey, & Yi, 2002). The model misspecification of multilevel MIMIC model has not been discussed in literature. Model misspecification can take various forms. In this study, model misspecification will focus on the omission of the covariates interaction effect in multilevel MIMIC models. The impact of omitting covariates interaction at one level on the parameter estimates at another level as well as at the same level in multilevel MIMIC models has not been examined in literature. The intent of Study Two is to investigate the impact on parameter estimates at both levels in multilevel MIMIC models when the covariates interaction was omitted from the model but it was present in the population model. Additionally, the sensitivity of model fit indices (e.g., CFA, RMSEA, SRMR-W, SRMR-B, AIC, BIC, and SaBIC) to omitting the covariates interaction effect in multilevel MIMIC model is also examined.

Sensitivity of Model Fit Indices in Multilevel SEM

Similar to the situation of insufficient research in model misspecification impact in multilevel SEM, the sensitivity of model fit indices in multilevel SEM has not attracted as much attention as single level SEM. In some empirical research of multilevel CFA models, applied researchers still used the cutoff criteria from conventional SEM to evaluate multilevel SEM model fit. However, methodological researchers started to investigate the appropriateness of using single level model fit indices to the evaluation of multilevel SEM. Ryu and West (2009) examined the sensitivity of two fit indices (i.e., RMSEA and CFI) in detecting model misspecification in multilevel SEM. They showed that RMSEA and CFI failed to detect the misspecification at the between level model, although the two fit indices were able to detect the misspecification at the within level model. This was a great starting point for evaluating fit indices in multilevel SEM. However, Ryu and West (2009) used a single ICC level of 0.5, which was very high in educational and psychological data. Also, they included only one type of model misspecification; that is, the correlation of the two factors was specified higher than the true correlation in the population. Lastly, they examined only two fit indices (i.e., RMSEA and CFI). Hsu, Kwok, Lin, and Acosta (2015) expanded Ryu and West's (2009) simulation study by including more design factors, such as, number of clusters, cluster size, ICC, and misspecification type. Their findings showed that CFI, TLI, and RMSEA could only detect the misspecification at the within level, but not the between level. Moreover, CFI, TLI, and RMSEA were more sensitive to misspecification of pattern coefficients whereas SRMR-W was more sensitive to misspecification of factor covariance, and SRMR-B was the only fit index that was sensitive to the model misspecification in the between level model.

Model Misspecification in Interaction Effect

Among the studies about various types of model misspecification in SEM, most of them examined the CFA model, and the misspecification type was manifested as misspecified factor loadings, factor variance covariance, and residuals correlation (Fan & Sivo, 2007; Hu & Bentler, 1999). None of them examined misspecification of covariates interaction. However, in empirical research, researchers tend to model only the main effects of the covariates, ignoring the covariates interaction effect. If that is the case, the interpretation of the main effects of the covariates is misleading. Moreover, the estimation of the interaction effect between variables has important theoretical, substantive, and empirical implications in psychology and other social sciences (Marsh, Wen, & Hau, 2004), and the interaction effect is even more interesting than the main effects of the covariates (Wen, Marsh, Hau, Wu, Liu, & Morin, 2014). Little is known about the impact of omitting an existing covariates interaction effect in SEM models, especially multilevel SEM models. In Muthén and Asparouhov's (2003) summary of all types of possible interaction in SEM models, they called for Monte Carlo study to investigate the mis-estimation of parameters as a result of omitting interaction effect when the interaction effect was present in the population model.

Chapter Three: Method

Two studies were conducted using simulation. In the first study, the performance of multilevel MIMIC models in testing covariates interaction effect was examined under different simulation conditions. The two covariates were both at the between level, both at the within level, and one at the between level and the other at the within level producing cross level interaction effect. In the second study, the impact of model misspecification on other parameters in multilevel MIMIC models was investigated. To be more specific, the impact of the omission of an existing covariates interaction effect in multilevel MIMIC models was examined. The three scenarios of the two covariates and design factors were the same as in Study One.

Study One: Testing Covariates Interaction Effect in MIMIC Models with Multilevel Data

In this simulation study, two-level data with two dichotomous covariates were generated using PROC IML in SAS/IML 9.4. For both the between and the within level models, a single factor with six continuous indicators was generated. At the within level, factor loadings were all .80, so the residual variances were .36. The factor variance was set to be 1.00 at the within level. At the between level, the six indicators were regarded as latent variables (random intercepts) in the multilevel SEM. By assuming measurement invariance across levels, the six factor loadings were set to be equal at .80 at the between level, and the between level residual variances of all six indicators were set to be equal at .02, .01 and .004 for the large, medium, and small factor ICC, respectively. The between level residual variances were selected based on the following considerations. Satisfying scalar measurement invariance in two-level SEM models required to have between level residual variances of zero based on the research on measurement bias in multilevel data conducted by Jak et al. (2014). In real world data, it is very difficult to have scalar invariant construct when there are many groups (represented by clusters in this study), so a close to zero (.02, .01, and .004) but not zero was selected as the between level residual variance. The between level factor variance was varied to form different ICCs, which will be explained later.

When the two covariates were at the same level (i.e., both at the within level, or at the between level), the main path coefficients of the two covariates to the factor were not varied. In all conditions they were set to be .30 and .40 respectively, with .30 representing a moderate-sized relationship in prior research (Maas & Hox, 2005). These numbers also accorded with the values of some empirical research in applied psychology literature. Mathieu, Aguinis, Culpepper, and Chen (2012) conducted a review of studies published in the *Journal of Applied Psychology* involving tests of cross level interactions, and they reported a within level slope range between -.06 and .45 and a between level slope range between -.23 and .35. On the other hand, when one of the two covariates was at the within level and the other one at the between level, the within level main effect was set at .40, and the between level main effect was set at .50. The population parameters are summarized in Table 1. When the two covariates were at the between level, refer to Figure 1 and Equation 4-1 for the measurement and structural relations in the between-level MIMIC model. When the two covariates were at the within level, Figure 2 and Equation 4-2 described the model in the within-level MIMIC model. When one of the covariates was the within level and the other one at the between level, Figure 3 and Equation 4-6 showed the cross-level covariates interaction effect.

Table 1. Parameters in the population data generation

	Location of the two covariates		
	Within	Between	Cross
number of indicators	6	6	6
factor loadings	.8	.8	.8
within level item residual variance	.36	.36	.36
between level item residual variance	.02, .01, and .004 for L, M, and S ICC	.02, .01, and .004 for L, M, and S ICC	.02, .01, and .004 for L, M, and S ICC
within level factor variance	1	1	1
main effects of the two covariates	.3, .4	.3, .4	.4 (W), .5 (B)

Note. L=large; M=medium; S=small; W=at the within level; B=at the between level.

Data Generation

PROC IML was used to generate the simulated data meeting the design factors. The between-level measurement and the within-level measurement parts were simulated first using the population parameters in Table 1. Basically, the two-level CFA measurement part was generated first, and then the effects of the covariates on the factor mean was manipulated. To simulate the interaction effect of the two dichotomous covariates, the factor means of the four cells, which represent the four groups created by two dichotomous variables (for example, female with high SES, female with low SES, male with high SES, and male with low SES), were manipulated to reflect the covariates interaction effect as presented in data generation code in

Appendix 1. When the two covariates were at the between level, the between-level factor means of the four cells varied to create the interaction effect. To be specific, when the main effects were .30 and .40, and the covariates interaction was .30, in the first cell (X1=0, X2=0), the factor mean was 0; in the second cell (X1=0, X2=1, the factor mean was .40; in the third cell (X1=1, X2=0), the factor mean was .30; in the fourth cell (X1=1, X2=1), the factor mean was 1.0. The difference between cell Two and cell One was the main effect of X2, and the difference between cell Three and cell One was the main effect of X1. When X2=0, the effect of X1 on factor mean was .30 (Cell Three minus Cell One), whereas when X2=1, the effect of X1 on factor mean was .60 (Cell Four minus Cell Two). Thus the effect of covariate X1 on the between-level factor mean was moderated by covariate X2, and the moderation effect was .30 as presented in Figure 4. When the interaction effect was at .60, the factor mean of Cell Four changed to 1.30. When there was no interaction effect, the effect of X1 on factor mean was the same regardless of the value of X2, and the factor mean of cell Four was .70. When the two covariates were at the within level, very similar manipulation of the within-level factor means was conducted to create the within-level covariates interaction effect. The only difference was that within-level factor mean was manipulated instead of between-level factor mean. The procedure was very similar to that in the between-level interaction effect and it was not going to be repeated here.

	X2=0	X2=1
X1=0	Mean=0	Mean=0.4
X1=1	Mean=0.3	Mean=1.0

Figure 4. Marginal factor means for the between-level and within-level covariates interaction

When the random slope of the within-level factor regressed on the within-level covariate was moderated by the between-level covariate (cross-level interaction), a slightly different manipulation of the factor means of the between-level factor and the within-level factor was utilized to create the cross-level covariates interaction effect. Both the between-level factor mean and the within-level factor mean were manipulated to reflect the cross-level interaction. As shown in Figure 5 to create two main effects of .40 and .50 , and an interaction effect of .20, the within-level covariate main effect was .40 (cell Three minus Cell One), and the between-level covariate main effect was .50 (cell Two minus Cell One). For the between-level factor mean, When X_B=0, the average factor mean was .20 (the sum of cell One and cell Three divided by 2: (0+.4)/2); when X_B=1, the average factor mean was .80 (the sum of cell Two and cell Four divided by 2: (.5+1.1)/2). Thus, the between-level factor mean was .60 higher for X_B=1 than for X_B=0 as shown in Appendix 1 for the SAS code of cross-level covariates interaction. For the within-level factor mean, when X_B=0, the factor mean difference of the two within-level cells was .40; when X_B=1, the factor mean difference of the two within-level cells was .60 as shown in Appendix 1 for the cross-level interaction effect generation code. Similarly, when the covariates interaction was .40, the factor mean in cell Four was 1.3, and all the other three cells remained the same. Therefore, the between-level factor mean was .70 higher for X_B=1 than for X_B=0. For the within-level factor mean, when X_B=0, the factor mean difference of the two within-level cells was .40; when X_B=1, the factor mean difference of the two within-level cells was .80

	X_B=0	X_B=1
X_W=0	Mean=0	Mean=0.5
X_W=1	Mean=0.4	Mean=1.1

Figure 5. Marginal factor means for the cross-level covariates interaction

Simulation Conditions

To examine the performance of multilevel MIMIC models in detecting the covariates interaction effect, a set of design factors were taken into account. Simulation conditions included the location of the interaction effect with 3 levels (at the between level, at the within level, and at the cross level), number of clusters (CN), cluster size (CS), intraclass correlation (ICC), and magnitude of the covariates interaction effect.

Location of the Interaction Effect

There were three levels of the location of the interaction effect: the covariates interaction effect at the between level, the covariates interaction effect at the within level, and the cross level covariates interaction effect. When the location of the covariates interaction effect was different, some of the other design factors varied, too, which will be explained in the coming session. For the two-level random effect model, the within-level random slope was specified to be predicted by the between-level covariate (cross-level covariates interaction effect) as mentioned above. However, the random slope was not explained 100% by the between-level covariate, and the residual variance of this random slope was specified to be at .10, which is a reasonable value for the magnitudes of the cross-level interaction effect. Also, the specification of the residual variance of the random slope is very important for the model to converge.

Number of Clusters (CN)

CN was included because previous simulation studies on multilevel modeling found that CN had an impact on estimating parameters of multilevel models (Hox & Maas, 2001; Kim et al., 2012; Maas & Hox, 2005). Because CN and CS influence sample size, it is expected that they are directly related to the power to detect the covariates interaction effect in multilevel MIMIC models. When the two covariates were at the between level, three levels of numbers of clusters were simulated (40, 80, and 120). Note that two dichotomous covariates were at the between level, creating four groups. Four balanced groups were generated, with 10, 20, and 30 in each group for the three levels of number of clusters, respectively. When the two covariates are at the within level, the number of clusters are set at 20, 40, and 60, respectively. The number of clusters at the within level covariates interaction was only half of the clusters at the between level covariates interaction to control for the total sample size. The cluster size of the within level covariates interaction was different from the cluster size of the between level covariates interaction, which will be explained in the next section. When one covariate is at the between level and one at the within level, the number of clusters was set to be identical as the conditions of two covariates at the between level, that is, 40, 80, and 120 clusters.

Cluster Size (CS)

When the two covariates were at the between level, CS was examined at two levels (10 and 20). A cluster size of 10 and 20 was suggested by Hox (1998) as a typical cluster size for designing multilevel research. Multiplying CN by CS, the minimum total sample size was 400 (CN=40, CS=10), and the maximum sample size was 2400 (CN=120, CS=20). When the two dichotomous covariates were at the within level, cluster size was set to be 20 and 40, so that the 4 balanced groups had a cluster size of 5 and 10, respectively. The minimum total sample size is also 400 (CN=20, CS=20), and the maximum sample size is 2400 (CN=60, CS=40). Note that when the two covariates were at the between level, the cluster number was twice that at the within level, but the cluster size was only half of that at the within level. Thus, the total sample size was the same for the two different locations of the interaction effect. The major reason for adopting this method was because the grouping was at different levels. When one of the covariates is at the within level, and the other one at the between level, the cluster size was the same as the situation of both covariates at the between level: 10 and 20. In sum, the total sample size remained the same when the two covariates were at the between level, at the within level, and at the cross level.

Intraclass Correlation (ICC)

ICC was also manipulated, as the previous studies (e.g., Kim, Kwok, & Yoon, 2012) showed that ICC had an impact on Type I error and power of measurement invariance tests in multilevel CFA models. It may have an effect on detecting the covariates interaction effect in multilevel MIMIC models. ICC is defined as the ratio of the between factor variance over the total factor variance, which is the sum of the between factor variance and the within factor variance. As previously stated, the within-level factor variance was set to be 1, so varying ICC levels depend on different values of the between-level factor variance. In this study, the between factor variance was set to be .10, .25, and .50, resulting in three ICCs of .09, .20, and .33, respectively as the small, medium, and large ICC. These ICC levels are typical of educational and psychological data and were employed in previous simulation studies (e.g., Hox & Maas, 2001; Kim et al., 2012; Maas & Hox, 2005). The corresponding item level ICCs were .07, .15, and .25 for all six items, respectively, because all the factor loadings of the six items were the same using the item level ICC formula ICC=B/(B+W), where B was the between variance for the observed variable and W was the within variance for the observed variable. Item level ICC represented the proportion of variances in the indicators at the between level. Item level ICCs were positively related to factor ICC.

Magnitude of the Covariates Interaction

Prior research showed that the power of detecting cross level interaction in multilevel multiple regression depended mainly on the magnitude of the effect size (Mathieu et al., 2012). Thus, the magnitude of the covariates interaction was manipulated in this research. When the two covariates were both at the between level and both at the within level, the covariates interaction effect was set to 0, .30, and .60, which are consistent with the range (-.06 to .45) reported in the literature review by Mathieu et al. (2012), and in their simulation study, they manipulated the cross level interaction effect range from 0 to .75. The value of zero for the covariates interaction was used to assess Type I error rate when in the population there was no covariates interaction. And the values of .30 and .60 were used to assess power, with the former value representing a medium moderating effect and the latter representing a large moderating effect. The cross level covariates interaction effect was set to be at 0, .20, and .40, with .20 representing a small cross level covariates interaction and .40 indicating a moderate cross level interaction effect. One of the major purposes of the study is to examine the performance of multilevel MIMIC in detecting cross level covariates interaction - even a small interaction effect. Thus, only small and moderate cross level interactions were simulated to examine the performance of multilevel MIMIC models, since we expect that the performance will improve with a larger interaction effect.

In short, there were 3 levels of location of the covariates interaction effect, 3 levels of cluster number, 2 different cluster sizes, 3 varying ICC levels, and 3 levels of the interaction effect size. Thus, there were a total of 162 conditions (3×3×2×3×3). For each condition, 1000 replications were generated.

Analysis of Simulation Results and Outcome Variables

Each replication was analyzed using multilevel MIMIC models with correct specification based on the population model and estimated with the robust maximum likelihood (MLR) estimation which is the default estimation method of Mplus 7.1 (Muthén & Muthén, 1998-2012). Outcomes variables included admissible solution rates (ASR), power, and Type I error. The replications were classified as inadmissible solutions if they had one of these following scenarios: no output was produced, i.e., no parameter estimates or standard errors were obtained; the estimated parameters did not make sense, such as negative variance. The admissible rate was defined as the proportion of replications that produced admissible solutions. Power was defined as the proportion of replications in which multilevel MIMIC models detected the covariates interaction effect as statistically significant. Type I error was defined as the proportion of replications in which multilevel MIMIC models falsely flagged a statistically significant covariates interaction when there was no covariates interaction in the population model.

Standard error of estimate (SE) is used to test if the coefficient is significantly from zero by using the ratio of the parameter estimate over the standard error. Bias, relative bias, and root mean square error (RMSE) were used to evaluate the accuracy and precision of a parameter estimate. The parameter of interest in this study was the covariates interaction effect, $\gamma_{3\mathrm{B}}$, $\gamma_{3\mathrm{W}}$, and Υ_C in Equations 4-1, 4-2, and 4-3, respectively. Standard error of the parameter estimate was extracted from the Mplus output. Bias was calculated as the deviation of the parameter estimate from the population parameter (θ) on average across replications (r = 1, 2, …,R):

$$B(\theta) = R^{-1}\sum_{r=1}^{R}(\hat{\theta}_r - \theta)$$

Relative bias was computed as the ratio of the raw bias (deviation of the parameter estimate from the population parameter (θ) across replications) over the population parameter (θ) itself (Hoogland & Boomsma, 1998):

$$RB(\theta) = R^{-1}\sum_{r=1}^{R}(\hat{\theta}_r - \theta)/\,\theta$$

RMSE was defined as the average estimate error showing the variability of the estimates:

$$RMSE(\theta) = (R^{-1}\textstyle\sum_{r=1}^{R}(\hat{\theta}_r - \theta)^2)^{1/2}$$

To examine the impact of design factors (i.e., cluster number, cluster size, magnitude of interaction effect, and ICC level) on outcome variables (i.e., ASR, type I error, power, SE, bias, relative bias, and RMSE). Factorial analysis of variance (ANOVA) with generalized eta-squared (η^2) was used. η^2 was derived by dividing the Type III sum of squares of a particular design factor or interaction of design factors by the corrected total sum of squares.

Study Two: The Impact of Model Misspecification in Multilevel MIMIC Models

The purpose of Study Two was to assess the impact of model misspecification in multilevel MIMIC models on parameter estimates and the sensitivity of model fit indices. Specifically, the impact on parameter estimates when the covariates interaction was omitted in multilevel MIMIC models was examined. In Study Two, the simulated model population parameters, simulation conditions, and design factors were the same as those in Study One except that the interaction effect value of 0 was excluded for all three models (i.e., the between-level covariates interaction, the within-level covariates interaction, and the cross-level covariates interaction). This was because the focus of Study Two was on the impact of omitting the interaction effect on estimates of all the other parameters in the model. When there was no covariates interaction in the population model, omitting the covariates interaction effect cannot be considered as model misspecification. Therefore, there was a total of 108 (3×3×2×3×2) conditions in Study Two.

For each replicated data set generated by the population model with covariates interaction effect present, two models were run. One estimated model was the correct model which specified the covariates interaction as in Study One, and the other estimated model was misspecified by omitting the covariates interaction. All parameter estimates as well as model fit indices of both correct and misspecified models of the same data were saved and stacked in one file for comparison.

Outcome Variables

Firstly, model fit indices such as SEM-based CFI, RMSEA, SRMR (between), and SRMR (within) were compared between the correctly specified model and misspecified model to see if the model fit indices were sensitive to the omission of the covariates interaction in multilevel MIMIC models. The average SEM-based model fit indices across the replications and their standard deviations were reported. Also, Akaike Information Criterion (AIC; Akaike, 1974), Bayesian Information Criterion (BIC; Schwarz, 1978), and sample-size adjusted BIC (SaBIC; Sclove, 1987) were compared for the correct and misspecified models to examine if they were sensitive to the model misspecification. For AIC, BIC, and SaBIC, a smaller value indicates a better fit of the model to the data. When the AIC, BIC, and SaBIC were larger in the misspecified model than in the correct model, they selected the correct model. The proportion of replications in which the correct model was selected in model comparisons was computed for AIC, BIC, and SaBIC.

Secondly, the estimation bias of a series of parameters, including factor mean, factor variance, factor loading, intercept, residual of both levels, and the main effect of the two covariates, in the correctly specified and misspecified model was compared to examine the impact of omitting the covariates interaction effect on the estimation of the other parameters in the model. Special attention was paid to the cross-level covariates interaction model. When the cross-level covariates interaction was omitted from the model, what parameters were impacted the most?

Thirdly, the impact of the simulation design factors (i.e., cluster number, cluster size, magnitude of interaction effect, and ICC level) on the estimation would be investigated using factorial ANOVA with eta-squared (η^2) if the bias in some parameters varied as a result of the design factors in Study Two.

Chapter Four: Results

Results of Study One

The outcome variables of Study One include the admissible rates, power, Type I error, bias, relative bias, standard error of estimates, and RMSE. The results of between-level covariates interaction conditions, within-level covariates interaction conditions, and cross-level covariates interaction conditions are presented in Tables 2, 3, and 4, respectively.

Admissible Solution Rate

In multilevel MIMIC modeling with two covariates at the between level, the proportions of the admissible solutions rate (ASR) seemed to be related to ICC and the total sample size as presented in Table 2. The ASRs were higher with the small ICC. Simulation conditions with the combination of large ICC and small total sample size produced relatively lower percentage of admissible solutions (76% with CN=40, CS=10 in large ICC when covariates interaction effect was at .30). Under all conditions in small ICC, the ASRs were above 97%. As the total sample size increased, the ASR increased in medium and large ICCs, but the pattern was not very consistent. In the biggest total sample size of 2400, the ASRs were above 0.98 regardless of the ICC levels. The ASRs were very identical for the two different covariates interaction magnitude of .30 and .60. When the two covariates were at the within level, ASRs were all 100% except one condition that was at 99%, across all simulation conditions. When the two covariates were at cross level, ASRs were all 100% except some conditions with smaller sample size. In sum, ASRs were higher when the two covariates were at the within level or at the cross level than the two covariates at the between level. When the two covariates were at the within level or at the cross level, ICC had no impact on the ASR. However, when the two covariates were at the between level, lower ASR was associated with larger ICC.

Type I Error

Type I error rates of falsely detecting covariates interaction effect when there was no covariates interaction effect in multilevel MIMIC with two covariates at the between level ranged from .04 to .08, most of which were around .06. In general, multilevel MIMIC with between-level covariates adequately controlled the Type I error rates, as presented in Table 2. The design factors appeared to have no impact on Type I error rates. When the two covariates were at the within level, Type I error rates were very similar to those observed in between-level covariates interaction conditions. In cross-level covariates interaction conditions, Type I error rates were controlled well, ranging from .04 to .06. Thus, in all simulation conditions of Study one, Type I error rates were under control in ML MIMIC modeling. This was consistent with the result of the previous study that showed Type I error was well controlled when the hierarchical data structure was properly modeled (Finch & French, 2011).

Power

When the two covariates were at the between level, the power of detecting the between-level covariates interaction varied considerably as a result of the design factors with a minimum at .14 and a maximum at 1.00. The power decreased when the ICC level increased from small, to medium, and large after controlling for the other design factors. The power increased concomitantly with sample size. The lower end of the power stratum (i.e., .14 and .15) was found when sample size was small in large ICC of the covariates interaction of .30 (CN=40 and CS=10/20). CN seemed to have a stronger impact on power than CS or total sample size. To state it in another way, when total sample size was the same, a higher power was observed with more clusters (e.g., CN=80, CS=10) than with bigger cluster size (e.g., CN=40, CS=20). When CN was constant, a bigger power was observed with a bigger cluster size. When all the other factors were kept constant, the power of detecting a bigger magnitude of covariates interaction, .60 was higher than that of the smaller magnitude, .30.

When the two covariates were at the within level, the power of detecting within level covariates interaction effect differed substantially for the interaction effect magnitudes of .30 and .60. When the interaction magnitude was at .60, the power rates in all conditions were above .95, and most of which were at 1.00. On the other hand, when the effect size was at .30, a considerable variability was observed across the simulation conditions. ICC appeared to have no impact on the power, which varied depending on the total sample size. Cluster number seemed to have a positive impact effect on power, but not as big as cluster size. When the total sample size was the same, the power was always higher with bigger cluster number when the interaction effect was at .30.

The power of detecting cross-level covariates interaction ranged from .22 to .83 when the covariates interaction effect was at .20. When the interaction effect increased to .40, the power went up considerably, ranging from .60 to 1.00. As total sample size increased, the power increased. When the total sample size was the same, the impact of cluster number on the power was not consistent across different simulation conditions. When the cross-level interaction effect reached .40, and the sample size was bigger than 800, the power was above .89. With a smaller interaction effect of .20, a sample size of 2400 was needed to achieve a power of above .80.

With respect to the power, the location of the interaction effect was the most important factor. When the covariates interaction occurred at the within level or cross level, the power of detecting the interaction effect was much higher than the between level covariates interaction effect because of the sample size. This was expected because the sample size at the between level was smaller than the total sample size because the unit of analysis was the clusters (not individuals); however, the unit of analysis at the within level was individuals utilizing the total sample size when the interaction effect was at the within level or at the cross level.

Standard Error

Standard error (SE) of parameter estimate is used to test the statistical significance for which the parameter estimate is divided by the standard error. In the between-level covariates interaction conditions, SE ranged from .11 to .38, varying as a function of the ICC and the total sample size. SE increased with the increasing level of ICC, and it decreased as the total sample size increased. The smallest SE, .11, was observed in a total sample size of 2400 in small ICC, while the biggest SE, .38, was found in a total sample size of 400 in large ICC. SEs were very similar in the two different covariates interaction magnitudes of .30 and .60.

In the within-level covariates interaction conditions, SE ranged from .07 to .17, smaller than that of the between-level covariates interaction conditions, after controlling for all other simulation factors. SE was influenced by the total sample size. As total sample increased, SE decreased. ICC had no impact on the SE. In the conditions of the cross-level covariates interaction, SE was very similar to the conditions in which the covariates interaction effect was at the within level.

Bias and Relative Bias

Bias was calculated as the deviation of the estimated covariates interaction effect from the simulated value in the population. When the interaction effect occurred at the between level, bias rates were from -.03 to .02 with most of them at -.01 and zero, implying that the estimated parameter was around the simulated effect of covariates interaction regardless of the simulation conditions. For the within-level covariates interaction conditions, bias rates were mostly at zero, indicating that bias rates were negligible. With respect to the cross-level covariates interaction, bias rates ranged from zero to .02. In sum, bias rates were negligible across all simulation conditions. With respect to relative bias, in the conditions of between-level covariates interaction, it ranged from -.06 to .08 when the interaction effect was at .30, and it was smaller when the interaction effect was at .60, ranging from -.04 to zero. In the conditions of the within-level or cross-level covariates interaction, relative bias was all around zero.

RMSE

RMSE was employed to measure the precision of the parameter estimate. The smaller the RMSE was, the more precise the parameter estimate was. RMSE varied as a function of ICC level and the sample size, ranging from .07 to .39, when the covariates interaction effect was at the between level. A smaller ICC level was associated with a smaller RMSE value controlling for the other design factors. As the total sample size increased, cluster number in particular, RMSE decreased. RMSE was similar for covariates interaction effect of .30 and .60. When the covariates interaction effect was at the within level, RMSE was smaller than that of the between-level covariates interaction effect conditions, ranging from .07 to .17. RMSE varied as a function of only the total sample size. While ICC and covariates interaction magnitude seemed not to affect RMSE. Similar value and pattern of RMSE were observed in cross-level covariates interaction conditions as in within-level covariates interaction ones.

The second purpose of Study One was to examine the impact of design factors on outcome variables (i.e., ASR, Type I error, power, SE, bias, relative bias and RMSE). Factorial analysis of variance (ANOVA) with generalized eta-squared (η^2) was used to examine the impact of design factors on outcome variables. η^2 indicated the proportion of the variance explained by a specific design factor or the interaction of two or more of the design factors, and it was obtained by dividing the Type III sum of squares of a particular design factor or interaction of design factors by the corrected total sum of squares. The Cohen's (1973) moderate effect size of .0588 was applied as the practical significance level. ASR, Type I error, bias, relative bias did not vary much across different simulation conditions. Thus, only power, SE, and RMSE were used as the dependent variable in the ANOVA analyses.

Sources of Variance in Power

For the power, when the covariates interaction effect was at the between level, covariates interaction effect magnitude, ICC level, and cluster number were significantly associated with the power as shown in Table 5. The three factors accounted for 95% of the variance of power as shown in Table 5. Interaction effect magnitude, ICC, and cluster number explained 56%, 20%, and 19% of the variance of the power, respectively. Cluster size was not significantly associated

Table 2. Power, Type I error, bias, relative bias, standard error of estimate, and root mean square error (RMSE) under design factors with between-level covariates interaction effect at .30 and .60

			Between-level covariates interaction											
			Covariates interaction of .30						Covariates interaction of .60					
ICC	CN/CS	Error	ASR	Power	S.E.	Bias	R Bias	RMSE	ASR	Power	S.E.	Bias	R Bias	RMSE
Small	40/10	.07	.97	.23	.24	-.02	-.06	.23	.98	.72	.23	-.01	-.02	.24
	40/20	.07	.98	.33	.19	-.01	-.04	.20	.97	.87	.19	.00	.00	.20
	80/10	.05	.98	.44	.16	.00	-.01	.17	.99	.96	.16	.00	.00	.17
	80/20	.06	.97	.59	.14	.00	.00	.14	.97	.99	.14	-.01	-.02	.14
	120/10	.04	.98	.60	.13	.00	.01	.13	.98	1.00	.13	-.01	-.02	.13
	120/20	.06	.97	.75	.11	.00	-.01	.11	.97	1.00	.11	-.01	-.02	.12
Medium	40/10	.05	.90	.17	.30	-.01	-.04	.30	.90	.49	.30	-.02	-.03	.30
	40/20	.06	.84	.20	.27	-.01	-.03	.28	.85	.57	.27	-.03	-.04	.28
	80/10	.06	.89	.27	.21	-.01	-.05	.21	.89	.80	.21	-.01	-.02	.21
	80/20	.05	.91	.31	.19	-.01	-.05	.20	.93	.84	.19	-.02	-.03	.20
	120/10	.06	.90	.39	.17	-.01	-.03	.18	.90	.93	.17	.00	-.01	.17
	120/20	.06	.97	.47	.16	.00	-.01	.16	.97	.97	.16	.00	.00	.16
Large	40/10	.06	.76	.14	.38	.01	.02	.39	.78	.35	.38	-.01	-.02	.39
	40/20	.08	.91	.15	.36	.00	.01	.38	.88	.39	.36	.00	.00	.40
	80/10	.05	.89	.18	.27	-.01	-.03	.27	.88	.59	.28	.00	.00	.27
	80/20	.06	1.00	.23	.26	.02	.08	.27	1.00	.63	.26	.00	.00	.26
	120/10	.06	.97	.26	.22	-.01	-.03	.23	.96	.75	.22	.00	.00	.23
	120/20	.06	1.00	.33	.21	.01	.04	.21	1.00	.80	.21	.00	.00	.21

Note. ICC = intraclass correlation; CN= number of clusters per group; CS= cluster size; ASR = admissible solution rate; Error = Type I error; S.E. = standard error of the parameter estimate (i.e., latent group mean difference); R Bias=relative bias; RMSE=root mean square error.

Table 3. Power, Type I error, bias, relative bias, standard error of estimate, and root mean square error (RMSE) under design factors with within-level covariates interaction effect at .30 and .60

			Within-level covariates interaction											
			Covariates interaction of .30						Covariates interaction of .60					
ICC	CN/CS	Error	ASR	Power	S.E.	Bias	R Bias	RMSE	ASR	Power	S.E.	Bias	R Bias	RMSE
Small	20/20	.06	1.00	.46	.16	.00	.01	.17	1.00	.95	.16	.00	.01	.17
	20/40	.07	.99	.74	.11	.00	.02	.12	1.00	1.00	.12	.00	.00	.12
	40/20	.06	1.00	.69	.12	-.01	-.02	.12	1.00	1.00	.12	.01	.01	.12
	40/40	.06	1.00	.95	.08	.00	.01	.08	1.00	1.00	.08	.00	.00	.08
	60/20	.05	1.00	.88	.10	.00	-.01	.10	1.00	1.00	.10	.00	-.01	.10
	60/40	.06	1.00	.99	.07	.00	.01	.07	1.00	1.00	.07	.00	.01	.07
Medium	20/20	.06	1.00	.48	.17	.01	.02	.17	1.00	.95	.16	.01	.01	.17
	20/40	.07	1.00	.74	.11	.00	.00	.12	1.00	1.00	.11	-.01	-.01	.12
	40/20	.06	1.00	.73	.12	.00	.00	.12	1.00	1.00	.12	.00	.01	.12
	40/40	.07	1.00	.95	.08	.00	.00	.08	1.00	1.00	.08	.00	.00	.08
	60/20	.06	1.00	.87	.10	.00	.00	.10	1.00	1.00	.10	-.01	-.01	.09
	60/40	.05	1.00	.99	.07	.00	.00	.07	1.00	1.00	.07	.00	.01	.07
Large	20/20	.07	1.00	.51	.16	.01	.05	.17	1.00	.95	.16	.00	.00	.17
	20/40	.06	1.00	.74	.11	.00	.01	.12	1.00	1.00	.11	.00	.01	.12
	40/20	.05	1.00	.70	.12	.00	-.01	.12	1.00	1.00	.12	.00	.00	.12
	40/40	.05	1.00	.96	.08	.00	.00	.08	1.00	1.00	.08	.00	.01	.08
	60/20	.05	1.00	.87	.10	.00	.00	.09	1.00	1.00	.10	.00	.00	.10
	60/40	.06	1.00	1.00	.07	.00	.01	.07	1.00	1.00	.07	.00	.00	.07

Note. ICC = intraclass correlation; CN= number of clusters per group; CS= cluster size; ASR = admissible solution rate; Error = Type I error; S.E. = standard error of the parameter estimate (i.e., latent group mean difference); R Bias=relative bias; RMSE=root mean square error.

Table 4. Power, Type I error, bias, relative bias, standard error of estimate, and root mean square error (RMSE) under design factors with cross-level covariates interaction effect at .20 and .40

			Cross-level covariates interaction											
			Covariates interaction of .20						Covariates interaction of .40					
ICC	CN/CS	Error	ASR	Power	S.E.	Bias	R Bias	RMSE	ASR	Power	S.E.	Bias	R Bias	RMSE
Small	40/10	.06	.91	.22	.19	.02	.11	.18	.93	.62	.19	.02	.04	.18
	40/20	.05	1.00	.38	.12	.01	.04	.12	.99	.92	.12	.01	.02	.12
	80/10	.05	1.00	.38	.12	.00	.01	.13	.99	.91	.12	.01	.03	.13
	80/20	.05	1.00	.65	.09	.00	.00	.09	1.00	1.00	.09	.01	.02	.09
	120/10	.05	1.00	.56	.10	.01	.05	.10	1.00	.98	.10	.01	.03	.10
	120/20	.05	1.00	.83	.07	.00	.00	.07	1.00	1.00	.07	.00	.01	.07
Medium	40/10	.06	1.00	.20	.22	.02	.12	.18	1.00	.61	.19	.02	.04	.18
	40/20	.06	1.00	.37	.13	.01	.05	.12	1.00	.89	.13	.00	.01	.13
	80/10	.04	1.00	.36	.13	.01	.05	.12	1.00	.91	.13	.01	.02	.12
	80/20	.04	1.00	.67	.09	.01	.03	.08	1.00	1.00	.09	.01	.02	.09
	120/10	.06	1.00	.54	.10	.01	.03	.10	1.00	.98	.10	.00	.01	.10
	120/20	.05	1.00	.83	.07	.00	.01	.07	1.00	1.00	.07	.01	.01	.07
Large	40/10	.05	1.00	.21	.19	.02	.09	.18	1.00	.60	.19	.02	.05	.19
	40/20	.05	1.00	.38	.13	.01	.06	.12	1.00	.89	.13	.01	.02	.13
	80/10	.05	1.00	.39	.13	.01	.03	.12	1.00	.91	.13	.01	.02	.12
	80/20	.03	1.00	.62	.09	.00	.02	.09	1.00	.99	.09	.01	.01	.09
	120/10	.04	1.00	.54	.10	.01	.04	.10	1.00	.98	.10	.00	.00	.10
	120/20	.05	1.00	.82	.07	.00	.02	.07	1.00	1.00	.07	.01	.02	.07

Note. ICC = intraclass correlation; CN= number of clusters per group; CS= cluster size; ASR = admissible solution rate; Error = Type I error; S.E. = standard error of the parameter estimate (i.e., latent group mean difference); R Bias=relative bias; RMSE=root mean square error.

with the power with an eta-squared value of .01. Also, only the three main effects were significant, and no interactions between the factors were significantly associated with the power.

When the covariates interaction effect was at the within level, the magnitude of covariates interaction effect, cluster number, and cluster size, as well as the interaction between the magnitude and cluster number, and the interaction between magnitude and cluster size were significantly associated with the power of detecting within-level covariates interaction effect. They accounted for 97% of the variance of the power as shown in Table 6. Note that cluster size was significant when the covariates interaction was at the within level.

When the covariates interaction effect was at the cross level, the combination of magnitude, cluster number, and cluster size accounted for 96% of the variance in the power as presented in Table 7. The proportion of variance explained by cluster number is larger than that explained by cluster size.

Table 5. Eta-squared (η^2) of different sources for power, standard error, and RMSE when the covariates interaction effect was at the between level

Sources	Outcome variable		
	power	standard error	RMSE
Overall η^2	96.99%	99.91%	99.67%
magnitude	**55.72%**	0.00%	0.01%
ICC	**19.65%**	**47.04%**	**45.89%**
CN	**19.43%**	**48.28%**	**49.27%**
CS	1.27%	2.06%	1.17%
magnitude*CN	0. 61%	0.01%	0.03%
magnitude*ICC	0.49%	0.01%	0.02%
ICC*CN	0.21%	2.09%	3.02%
ICC*CS	0.17%	0.14%	0.24%
magnitude*CS	0.04%	0.00%	0.01%
CN*CS	0.01%	0.28%	0.01%

Note. ICC = intraclass correlation; CN= number of clusters per group; CS= cluster size; RMSE= root mean square error.

Table 6. Eta-squared (η^2) of different sources for power, standard error, and RMSE when the covariates interaction effect was at the within level

Sources	Outcome variable		
	power	standard error	RMSE
Overall η^2	99.14%	99.79%	100.00%
magnitude	**39.83%**	0.00%	0.00%
ICC	0.02%	0.02%	0.01%
CN	**21.03%**	**52.79%**	**62.64%**
CS	**12.54%**	**44.83%**	**34.91%**
magnitude*CN	**15.05%**	0.00%	0.00%
magnitude*ICC	0.02%	0.06%	0.04%
ICC*CN	0.04%	0.04%	0.03%
ICC*CS	0.01%	0.14%	0.01%
magnitude*CS	**9.09%**	0.04%	0.4%
CN*CS	1.51%	1.87%	2.11%

Note. ICC = intraclass correlation; CN= number of clusters per group; CS= cluster size; RMSE= root mean square error.

Table 7. Eta-squared (η^2) of different sources for power, standard error, and RMSE when the covariates interaction effect was at the cross level

Sources	Outcome variable		
	power	standard error	RMSE
Overall η^2	97.48%	99.26%	99.73%
magnitude	**57.70%**	0.04%	0.10%
ICC	0.03%	0.36%	0.05%
CN	**24.30%**	**62.46%**	**62.76%**
CS	**12.28%**	**31.44%**	**33.29%**
magnitude*CN	1.77%	0.09%	0.09%
magnitude*ICC	0.00%	0.09%	0.05%
ICC*CN	0.01%	0.31%	0.32%
ICC*CS	0.01%	0.09%	0.05%
magnitude*CS	1.01%	0.04%	0.02%
CN*CS	0.37%	4.34%	3.00%

Note. ICC = intraclass correlation; CN= number of clusters per group; CS= cluster size; RMSE= root mean square error.

Sources of Variance in Standard Error of Estimation

With respect to SE, when the covariates interaction effect was at the between level, cluster number and ICC accounted for 95% percent of the variance in SE as shown in Table 5. Cluster size accounted for only 2% of the variance in SE, and the magnitude of covariates interaction effect didn't explain any of the variance.

When the covariates interaction effect was at the within level, cluster number, and cluster size accounted for 97% of the variance in SE as shown in Table 6. All the other design factor had negligible effect on the variance in SE.

When the interaction effect was at the cross level, cluster number and cluster size accounted for 93% percent of the variance in SE as shown in Table 7. Cluster number explained 62% of the variance, twice as large as that of cluster size. The other main effects and interactions were not significantly associated with SE.

Sources of Variance in RMSE

With respect to RMSE, for the between-level covariates interaction effect, ICC and cluster number were significantly associated with RMSE as shown in Table 5, explaining about 63% and 35% of variance in RMSE, respectively. For the within-level and cross-level covariates interaction conditions, cluster number and cluster size explained around 96% of the variance in RMSE, and the other design factors were not significantly associated with RMSE.

In sum, for all the three outcome variables (i.e., power, SE, and RMSE), the magnitude of covariates interaction, ICC level, and cluster number explained most of the variance of them when the covariates interaction effect was at the between level. On the other hand, when the covariates interaction was at the within and the cross level, cluster number, and cluster size explained most of variance in SE and RMSE; the magnitude of interaction effect, cluster number, and cluster size accounted for most of the variance in the power.

Discussion

MIMIC models are flexible in modeling covariates that are related to the latent variables in educational and psychological research. However, the performance of ML MIMIC models in detecting the covariates interaction effect has not been examined yet. The purpose of Study one was to examine the performance of ML MIMIC models in detecting covariates interaction effect. In multilevel data, the covariates can be all at the within level, all the between level, and some covariates are at the within level and some covariates are at the between level, thus allowing for a more systematic investigation of the variance in latent variables. The location of the covariates creates interaction effect at different levels. Specifically, this study examined the efficacy of ML MIMIC in detecting covariates interaction effect when the interaction effect was at the between level, at the within level, and at the cross level. Prior research has demonstrated the importance of using multilevel MIMIC modeling rather than traditional single level MIMIC modeling in the presence of multilevel data structure when there was only one within-level or between-level covariate related to the latent variable (Finch & French, 2011). This study extended the prior research by examining the performance of ML MIMIC models in estimating covariates interaction effect when the covariates were at the within level, or at the between level, or at both levels, creating three possible types of covariates interaction effect.

The results of Study One suggested that when the covariates interaction effect was at the between level, the power of ML MIMIC in detecting the covariates interaction depended on magnitude of the interaction effect, the ICC level, and the cluster number. When the interaction effect was at .30, a large sample size (e.g., 2400) was required to get a power of around .70 for a small ICC level of .09. In conditions of large ICC, even the largest sample size produced a power of .33 of detecting the interaction effect. For an interaction effect at .60, the power of ML MIMIC in detecting the interaction effect was very decent for the small ICC level, ranging from .72 to 1.00. However, when the ICC level became larger, the power decreased. When the other design factors were kept constant, ICC had a negative impact on power in detecting interaction effect in ML MIMIC models. This was expected, given the formula of design effect of multilevel model. Design effect (deff) is defined as: deff=1+(n-1)*ICC, where n represents cluster size, and ICC is the intraclass correlation. Deff gauges the extent to which the sampling error occurs in sampling individuals from a specific sampling design that differs from the traditional simple random sampling in which each individual has equal chance of being selected. (Heck & Thomas, 2015). For the within-level modeling, the within-level sample size (total sample size) is of main importance, and this design effect is not shown in the within level. Design effect increases as the ICC level increases, holding cluster size constant. When the data has no clustering effect (i.e., ICC=0), design effect equals 1. In this case, the data are simple random sampling. A design effect of 3.0 would suggest that a sample three times as large as the sample in simple random sampling is required to produce identical sample variability to simple random sampling (Heck & Thomas, 2015). Thus, the larger the design effect is, the bigger the sample size is required. In this study, the smallest design effect was 1.81 (ICC=.09, CS=10), and the largest design effect was 7.27 (ICC=.33, CS=20). ICC, the ratio of the between-level variability to total variability, also indicates the correlation of individuals within the same cluster (Hox, 2010). An ICC of .33 means that 33% of the variance of an outcome variable exists at the between level, and it also indicates that the correlation between any two selected individuals in a cluster is .33. The higher the within cluster correlation is, the more homogeneous the individuals in the cluster are, implying that adding sampling individuals in the cluster doesn't add much information. Therefore, as design effect increases, the effective sample size decreases. When the effective sample size became smaller as a result of the increasing ICC, the power of detecting covariates interaction effect decreased in ML MIMIC models. For the same reason, the ASR rate was lower in large ICC conditions than in smaller ICC conditions when the total sample size was smaller. When the sample size was large, for example, 2400, the ASR was not affected much by the ICC level. Similarly, standard error and RMSE were affected by ICC as well controlling for all other design factors. As expected, when the ICC level increased, standard error and RMSE increased. It is of note that standard error and RMSE were not affected by interaction effect magnitude, but by ICC, and sample size.

The performance of ML MIMIC was more impressive with within-level covariates interaction effect than with between-level covariates interaction effect. The power of detecting the interaction effect was around .95 when the sample size reached 1600 for an interaction effect of .30. When the interaction effect was at .60, the power was 1.00 except the smallest sample size of 400. All the outcome variables were affected by sample size, but not affected by the ICC level. This is also expected, as the interaction effect was at the within level and the estimation of the interaction effect utilized the within-level total sample size. The magnitude of the ICC level was not associated with the ASR, power, standard error, bias, relative bias, and RMSE.

The focus of the study was the performance of ML MIMIC modeling in detecting cross-level covariates interaction effect. For a small interaction effect at .20, the power varied as a function of total sample size, ranging from .21 to .83. Note that when the total sample size was identical, cluster number had no impact on power, that is, the power of 40 clusters with a cluster size of 20 was very similar to that of 80 clusters with a cluster size of 10. When sample size was as big as 2400, the power of ML MIMIC in detecting cross-level covariates interaction effect could be around .83. Therefore, even with a small cross-level interaction effect, the power of ML MIMIC in detecting it could be above .80 when the sample size was 2400. For a moderate cross-level interaction effect at .40, the power of ML MIMIC in detecting a significant cross-level covariates interaction effect was above .90 when the sample size reached 800 or above. In sum, the performance of ML MIMIC in detecting and estimating cross-level covariates interaction effect was promising. When the interaction effect was at the cross level, the ICC level didn't have an impact on the outcome variables. Type I error, Power, standard error, bias, and RMSE were similar across different ICC levels when the other design factors were held constant. This tendency was more similar to that of the performance of MLMIMIC with within-level covariates interaction effect than to that of the between-level covariates interaction effect. Also, the magnitude of standard error and RMSE was also similar to the results in the within-level covariates interaction conditions.

Type I error was well controlled in all conditions for the between-level covariates interaction effect model, the within-level covariates interaction effect model, and the cross-level covariates interaction effect model. This was very important for the interpretation of power, because the interpretation of power was meaningless if Type I error rate was high.

Previous research has demonstrated the lower power of ML MIMIC in detecting the effect of the between-level covariate on the latent factor, especially when the ICC was large and sample size was small (Finch & French, 2011). Similarly, lower power was observed in detecting between-level covariates interaction effect in ML MIMIC than in detecting within-level covariates interaction effect and cross-level covariates interaction effect. Practitioners using ML

MIMIC to examine covariates interaction effect should be cognizant of the location of the interaction effect. The performance ML MIMIC in detecting the covariates interaction effect was more promising when the covariates interaction effect was at the within level or at the cross level.

The decent performance of the ML MIMIC modeling in detecting and estimating the cross-level covariates interaction effect assured the researchers and practitioners of its correct statistical modeling of the data structure, but also led to greater insights into examining the influence of covariates from both the between level and within level on latent factors. The substantively meaningful cross-level covariates interaction, that is, the association between within-level covariate and within-level outcome varying as a function of the value of the between-level, or contextual covariates, can be modeled flexibly and estimated accurately in ML MIMIC with a relatively big sample size. With the increasing prevalence of examining cross-level interaction effect in applied psychology (Mathieu et al., 2012), the decent performance of ML MIMIC in estimating cross-level covariates interaction renders the practitioners to apply it more comfortably. On the contrary, the estimation and interpretation of the between-level covariates interaction effect should be made with caution, particularly when the ICC level is high and cluster number is small.

The impact of cluster number and cluster size on power in detecting covariates interaction effect varied depending on the location of the interaction effect in ML MIMIC. To be more specific, cluster number was more important than cluster size in detecting between-level covariates interaction effect, while both cluster number and cluster size were important in estimating within-level covariates interaction effect and cross-level covariates interaction effect. These shed lights to the sampling strategies and allocation for resources for researchers. If the researcher is interested in estimating between-level covariates interaction effect, more clusters should be sampled if resources allow (sampling more clusters is more expensive than sampling more individuals from a smaller number of clusters). If the researcher is interested in estimating within-level or cross-level covariates interaction, more thorough sampling of individual in clusters (less costly than sampling more clusters with a smaller cluster size) enables an enhanced power because the total sample size is an important factor at the within level.

As mentioned in the data generation part, cluster invariance was met in ML MIMIC modeling. To be more specific, factor loadings were constrained to be equal across levels (within level factor loadings are identical to between-level factor loadings). Intercepts had a very small variance at the between level (.02, .01, and .004 for large, medium, and small ICC levels, respectively), implying that intercepts were very similar across clusters. Moreover, the residual variances were constant across clusters. The impact of violating cluster invariance on the performance of ML MIMIC in detecting covariates interaction effect is unknown.

The generalizability of the performance of ML MIMIC should not be extended beyond the design of the measurement part and structural part of ML MIMIC modeling in this study. Specifically, there was only one factor with 6 continuous indicators with factor loadings all at .80. The main effects of the covariates were .30 and .40 when the two covariates were at the same level, and .40 (within-level covariate main effect) and .50 (between-level covariate main effect) for the cross-level. The impact of changing the design factors mentioned above on the performance of ML MIMIC is unknown. More comprehensive simulation studies varying the measurement and structural parts of the model are called for.

Results of Study Two

The purpose of Study Two was to investigate the impact of ignoring covariates interaction on model fit indices and model parameter estimates in ML MIMIC models. In Study Two, the population models were the models in Study One, with covariates interaction effect at the between level, at the within level, and at cross level. The population models in Study One with no covariates interaction effect used to evaluate Type 1 error were excluded from Study Two. When the two covariates were both at the between level or at the within level, the analysis models specified the main effects of the two covariates on the between level factor or the within level factor, but omitted the interaction effect (i.e., the product term of the two covariates) on the latent factor. Similarly, in the case of cross-level interaction effect the analysis model specified that the within-level factor and the between-level factor were regressed on the within-level covariate and the between-level covariate, respectively; however, the within-level regression effect was not specified to be predicted by the between-level covariate, that is, the interaction effect was ignored.

The impact of omitting covariates interaction effect in ML MIMIC has not been examined in literature. It is expected that some of the parameters in the models are to be affected as a result of the model misspecification. It remains unknown which of the parameters are going to be impacted and how they are affected. In SEM, a set of fit indices (CFI, TLI, RMSEA, SRMR) can be used to evaluate model adequacy. It is yet known whether the fit indices can detect the misspecification as compared to the results from the analysis of the correct model in Study One.

The results of Study Two were presented in the sequence of omitting between-level covariates interaction, omitting within-level covariates interaction, and omitting cross-level covariates interaction.

Omitting the Between-level Covariates Interaction

When the between-level covariates interaction was omitted, among all the SEM-based fit indices examined in this study: CFI, TLI, RMSEA, SRMR-B, and SRMR-W, none of them was found to be sensitive to the omission of the covariates interaction effect as presented in Table 8. The CFI values of both correct model and misspecified model were identical, at .98 or .99, and the same pattern was observed for TLI, ranging from .96 to .99 for both correct and misspecified model, indicating a very good model fit. TLI was very similar to CFI, and it was not presented in Table 8. RMSEA for both correct and misspecified model was around .05. SRMR-B produced in the correct model and the misspecified model was very similar, ranging from .01 to .05. SRMR-W was at .05 in all conditions, regardless of the model specified. CFI, TLI, RMSEA, SRMR-B, and SRMR-W were very similar for interaction effect at .30 and .60. The standard deviation of these fit indices in the correct model was very similar to that in the misspecified model.

The proportion of the replications that AIC, BIC, and SaBIC selected the correct model (i.e., a model with an interaction effect) were computed to denote their sensitivity to model misspecification. When the omitted covariates interaction effect was .30, the proportions ranged from .30 to .90, from .03 and .42, from .18 to .68 for AIC, BIC, and SaBIC, respectively. The results showed that their proportions increased as sample size increased, especially as cluster number increased. ICC had a negative impact on their sensitivity to model misspecification. When ICC increased, their proportions decreased, controlling for all the other design factors.

Therefore, AIC, BIC, and SaBIC had the best performance in the condition with the biggest sample size and small ICC. Although the pattern of the three fit indices was very similar to each other, AIC was more sensitive to the misspecification than SaBIC which was more sensitive than BIC. When the omitted covariates interaction was at .60, all the three fit indices were more sensitive to the model misspecification than the covariates interaction of .30. The proportions of replications that selected the correct model for AIC, BIC, and SaBIC ranged from .54 to 1.00, from .19 to .98, from .41 to 1.00, respectively. In some conditions with bigger sample size and small ICC, AIC, and SaBIC selected the correct model in all replications. The impact of sample size and ICC on the performance of the three fit indices was similar to that in the covariates interaction effect of .30.

AIC and SaBIC outperformed BIC in terms of selecting the correct model. BIC is defined as:

$$BIC=-2\log L(\theta) + \log(n)r,$$

where $\log L(\theta)$ denotes the log likelihood; n represents sample size; r is the number of parameters in the model. Lukočiené and Vermunt (2010) discussed a specific issue when using BIC in the case of multilevel analysis: It is not clear whether the sample size, n, should be the cluster number, J, or the total sample size at the within level. Pauler (1998) suggested the using of J for decisions about between-level model features and n for within-level model features in the context of linear mixed models. In Mplus output, BIC was calculated using the total sample size, n. To evaluate the sensitivity of BIC using the cluster number as sample size in the calculation, BIC(J) was computed and listed in Table 8. BIC(J) performed better than BIC in selecting the correct model, but still not as good as AIC and SaBIC. Note that BIC(J) was calculated only for the between-level covariates interaction model, because in the within-level and cross-level covariates interaction models total the sample size was used for decisions about model features.

The omission of the covariates interaction effect was expected to impact the other parameter estimates in the model. Results consistently showed that the main effects of the two covariates were overestimated as presented in Table 9. In the correct model, the main effects of the two covariates were around .30 and .40, respectively. In the misspecified model omitting the covariates interaction effect, the two main effects were overestimated around .44, and .54, respectively, when the interaction effect was at .30. The two main effects were estimated to be around .59 and .69 when the interaction effect was at .60, under all conditions. This indicated that the omitted interaction effect was carried through the main effects, divided almost equally between the two main effects. Between-level factor mean was underestimated in all conditions when the interaction effect was omitted. In the correct model, the between-level factor mean was around zero in all conditions. In the misspecified model omitting the covariates interaction effect, the between-level factor mean was estimated to be around -.07, and -.14, when the covariates interaction was at .30, and .60, respectively. In addition to the affected three parameters above, between-level factor residual variance was slightly impacted as shown in Table 9. The between-level residual variance was overestimated by approximately .01 and .02 for covariates interaction of .30 and .60, respectively. The estimates of all other parameters were not impacted by the misspecification, including the between-level factor loadings, the between-level intercepts, the between-level residual variances, the within-level factor loadings, the within-level factor mean, and the within-level residual variances.

Table 8. The proportion of the replications that AIC, BIC, BIC(J), and SaBIC selected the correct model, as well as CFI, RMSEA, SRMR-B, and SRMR-W and their standard deviations of correct and misspecified models with between-level covariates

							Between-level covariates interaction							
							SEM based fit indices (correct model)				SEM based fit indices (misspecified model)			
effect	ICC	CN/CS	AIC	BIC	BIC (J)	Sa BIC	CFI (sd)	RM (sd)	SRb (sd)	SRw (sd)	CFI (sd)	RM (sd)	SRb (sd)	SRw (sd)
.30	S	40/10	.45	.14	.28	.36	.98(.02)	.06(.02)	.05(.02)	.05(.01)	.98(.04)	.07(.04)	.04(.01)	.05(.01)
		40/20	.54	.15	.35	.37	.98(.01)	.05(.01)	.03(.01)	.05(.01)	.98(.02)	.06(.01)	.03(.01)	.05(.01)
		80/10	.68	.24	.41	.50	.99(.00)	.05(.01)	.04(.01)	.05(.01)	.99(.01)	.05(.01)	.03(.01)	.05(.01)
		80/20	.77	.28	.54	.56	.99(.00)	.04(.01)	.02(.00)	.05(.01)	.99(.00)	.05(.01)	.02(.00)	.05(.01)
		120/10	.81	.36	.52	.62	.99(.00)	.04(.01)	.03(.01)	.05(.01)	.99(.00)	.05(.01)	.03(.01)	.05(.01)
		120/20	.90	.42	.66	.68	.99(.00)	.04(.00)	.02(.00)	.05(.00)	.99(.00)	.04(.00)	.02(.00)	.05(.01)
	M	40/10	.37	.09	.20	.27	.99(.01)	.05(.01)	.04(.01)	.05(.01)	.98(.01)	.06(.02)	.04(.01)	.05(.01)
		40/20	.38	.07	.22	.24	.99(.01)	.05(.01)	.03(.01)	.05(.01)	.98(.01)	.06(.02)	.04(.01)	.05(.01)
		80/10	.48	.11	.24	.31	.99(.00)	.04(.01)	.03(.01)	.05(.01)	.99(.01)	.05(.01)	.03(.01)	.05(.01)
		80/20	.53	.12	.27	.28	.99(.00)	.04(.00)	.02(.00)	.05(.01)	.99(.00)	.04(.01)	.02(.00)	.05(.01)
		120/10	.59	.18	.32	.39	.99(.00)	.04(.01)	.02(.00)	.05(.01)	.99(.00)	.04(.01)	.02(.00)	.05(.01)
		120/20	.68	.18	.38	.40	.99(.00)	.04(.00)	.02(.00)	.05(.00)	.99(.00)	.04(.00)	.01(.00)	.05(.00)
	L	40/10	.33	.03	.16	.23	.98(.01)	.05(.01)	.03(.01)	.05(.01)	.98(.01)	.05(.01)	.03(.01)	.05(.01)
		40/20	.30	.06	.16	.18	.99(.00)	.04(.01)	.03(.01)	.05(.01)	.99(.00)	.04(.01)	.02(.01)	.05(.01)
		80/10	.38	.07	.16	.20	.99(.00)	.04(.01)	.02(.01)	.05(.01)	.99(.00)	.04(.01)	.02(.00)	.05(.01)
		80/20	.43	.08	.20	.21	.99(.00)	.04(.00)	.02(.00)	.05(.01)	.99(.00)	.04(.00)	.02(.00)	.05(.01)
		120/10	.47	.10	.20	.26	.99(.00)	.04(.01)	.02(.00)	.05(.01)	.99(.00)	.04(.01)	.02(.00)	.05(.01)
		120/20	.53	.08	.24	.25	.99(.00)	.04(.00)	.01(.00)	.05(.00)	.99(.00)	.04(.01)	.01(.01)	.05(.01)

Table 8 (Continued). The proportion of the replications that AIC, BIC, BIC(J), and SaBIC selected the correct model, as well as CFI, RMSEA, SRMR-B, and SRMR-W and their standard deviations of correct and misspecified models with between-level covariates

			Between-level covariates interaction												
							SEM based fit indices (correct model)				SEM based fit indices (misspecified model)				
effect	ICC	CN/CS	AIC	BIC	BIC (J)	Sa BIC	CFI (sd)	RM (sd)	SRb (sd)	SRw (sd)	CFI (sd)	RM (sd)	SRb (sd)	SRw (sd)
.60	S	40/10	.87	.57	.75	.81	.98(.02)	.05(.01)	.03(.01)	.05(.01)	.98(.01)	.06(.02)	.03(.01)	.05(.01)
		40/20	.96	.68	.88	.89	.98(.01)	.05(.01)	.02(.01)	.05(.01)	.98(.01)	.05(.01)	.02(.01)	.05(.01)
		80/10	.99	.86	.95	.97	.99(.00)	.05(.01)	.02(.01)	.05(.01)	.99(.00)	.05(.01)	.02(.01)	.05(.01)
		80/20	1.00	.92	.99	.99	.99(.00)	.04(.01)	.02(.00)	.05(.01)	.99(.00)	.05(.01)	.01(.00)	.05(.01)
		120/10	1.00	.96	.99	1.00	.99(.00)	.04(.01)	.02(.00)	.05(.01)	.99(.00)	.05(.01)	.02(.00)	.05(.01)
		120/20	1.00	.98	1.00	1.00	.99(.00)	.04(.00)	.01(.00)	.05(.00)	.99(.00)	.04(.00)	.01(.00)	.05(.00)
	M	40/10	.71	.33	.52	.60	.98(.01)	.05(.01)	.03(.01)	.05(.01)	.98(.01)	.06(.01)	.03(.01)	.05(.01)
		40/20	.74	.33	.59	.61	.99(.01)	.04(.01)	.02(.01)	.05(.01)	.99(.01)	.05(.01)	.02(.01)	.05(.01)
		80/10	.92	.58	.77	.83	.99(.00)	.04(.01)	.02(.00)	.05(.01)	.99(.00)	.04(.01)	.02(.00)	.05(.01)
		80/20	.94	.60	.80	.81	.99(.00)	.04(.01)	.02(.00)	.05(.01)	.99(.00)	.04(.01)	.01(.00)	.05(.01)
		120/10	.98	.77	.89	.92	.99(.00)	.04(.01)	.02(.00)	.05(.01)	.99(.00)	.04(.01)	.02(.00)	.05(.01)
		120/20	.99	.83	.95	.95	.99(.00)	.04(.00)	.01(.00)	.05(.00)	.99(.00)	.04(.00)	.01(.00)	.05(.00)
	L	40/10	.54	.19	.36	.45	.98(.01)	.05(.01)	.03(.01)	.05(.01)	.99(.01)	.05(.01)	.03(.01)	.05(.01)
		40/20	.58	.20	.40	.41	.99(.01)	.04(.01)	.02(.01)	.05(.01)	.99(.00)	.04(.01)	.02(.01)	.05(.01)
		80/10	.79	.34	.54	.62	.99(.00)	.04(.01)	.02(.00)	.05(.01)	.99(.00)	.04(.01)	.02(.00)	.05(.01)
		80/20	.81	.32	.58	.60	.99(.00)	.04(.00)	.02(.00)	.05(.01)	.99(.00)	.04(.01)	.01(.00)	.05(.01)
		120/10	.88	.50	.69	.75	.99(.00)	.04(.01)	.02(.00)	.05(.01)	.99(.00)	.04(.01)	.02(.00)	.05(.01)
		120/20	.91	.49	.72	.74	.99(.00)	.04(.00)	.01(.00)	.05(.00)	.99(.00)	.04(.00)	.01(.00)	.05(.00)

Note. ICC = intraclass correlation; CN= number of clusters per group; CS= cluster size; S=small; M=medium; L=large; AIC=Akaike information criterion; BIC=Bayesian information criterion; BIC(J)=BIC with between-level sample size; SaBIC=sample-size adjusted BIC; CFI=Comparative Fit Index; sd= standard deviation; RM=root mean square error of approximation; SRb= standardized root mean square residual for the between level; SRw= standardized root mean square residual for the within level.

Table 9. The estimated values of main covariates effects and between-level factor mean of correct and misspecified models with between-level covariates

		Betwee-level covariates interaction															
		Covariates interaction of .30								Covariates interaction of .60							
		Correct model				Misspecified model				Correct model				Misspecified model			
ICC	CN/CS	M1	M2	fb	fbv	M1	M2	fb	fbv	M1	M2	fb	Fbv	M1	M2	fb	fbv
Small	40/10	.30	.40	.00	.03	.44	.54	-.07	.04	.30	.39	.00	.03	.60	.69	-.14	.06
	40/20	.29	.40	.01	.04	.44	.54	-.06	.05	.29	.39	.01	.04	.59	.68	-.14	.07
	80/10	.29	.39	.01	.04	.44	.54	-.07	.04	.29	.39	.01	.04	.59	.69	-.14	.06
	80/20	.29	.39	.01	.05	.44	.54	-.07	.06	.30	.40	.00	.05	.59	.69	-.14	.07
	120/10	.29	.39	.01	.04	.44	.54	-.07	.05	.30	.40	.00	.04	.59	.69	-.14	.06
	120/20	.30	.39	.01	.05	.44	.54	-.07	.06	.30	.40	.00	.05	.60	.69	-.14	.07
Medium	40/10	.29	.39	.01	.11	.43	.53	-.06	.12	.30	.40	.01	.12	.59	.69	-.14	.14
	40/20	.30	.40	.00	.13	.45	.54	-.07	.14	.31	.41	.00	.13	.60	.70	-.14	.15
	80/10	.30	.39	.01	.12	.44	.54	-.06	.13	.29	.40	.01	.13	.59	.69	-.14	.15
	80/20	.30	.40	.00	.14	.45	.54	-.07	.14	.30	.40	.01	.14	.59	.69	-.14	.16
	120/10	.29	.39	.01	.13	.44	.54	-.06	.14	.29	.39	.01	.13	.59	.69	-.14	.16
	120/20	.30	.39	.00	.14	.45	.54	-.07	.15	.29	.39	.01	.14	.59	.69	-.14	.17
Large	40/10	.29	.38	.01	.25	.45	.53	-.06	.27	.28	.39	.01	.26	.58	.69	-.14	.29
	40/20	.30	.40	.00	.27	.45	.56	-.07	.29	.29	.39	.01	.28	.59	.69	-.14	.31
	80/10	.29	.40	.00	.27	.44	.54	-.07	.28	.29	.39	.01	.28	.59	.69	-.14	.31
	80/20	.28	.38	.02	.29	.44	.54	-.06	.30	.30	.40	.00	.29	.60	.70	-.15	.32
	120/10	.30	.39	.01	.28	.44	.54	-.07	.29	.29	.39	.01	.28	.59	.69	-.14	.31
	120/20	.29	.39	.01	.29	.45	.54	-.07	.30	.30	.39	.00	.29	.60	.69	-.14	.32

Note. ICC = intraclass correlation; CN= number of clusters per group; CS= cluster size; M1=main effect of covariate 1 (.30); M2=main effect of covariate 2 (.40); fb= between-level factor mean; fbv=between-level factor residual variance.

Omitting the Within-level Covariates Interaction

When the within-level covariates interaction effect was omitted in the model, CFI, TLI, RMSEA, SRMR-B, SRMR-W produced in the correct model were identical to those in the misspecified model as shown in Table 10. None of these fit indices detected the misspecification. CFI and TLI ranged from .99 to 1.00 in both correct and misspecified model; RMSEA was between zero and .03 in the correct model, and between .01 and .04 in the misspecified model; SRMR-B was between .01 and .06 (in the smallest sample size) in both correct and misspecified models; SRMR-W was between .01 and .02 in both correct and misspecified models. The proportions of replications that AIC, BIC, and SaBIC selected the correct model increased as the total sample size increased. Note that cluster number didn't have an effect on them. ICC was not associated with the sensitivity of the three fit indices. The proportions were very similar in small, medium, and large ICC when the other design factors were constant. When the omitted covariates interaction effect was at .30, the proportions of AIC, BIC, and SaBIC ranged from .66 to 1.00, from .24 to.96, from .54 to.99, respectively. When the omitted covariates interaction was at .60, all three fit indices selected the correct model in 100% of the replications except in the conditions with the smallest sample size.

With respect to parameter estimates, biased parameter estimates included the main effects of the two within-level covariates, and the between-level factor mean. Similar to the scenario of omitting between-level interaction effect, the omitted covariates interaction effect was carried through the two main effects. Note that the two main effects received exactly half of the interaction effect. For instance, when the interaction effect was at .30, the two main effects in the misspecified model were 0.15 higher than in the correct model; when the interaction effect was .60, the two main effects were 0.30 higher than in the correct model. The between-level factor mean was underestimated when the within-level covariates interaction effect was omitted. In the correct model the between-level factor mean was around -.20, which was estimated to be around -.27, and around -.35 when the interaction effect was .30, and .60, respectively. The other parameters in the multilevel MIMIC model were not affected.

Omitting the Cross-level Covariates Interaction

The two-level random effect modeling didn't produce CFI, RMSEA, SRMR, fit indices that are commonly used in SEM; however, it did produce AIC, BIC, and SaBIC. Therefore, checking the sensitivity of fit indices to misspecification in cross-level covariates interaction conditions applies only to AIC, BIC, and SaBIC. When the cross-level covariates interaction of .20 was omitted, AIC (from .08 to .31), BIC (from .02 to .10), and SaBIC (from .05 to .25) often failed to select the correct model. However, when the omitted covariates interaction effect increased to .40, the sensitivity of the three fit indices improved substantially. The proportions of replications that AIC, BIC, and SaBIC selected the correct model ranged from .54 to .90, from .30 to .78, and from .49 to .85, respectively. The sensitivity of the three fit indices increased as sample size increased. As ICC increased, the sensitivity of the fit indices improved a little bit.

With respect to the bias in parameter estimates resulted from the omission of the cross-level covariates interaction effect, more parameters were affected than in the conditions of omitting within-level or between-level covariates interaction effect. Again, the main effects of the within-level covariate and the between-level covariate were overestimated, receiving approximately half of the size of the omitted interaction effect, respectively. To be specific, when the omitted interaction effect was at .20, the within-level main effect was overestimated to

Table 10. The proportion of the replications that AIC, BIC, SaBIC selected the correct model, as well as CFI, RMSEA, SRMR-B, and SRMR-W and their standard deviations of correct and misspecified models with within-level covariates

Within-level covariates interaction

						SEM based fit indices (correct model)				SEM based fit indices (misspecified model)			
effect	ICC	CN/CS	AIC	BIC	SaBIC	CFI(sd)	RM(sd)	SRb(sd)	SRw(sd)	CFI(sd)	RM(sd)	SRb(sd)	SRw(sd)
.30	S	20/20	.66	.24	.54	.99(.04)	.03(.03)	.06(.06)	.02(.00)	.99(.02)	.04(.03)	.05(.06)	.02(.00)
		20/40	.87	.51	.76	1.00(.01)	.02(.02)	.03(.03)	.01(.00)	1.00(.01)	.02(.02)	.03(.03)	.01(.00)
		40/20	.84	.48	.70	1.00(.00)	.02(.01)	.03(.02)	.01(.00)	1.00(.00)	.02(.01)	.03(.02)	.01(.00)
		40/40	.98	.83	.94	1.00(.00)	.01(.01)	.02(.01)	.01(.00)	1.00(.00)	.01(.01)	.02(.01)	.01(.00)
		60/20	.95	.68	.87	1.00(.00)	.01(.01)	.03(.01)	.01(.00)	1.00(.00)	.01(.01)	.03(.01)	.01(.00)
		60/40	1.00	.96	.99	1.00(.00)	.01(.01)	.02(.01)	.01(.00)	1.00(.00)	.01(.01)	.02(.01)	.01(.00)
	M	20/20	.66	.27	.56	1.00(.00)	.03(.02)	.03(.02)	.02(.00)	.99(.02)	.03(.02)	.03(.02)	.02(.00)
		20/40	.88	.46	.75	1.00(.00)	.02(.01)	.02(.01)	.01(.00)	1.00(.00)	.02(.01)	.02(.01)	.01(.00)
		40/20	.88	.45	.75	1.00(.00)	.01(.01)	.02(.01)	.01(.00)	1.00(.00)	.01(.01)	.02(.01)	.01(.00)
		40/40	.99	.80	.92	1.00(.00)	.01(.01)	.01(.00)	.01(.00)	1.00(.00)	.01(.01)	.01(.00)	.01(.00)
		60/20	.95	.66	.87	1.00(.00)	.01(.01)	.01(.01)	.01(.00)	1.00(.00)	.01(.01)	.01(.01)	.01(.00)
		60/40	1.00	.95	.98	1.00(.00)	.00(.01)	.01(.00)	.01(.00)	1.00(.00)	.01(.01)	.01(.00)	.01(.00)
	L	20/20	.69	.28	.60	.99(.03)	.02(.02)	.02(.01)	.02(.00)	1.00(.00)	.02(.02)	.02(.01)	.02(.00)
		20/40	.87	.49	.75	1.00(.00)	.01(.01)	.01(.01)	.01(.00)	1.00(.00)	.01(.01)	.01(.01)	.01(.00)
		40/20	.87	.46	.74	1.00(.00)	.01(.01)	.01(.00)	.01(.00)	1.00(.00)	.01(.01)	.01(.00)	.01(.00)
		40/40	.99	.82	.95	1.00(.00)	.01(.01)	.01(.00)	.01(.00)	1.00(.00)	.01(.01)	.01(.00)	.01(.00)
		60/20	.95	.68	.87	1.00(.00)	.01(.01)	.01(.00)	.01(.00)	1.00(.00)	.01(.01)	.01(.00)	.01(.00)
		60/40	1.00	.96	.99	1.00(.00)	.00(.01)	.01(.00)	.01(.00)	1.00(.00)	.01(.01)	.01(.00)	.01(.00)

Table 10 (Continued). The proportion of the replications that AIC, BIC, SaBIC selected the correct model, as well as CFI, RMSEA, SRMR-B, and SRMR-W and their standard deviations of correct and misspecified models with within-level covariates

Within-level covariates interaction

						SEM based fit indices (correct model)				SEM based fit indices (misspecified model)			
effect	ICC	CN/CS	AIC	BIC	SaBIC	CFI(sd)	RM(sd)	SRb(sd)	SRw(sd)	CFI(sd)	RM(sd)	SRb(sd)	SRw(sd)
.60	S	20/20	.98	.88	.97	.99(.04)	.03(.03)	.06(.06)	.02(.00)	.99(.04)	.04(.04)	.06(.06)	.02(.00)
		20/40	1.00	.99	1.00	1.00(.01)	.02(.02)	.03(.03)	.01(.00)	1.00(.01)	.02(.02)	.03(.02)	.01(.00)
		40/20	1.00	1.00	1.00	1.00(.00)	.02(.01)	.03(.02)	.01(.00)	1.00(.00)	.02(.01)	.03(.02)	.01(.00)
		40/40	1.00	1.00	1.00	1.00(.00)	.01(.01)	.02(.01)	.01(.00)	1.00(.00)	.01(.01)	.02(.01)	.01(.00)
		60/20	1.00	1.00	1.00	1.00(.00)	.01(.01)	.03(.01)	.01(.00)	1.00(.00)	.01(.01)	.03(.01)	.01(.00)
		60/40	1.00	1.00	1.00	1.00(.00)	.01(.01)	.02(.01)	.01(.00)	1.00(.00)	.01(.01)	.02(.01)	.01(.00)
	M	20/20	.99	.88	.97	1.00(.02)	.03(.02)	.03(.02)	.02(.00)	.99(.01)	.03(.02)	.03(.01)	.01(.00)
		20/40	1.00	1.00	1.00	1.00(.00)	.02(.01)	.02(.01)	.01(.00)	1.00(.00)	.02(.01)	.02(.01)	.01(.00)
		40/20	1.00	1.00	1.00	1.00(.00)	.01(.01)	.02(.01)	.01(.00)	1.00(.00)	.01(.01)	.02(.01)	.01(.00)
		40/40	1.00	1.00	1.00	1.00(.00)	.01(.01)	.01(.00)	.01(.00)	1.00(.00)	.01(.01)	.01(.00)	.01(.00)
		60/20	1.00	1.00	1.00	1.00(.00)	.01(.01)	.01(.00)	.01(.00)	1.00(.00)	.01(.01)	.01(.00)	.01(.00)
		60/40	1.00	1.00	1.00	1.00(.00)	.00(.01)	.01(.00)	.01(.00)	1.00(.00)	.01(.01)	.01(.00)	.01(.00)
	L	20/20	.98	.87	.97	1.00(.00)	.02(.02)	.02(.01)	.02(.00)	1.00(.00)	.02(.02)	.02(.01)	.02(.00)
		20/40	1.00	.99	1.00	1.00(.00)	.01(.01)	.01(.01)	.01(.00)	1.00(.00)	.01(.01)	.01(.01)	.01(.00)
		40/20	1.00	1.00	1.00	1.00(.00)	.01(.01)	.01(.00)	.01(.00)	1.00(.00)	.01(.01)	.01(.00)	.01(.00)
		40/40	1.00	1.00	1.00	1.00(.00)	.01(.01)	.01(.01)	.01(.00)	1.00(.00)	.01(.01)	.01(.00)	.01(.00)
		60/20	1.00	1.00	1.00	1.00(.00)	.01(.01)	.01(.00)	.01(.00)	1.00(.00)	.01(.01)	.01(.00)	.01(.00)
		60/40	1.00	1.00	1.00	1.00(.00)	.00(.01)	.01(.00)	.01(.00)	1.00(.00)	.00(.01)	.01(.00)	.01(.00)

Note. ICC = intraclass correlation; CN= number of clusters per group; CS= cluster size; S=small; M=medium; L=large; AIC=Akaike information criterion; BIC=Bayesian information criterion; SaBIC=sample-size adjusted BIC; CFI=Comparative Fit Index; sd= standard deviation; RM=root mean square error of approximation; SRb= standardized root mean square residual for the between level; SRw= standardized root mean square residual for the within level.

Table 11. The estimated values of main covariates effects and between-level factor mean of correct and misspecified models with within-level covariates

		Within-level covariates interaction											
		Covariates interaction of .30						Covariates interaction of .60					
		Correct model			Misspecified model			Correct model			Misspecified model		
ICC	CN/CS	M1	M2	fb	M1	M2	fb	M1	M2	fb	M1	M2	fb
Small	40/10	.30	.40	-.20	.45	.55	-.27	.30	.39	-.20	.60	.70	-.35
	40/20	.30	.40	-.20	.45	.55	-.27	.30	.40	-.20	.60	.70	-.35
	80/10	.30	.40	-.20	.45	.55	-.27	.30	.40	-.20	.60	.70	-.35
	80/20	.30	.40	-.20	.45	.55	-.27	.30	.40	-.20	.60	.70	-.35
	120/10	.30	.40	-.20	.45	.55	-.27	.30	.40	-.20	.60	.70	-.35
	120/20	.30	.40	-.20	.45	.55	-.28	.30	.40	-.20	.60	.70	-.35
Medium	40/10	.30	.40	-.20	.45	.55	-.28	.30	.40	-.20	.60	.70	-.36
	40/20	.30	.40	-.20	.45	.55	-.27	.30	.40	-.20	.60	.70	-.35
	80/10	.30	.40	-.20	.45	.55	-.27	.30	.40	-.20	.60	.70	-.35
	80/20	.30	.40	-.20	.45	.55	-.27	.30	.40	-.20	.60	.70	-.35
	120/10	.30	.40	-.20	.45	.55	-.28	.30	.40	-.20	.60	.70	-.34
	120/20	.30	.40	-.20	.45	.55	-.28	.30	.40	-.20	.60	.70	-.35
Large	40/10	.30	.40	-.20	.45	.55	-.28	.30	.40	-.20	.60	.70	-.34
	40/20	.30	.40	-.20	.45	.55	-.28	.30	.40	-.20	.60	.70	-.35
	80/10	.30	.40	-.20	.45	.55	-.27	.30	.40	-.20	.60	.70	-.35
	80/20	.30	.40	-.20	.45	.55	-.27	.30	.40	-.20	.60	.70	-.35
	120/10	.30	.40	-.20	.45	.55	-.27	.30	.40	-.20	.60	.70	-.35
	120/20	.30	.40	-.20	.45	.55	-.28	.30	.40	-.20	.60	.70	-.35

Note. ICC = intraclass correlation; CN= number of clusters per group; CS= cluster size; M1=main effect of covariate 1 (.30); M2=main effect of covariate 2 (.40); fb= between-level factor mean.

be .50, which was .40 in the correct model; the between-level main effect was overestimated to be .60, which was .50 in the correct model. When the omitted interaction effect was at .40, the within-level main effect and the between-level main effect were overestimated to be .60 and .70, respectively, each receiving half of the magnitude of the interaction effect. Apart from the two main effects, the between-level factor residual variance, and the between-level intercepts of the six indicators were also affected. The between-level factor residual variance was underestimated. In the true model, it was around .05, .15, and .30 for the small, medium, and large ICC conditions, respectively. In the misspecified model, the between-level factor residual variance was underestimated around .02, .04, and .07 for the small, medium, and large ICCs, regardless of the size of the interaction effect. The severity of the underestimation became slightly worse as the sample size increased. The between-level intercepts of the six indicators were also underestimated. In the true model, they were estimated to be around -.20, but were underestimated to be around -.25, and -.30 for covariates interaction effect of .20 and .40, respectively. All the between-level intercepts of indicators were affected in a similar manner, thus, only indicator 3 was used as an example of the bias in intercepts of indicators. Note that the design factors in the simulation study, including cluster number, cluster size, and ICC, had little impact on the bias of parameter estimates in the misspecified models except between-level factor residual variance, as shown in Table 13.

Sources of Variance in the Sensitivity of AIC, BIC, and SaBIC

The proportions of replications that AIC, BIC, and SaBIC selected the correct model varied across the conditions. Eta-squared (η^2) was calculated to evaluate the proportion of

Table 12. The proportion of the replications that AIC, BIC, SaBIC selected the correct model with cross-level covariates

		Cross-level covariates interaction					
		Covariates interaction of .20			Covariates interaction of .40		
ICC	CN/CS	AIC	BIC	SaBIC	AIC	BIC	SaBIC
Small	40/10	.23	.07	.19	.54	.30	.49
	40/20	.16	.05	.11	.57	.36	.49
	80/10	.20	.06	.14	.68	.46	.61
	80/20	.11	.03	.06	.72	.53	.65
	120/10	.20	.07	.13	.76	.58	.69
	120/20	.08	.02	.05	.78	.62	.71
Medium	40/10	.28	.09	.23	.59	.35	.54
	40/20	.23	.07	.15	.65	.40	.58
	80/10	.26	.08	.17	.74	.50	.65
	80/20	.16	.06	.11	.80	.61	.73
	120/10	.24	.08	.16	.81	.61	.74
	120/20	.12	.04	.07	.86	.75	.81
Large	40/10	.31	.10	.25	.65	.40	.60
	40/20	.24	.08	.18	.71	.48	.63
	80/10	.32	.10	.26	.79	.55	.71
	80/20	.23	.08	.16	.86	.68	.79
	120/10	.30	.12	.21	.84	.64	.78
	120/20	.18	.07	.11	.90	.78	.85

Note. ICC = intraclass correlation; CN= number of clusters per group; CS= cluster size; S=small; M=medium; L=large; AIC=Akaike information criterion; BIC=Bayesian information criterion; SaBIC=sample-size adjusted BIC.

variance in the sensitivity of AIC, BIC, and SaBIC that the design factors and their interactions accounted for. As shown in Table 14, when omitting the between-level covariates interaction

Table 13. The estimated values of main covariates effects, intercept, and between-level factor residual variance of correct and misspecified models with cross-level covariates

			Cross-level covariates interaction							
			Correct model				Misspecified model			
effect	ICC	CN/CS	MB	MW	Intercept	fb_v	MB	MW	Intercept	fb_v
.20	S	40/10	.48	.39	-.19	.05	.60	.50	-.25	.03
		40/20	.49	.40	-.19	.05	.60	.50	-.25	.02
		80/10	.49	.40	-.20	.05	.60	.50	-.25	.02
		80/20	.50	.40	-.20	.06	.60	.50	-.25	.02
		120/10	.49	.40	-.20	.05	.60	.50	-.25	.02
		120/20	.50	.40	-.20	.06	.60	.50	-.25	.01
	M	40/10	.48	.39	-.18	.13	.59	.50	-.24	.06
		40/20	.49	.40	-.19	.14	.60	.50	-.25	.04
		80/10	.50	.40	-.20	.14	.60	.50	-.25	.04
		80/20	.50	.40	-.20	.15	.60	.50	-.25	.03
		120/10	.50	.40	-.20	.15	.60	.50	-.25	.03
		120/20	.50	.40	-.20	.15	.60	.50	-.25	.03
	L	40/10	.48	.39	-.19	.29	.59	.49	-.24	.09
		40/20	.48	.40	-.19	.29	.59	.50	-.25	.08
		80/10	.49	.40	-.19	.30	.59	.50	-.25	.07
		80/20	.49	.40	-.20	.30	.59	.50	-.25	.06
		120/10	.50	.40	-.20	.30	.60	.50	-.25	.05
		120/20	.50	.40	-.20	.31	.60	.50	-.25	.05
.40	S	40/10	.48	.40	-.19	.05	.70	.60	-.30	.03
		40/20	.50	.40	-.20	.05	.70	.60	-.30	.02
		80/10	.49	.40	-.20	.05	.70	.60	-.30	.02
		80/20	.50	.40	-.20	.05	.70	.60	-.30	.02
		120/10	.50	.40	-.20	.05	.70	.60	-.30	.02
		120/20	.49	.40	-.20	.06	.70	.60	-.30	.01
	M	40/10	.48	.39	-.19	.14	.70	.60	-.30	.06
		40/20	.49	.40	-.20	.14	.70	.60	-.30	.04
		80/10	.50	.40	-.20	.14	.70	.60	-.30	.04
		80/20	.50	.40	-.20	.15	.70	.60	-.30	.03
		120/10	.50	.40	-.20	.14	.70	.60	-.30	.03
		120/20	.50	.40	-.20	.15	.70	.60	-.30	.03
	L	40/10	.47	.39	-.18	.29	.69	.60	-.29	.09
		40/20	.50	.40	-.20	.30	.70	.60	-.31	.08
		80/10	.49	.40	-.19	.30	.70	.60	-.29	.07
		80/20	.50	.40	-.20	.31	.70	.60	-.30	.06
		120/10	.50	.40	-.20	.30	.70	.60	-.30	.05
		120/20	.50	.40	-.20	.31	.70	.60	-.30	.05

Note. ICC = intraclass correlation; CN= number of clusters per group; CS= cluster size; S=small; M=medium; L=large; MB=between-level covariate main effect (.50); Mw=within-level covariate main effect (.40); fb_v=between-level factor residual variance.

effect, the magnitude of the interaction effect, ICC, and cluster number explained around 54%, 21%, and 20% for AIC, respectively; 55%, 23%, and 13% of BIC, respectively, and 62%, 24%, and 10% for SaBIC, respectively. When the omitted covariates interaction effect was at the within level, magnitude of interaction effect, cluster number, cluster size, interaction between magnitude and cluster number, interaction between magnitude and cluster size were statistically significantly associated with the sensitivity of AIC, BIC, and SaBIC. Among these factors or interactions, the magnitude of the interaction effect explained the most proportion of the variance in the three fit indices. When the cross-level covariates interaction effect was omitted, the magnitude of the covariates interaction effect explained around 90% of the variance in the sensitivity of the three fit indices. All the other factors were not significant for AIC and SaBIC, but the interaction between the magnitude and the cluster number was significant for BIC.

Discussion

The purpose of Study Two was to examine the impact of omitting covariates interaction in multilevel MIMIC model on fit indices (RMSEA, CFI, TLI, SRMR-W, SRMR-B, AIC, BIC, and SaBIC) and parameter estimates in the model. It purported to enhance the understanding of the sensitivity of SEM based fit indices and other fit indices to omitting a parameter, and of the impact on other parameters in the model. The simulation results indicated that none of the SEM based fit indices (RMSEA, CFI, TLI, SRMR-W, and SRMR-B) detected the misspecified model in which the covariates interaction effect was omitted. All these fit indices produced in the misspecified model were identical to those in the correct model. Based on these fit indices, the misspecified model was concluded to fit the data very well, even though the existing covariates

Table 14. Eta-squared (η^2) of different sources for proportions of replications of AIC, BIC, and SaBIC that selected the correct model of between-level, within-level, and cross-level covariates interaction

	between Level fit indices			within level fit indices			cross level fit indices		
	AIC	BIC	SaBIC	AIC	BIC	SaBIC	AIC	BIC	SaBIC
Overall η2	96.89%	99.04%	98.05%	97.59%	98.83%	99.24%	99.90%	99.75%	99.87%
magnitude	**53.64%**	**54.60%**	**62.08%**	**31.21%**	**54.67%**	**46.51%**	**90.06%**	**85.15%**	**90.38%**
ICC	**20.52%**	**23.47%**	**23.74%**	0.08%	0.02%	0.06%	2.59%	1.59%	2.29%
CN	**20.43%**	**12.60%**	**10.37%**	**22.39%**	**17.67%**	**18.29%**	1.40%	4.97%	1.24%
CS	0.84%	0.13%	0.01%	**11.79%**	**10.68%**	**10.68%**	0.15%	0.37%	0.10%
magnitude*ICC	0.79%	4.36%	0.56%	0.07%	0.04%	0.06%	0.02%	0.34%	0.10%
magnitude*CN	0.37%	3.01%	0.87%	**18.76%**	**9.46%**	**14.40%**	3.91%	**5.75%**	4.24%
ICC*CN	0.11%	0.72%	0.20%	0.06%	0.01%	0.04%	0.03%	0.01%	0.04%
ICC*CS	0.10%	0.14%	0.21%	0.03%	0.01%	0.09%	0.02%	0.07%	0.03%
magnitude*CS	0.09%	0.02%	0.00%	**9.88%**	**6.19%**	**8.44%**	1.67%	1.46%	1.43%
CN*CS	0.01%	0.00%	0.02%	3.32%	0.09%	0.67%	0.06%	0.02%	0.02%

Note. ICC = intraclass correlation; CN= number of clusters per group; CS= cluster size; AIC=Akaike information criterion; BIC=Bayesian information criterion; SaBIC=sample-size adjusted BIC.

interaction effect was not specified in the model. Thus, obtaining very good fit indices didn't guarantee that the specified model was a very good fit to the data.

According to Ryu and West (2009), CFI and RMSEA produced in traditional standard method failed to detect model misspecification that occurred at the between level, but succeeded to detect model misspecification in the within-level model. The simulation results in this study were consistent with their results with respect to the misspecification at the between level (failure to detect the between-level misspecificaition), but conflicted with the results of their study with respect to misspecification at the within level (CFI and RMSEA in their study were sensitive to the within-level model misspecification, whereas they were not in this study). Note that the form of misspecification in their study was different from the misspecification form in this study. They misspecified the correlation between two factors, whereas in this study, the covariates interaction effect was omitted. The sensitivity of the fit indices may vary as a result of different forms of the misspecification. Hsu and colleagues (2015) conducted simulation studies to examine the common fit indices in detecting missepecified multilevel SEMs, and the results showed that CFI, RMSEA, and TLI could detect misspecification only in the within-level model but not misspecification in the between-level model; SRMR-B was the only fit index sensitive to misspecification in the between-level model. However, in our study, SRMR-B was not sensitive to the omission of covariates interaction in the between-level model. One explanation for the insensitivity of fit indices in multilevel SEMs was that the misspecification in one parameter (constraining the covariates interaction effect to be zero in this study) might be manifested as biased estimates of other related parameters without substantially changing the overall sample mean and covariance matrices (Gerbing & Anderson, 1993; Tomarken & Waller, 2003). When sample mean and covariance matrices were not affected much by the model misspecification, the

χ^2 statistic produced by the misspecified model was not very different from the χ^2 of the true model. Therefore, fit indices, such as CFI, RMSEA, TLI, which are a function of the overall model χ^2 test statistic, were not sensitive to the model misspecification.

The previous research about the sensitivity of fit indices to model misspecification focused on the misspecification of pattern coefficients (factor loadings) or factor covariance (e.g., Hsu et al., 2015; Hu & Bentler, 1998; Ryu & West, 2009). Misspecification of these forms had an effect on the sample covariance matrices, thus impacting the fit indices. On the other hand, omitting a covariates interaction effect in multilevel MIMIC had a less severe impact on the sample covariance matrices, and the fit indices were not sensitive enough to capture misspecification of this form. Therefore, the fit indices in SEM have their limitations in detecting model misspecification of a variety of forms. The fit indices are sensitive to some forms of model misspecification but not the other types. Good fit indices based on the gold rule criteria cannot guarantee that the relations of the variables in the model are correctly modeled.

In the cross-level covariates interaction model (two-level random slope modeling), no SEM based fit indices (i.e., CFI, RMSEA, TLI, and SRMR) were produced in the output. Instead, AIC, BIC, and SaBIC were used to make model comparison. In the between-level and within-level covariates interaction model, AIC, BIC, and SaBIC were also available. The performance of AIC, BIC, and SaBIC differed as a result of the location and size of the covariates interaction effect. When the between-level covariates interaction was omitted in the model, the proportions of replications that AIC, BIC, and SaBIC selected the correct model increased as sample size increased. The performance of the three fit indices also depended on the ICC (lower proportions as ICC increased), similar to the power of testing the covariates interaction effect. As expected, the larger the omitted covariates interaction effect was, the more sensitive the three fit indices were. In the conditions with small ICC and larger sample size, AIC and SaBIC selected the correct model in all replications. AIC performed better than SaBIC, which performed better than BIC. AIC, BIC, and SaBIC were more sensitive to the omission of the within-level covariates interaction effect than to the omission of the covariates interaction effect. When the omitted interaction effect reached .60, the three fit indices selected the correct model in almost 100% of the replications except the conditions with the smallest sample size. When the omitted within-level covariates interaction was .30, the sensitivity of three fit indices depended on sample size. When sample size was small, AIC was more sensitive than SaBIC, which was more sensitive than BIC; however, when sample size was big, the difference between the three fit indices was very small. ICC had no impact on the sensitivity of the three fit indices. The sensitivity of AIC, BIC, and SaBIC in omitting cross-level covariates interaction effect depended majorly on the magnitude of the interaction effect. When the omitted interaction effect was .20, a very small percentage (below 32%) of the replications selected the correct model, especially BIC. When the omitted covariates interaction was .40, the sensitivity of the three fit indices improved. A bigger proportion of replication that selected the correct model was associated with bigger sample size, and smaller ICC. Generally, AIC, SaBIC performed better than BIC in selecting the correct model. Also, when the covariates interaction effect was at the within level, AIC, BIC, and SaBIC were more sensitive to the misspecification than that was at the between-level and cross-level.

Compared to the large body of literature into the sensitivity of fit indices of SEM (e.g., Fan & Sivo, 2005, 2007; Hsu et al., 2015; Hu & Bentler, 1998, 1999; Ryu & West, 2009), only a small number of studies (Ferron, Dailey, & Yi, 2002; Enders & Tofighi, 2009; Kwok, West, & Green, 2007) examined the impact of model misspecification on estimates of other parameters in the model. The misspecification in these three studies were in the form of misspecifying the residual variances in two-level growth model. So far no studies have explored the impact of omitting covariates interaction effect in multilevel MIMIC models. The simulation results in this study indicated that the omitted interaction effect was manifested as biased estimates of the two main effects regardless of the location of the covariates interaction. It was of interest that each of the two main effects was overestimated by approximately half of the magnitude of the covariates interaction effect. Note that the direction of the bias in the two main effects was related to the direction of the omitted covariates interaction effect. The overestimation of the two main effects was due to the positive direction of the covariates interaction effect in the simulated data. If negative covariates interaction effect was omitted, the two main effects would be underestimated. Meanwhile the between-level factor mean was underestimated, and the underestimation varied as a function of the size of the interaction effect, when the covariates interaction was at the between level or at the within level. In addition to overestimated main effects, the between-level factor residual variance (no estimation of between-level factor mean was produced in the random slope modeling) and intercepts of the six indicators were underestimated when the covariates interaction was at the cross level. When necessary parameters are omitted from the model, not all parameters in the model are affected, but only affecting certain parameters that are closely related to the misspecification (Yuan & Marshall, 2003). Thus, in this study, only the parameters that were closely related to the omitted interaction effect were affected, including the two main effects of the two covariates and the between-level factor mean. When the cross-level interaction effect was omitted, the between-level factor residual variance and indicator intercepts were also affected, apart from the affected two main effects. In this two-level random slope modeling, the between-level factor mean was not estimated, and maybe that was why the indicators' intercepts were affected. In the between-level and within-level covariates interaction models, the omitted covariates interaction effect resulted in an underestimation of the between-level factor mean of around -.07 and -.15 for interaction effect of .30 and .60, respectively. When the cross-level covariates interaction effect was omitted, the six indicators' intercepts were underestimated by -.05 and -.10 for interaction effect of .20 and .40, respectively. The degree of bias in the between-level factor mean or the between-level indicators' intercepts was comparable given that the covariates interaction effect was different for different locations of covariates (.30 and .60 for the between-level and within-level covariates interaction effects; .20 and .40 for the cross-level covariates interaction effects).

The affected parameters were intuitive based on the path analysis proposed by Yuan, Marshall, and Bentler (2003). Using the path analysis, it was expected that parameters mentioned above (i.e., main effects and between-level factor mean) were to be affected, as they were closely related to the misspecification. However, to what degree they were to be affected was not certain. In this study, it showed that the two main effects of the two covariates carried approximately half of the interaction effect. The other two parameters affected by the omission of the interaction effect were the between-level factor mean and between-level factor residual variance when the two covariates were both at the between level; the between-level factor mean and within-level factor residual variance when the two covariates were both at the within level. The between-level factor mean was underestimated by approximately .08 and .15, when the omitted covariates interaction effect was at .30 and .60, respectively, in the conditions of the two covariates being at the between level or at the within level. When the omitted covariates interaction was at the between level, the between-level factor residual variance was overestimated by approximately .01 and .02, respectively, for omitted .30, and .60 interaction effect, in all conditions examined. When the omitted covariates interaction was at the within level, the within-level factor residual variance was overestimated by approximately .01 in some conditions for omitted .30 interaction effect, and was overestimated by .02 in all conditions for omitted .60 interaction effect. In sum, all the parameters affected by the omitted covariates interaction effect were closely related to the misspecification.

Chapter Five: Conclusions and Suggestions

Conclusions

The MIMIC model has been known for its flexibility in modeling multiple covariates in latent variable modeling, and the covariates can be continuous or categorical. Thus, it has been utilized frequently by methodological and empirical researchers in latent variable modeling. The extension of ML MIMIC with multilevel data enables the modeling of covariates from both the within level and the between level. It is common in educational and psychological research that the effect of some covariates on the latent factor are conditional on other covariates, that is, covariates interaction effect. However, there is a lack of research into the performance of ML MIMIC modeling in detecting the covariates interaction effect and the impact of omitting the existing covariates interaction effect. Two simulation studies were conducted to investigate the performance of ML MIMIC models in detecting covariates interaction effect and the impact of omitting the interaction effect. Study One aimed to examine the performance of ML MIMIC in detecting covariates interaction effect when the covariates interaction effect was at the between level, at the within level, and at the cross level, respectively. Study Two purported to examine the sensitivity of SEM fit indices (CFI, TLI, RMSEA, SRMR-W, and SRMR-B) and other fit indices, like AIC, BIC, and SaBIC, to the omission of the covariates interaction effect, as well as the impact of the omission on the estimates of other parameters in the model.

Simulation results of Study One showed that the performance of ML MIMIC in detecting the covariates interaction effect varied as a result of the location of the covariates interaction effect. Also, the impact of the design factor on the performance of ML MIMIC varied depending on the location of the covariates interaction effect. When the covariates interaction was at the between level, ASR was above 90% in most of the conditions except the large ICC combined with smaller sample size. Type I error rate was under control in all conditions. Power of detecting the covariates interaction effect varied substantially as a function of the magnitude of the covariates interaction effect, ICC level, and the sample size, especially cluster number. When the interaction effect was at .30, power ranged from .14 (large ICC combined with the smallest sample size) to .75 (small ICC combined with the largest sample size). As the covariates interaction effect increased to .60, power increased considerably, particular in small and medium ICC conditions, ranging from .72 to 1.00, from .49 to .97, and from .35 to .80 for small, medium, and large ICCs. Bias and relative bias of the covariates interaction effect were negligible in most conditions. The impact of cluster number on power was larger than that of cluster size. Standard error of parameter estimate and RMSE had very similar pattern, increasing as ICC became larger controlling for the other design factors, and decreasing as sample size increased. Standard error was similar for the two different magnitudes of the covariates interaction effect, whereas RMSE was smaller in the larger interaction effect of .60.

When the covariates interaction was at the within level, the performance of ML MIMIC was very decent. ASR was close to 100% in all conditions. Type I error rate was similar to that of the between-level interaction conditions. Power in detecting the covariates interaction effect ranged from .46 to .99 for the covariates interaction of .30, and was almost 1.00 in all conditions except the smallest sample size for the covariates interaction of .60. Power increased as the sample size became bigger, and ICC seemed to have no impact on power. Standard error of parameter estimate and RMSE decreased as sample size increased, and they were similar in the two different magnitudes of the covariates interaction effect.

For the cross-level covariates interaction effect conditions, the performance of ML MIMIC was adequate even with a small interaction effect of .20. Power was impacted only by total sample size, but not ICC level or cluster number. Relative bias was negligible except the smallest sample size conditions. When the interaction effect increased to .40, power ranged from .61 to 1.00. In all conditions power seemed to be impacted only by the total sample size, and so were the other outcome variables.

In sum, ML MIMIC had decent performance in detecting and estimating covariates interaction effect when the covariates were at the within level and at the cross level. However, when the covariates were at the between level, a large ICC and smaller number of cluster number compromised the performance of ML MIMIC. When ICC was small and cluster number was big, the performance of ML MIMIC was adequate. This finding provides useful implications for empirical researchers who are interested in estimating the covariates interaction effect in ML MIMIC. The location of the interaction effect, the ICC level, and the total sample size (especially cluster number) available render the research understand whether the chosen method was adequate or not. When the two covariates are at the between level, applied researchers need to check ICC level and cluster number before proceeding to estimating covariates interaction effect in ML MIMIC. However, when the two covariates are at the within level or cross level, researchers need to make sure that the total sample size is big enough to achieve decent power of detecting covariates interaction effect.

Simulation results in Study Two indicated that none of the SEM fit indices (CFI, TLI, RMSEA, SRMR-W, and SRMR-B) was sensitive to the omission of the covariates interaction effect, irrespective of the location of the covariates interaction effect. The insensitivity of these fit indices was partly because that the omitted interaction effect was carried through the other parameters in the model and the total variance-covariance matrices did not change substantially. This is similar to the equivalent or close to equivalent models in SEM, where structural relationship between variables (conceptual interpretation) is different but model fit is equivalent or close to equivalent, making it difficult to distinguish empirically which model is correct. These fit indices have been commonly used by applied researchers to evaluate SEM model fit. When all these fit indices are more than adequate based on the golden rule criteria, the applied researchers feel relieved that their models have good fit to the data. However, the simulation results in Study Two showed that very good fit indices did not guarantee that all the significant relationships between variables have been modeled correctly in the model. A missing interaction effect, which resulted in biased estimates of some other parameters in the model, was not detected by the fit indices.

The consequences of the insensitivity of fit indices to a missing covariates interaction effect pose serious issues for applied researchers to make references about the relationship between variables in the data. It is recommended that applied researchers use theoretical framework in specific research areas to guide the modeling of the variables in the model to avoid omitting some important effects between some variables. If no established theory has been in place, more explorations into the relationship between the variables are suggested to have a more complete understanding of the data. It is important to run both models with and without interaction effects. Because power of testing interaction effect was fairly decent at least at the within level and cross level, we can detect the presence of interaction effect by testing statistical significance of the interaction effects. Moreover, the estimate was also unbiased, and the magnitude of the interaction effect can be estimated accurately. For methodologist, it would be helpful to develop different but potentially more sensitive methods for detecting omitted interaction effect.

When the omitted interaction was small, none of AIC, BIC, and SaBIC selected the correct model consistently for the three scenarios of covariates interaction effect. On the other hand, when the omitted interaction effect reached .60, the three fit indices performed much better, especially for the omitted within-level covariates interaction effect (AIC, BIC, and SaBIC selected the correct model almost in all conditions). In sum, AIC performed better than SaBIC, which outperformed better than BIC. BIC(J) performed better than BIC in the model of omitting between-level interaction effect.

The biased parameter estimates in the model resulted from the model misspecification included the covariates main effects and the between-level factor mean. Of note that when between-level factor mean was not produced in the cross-level covariates modeling, all indicator intercepts were underestimated. Interestingly, the two main effects were overestimated by approximately half of the size of the covariates interaction effect. Between-level factor mean was underestimated. The underestimation of the between-level factor mean might be due to the overestimation of the two main effects. Of note that the direction of bias in between-level factor mean depended on the way in which the interaction effect was simulated (a positive interaction effect in this study). If a negative covariates interaction effect is simulated, the two main effects are to be underestimated by receiving half of the magnitude of the interaction effect, and I speculate that the between-level factor mean may be overestimated, correspondingly. This speculation was substantiated by another simulation in which the covariates interaction effect was negative. In the misspecified cross-level covariates interaction effect model, between-level factor residual variance was underestimated, in addition to the affected parameters mentioned above. The other parameters in the model were not affected much by the omission of the covariates interaction effect. Applied researchers need to interpret the main effects of the covariates and factor mean in ML MIMIC with caution if they do not rule out the possibility of covariates interaction effect in the model.

Limitations and Future Research Direction

The two studies have limitations that we should be aware of. First, in the ML MIMIC model, the measurement CFA part at both the between level and within level had good measurement quality with all factor loadings of the six indicators at .80. In real research scenarios it is too ideal to have factor loadings of all indicators so high. The impact of varying measurement quality (e.g., lower factor loadings) on the performance of ML MIMIC in detecting covariates interaction effect remains unknown. Future research on the performance of ML MIMIC in estimating covariates interaction effect with varying measurement quality of the CFA measurement part is encouraged.

Second, a limited number of manipulated factors were considered in these two studies. Some other parameters in the model were fixed at a common value for those parameters. The two main effects were fixed. Varying main effects to have different ratio of the main effects to the covariates interaction effect may impact the performance of ML MIMIC. Moreover, all the clusters had the same cluster size. In real research scenario, it is more common to have unbalanced design. More design factors can be included in the simulation to have a more comprehensive understanding of the performance.

Third, strict invariance in terms of the grouping covariates and strong measurement cluster invariance (identical cross-level factor loadings and intercepts) was assumed in ML MIMIC in the simulation studies. The violation of strong measurement invariance (e.g., noninvariant intercepts) in ML MIMIC may impact the performance of ML MIMIC in estimating covariates interaction effect. Future research on the impact of violating measurement invariance in ML MIMIC is called for.

Fourth, the fit indices examined in Study Two included SEM based fit indices, AIC, BIC, BIC(J), and SaBIC. Given the nested nature of the correct model and the misspecified model, likelihood-ratio test can be used to evaluate its performance in detecting the misspecified model.

Lastly, previous study showed the sensitivity of traditional SEM fit indices (i.e., CFI, RMSEA, TLI, SRMR) to the within level model misspecification (Hsu et al., 2015; Ryu & West, 2009). Of note that in their study model misspecification took the form of misspecified factor loadings and factor covariance. A quantification of the misspecification severity (Fan & Sivo, 2005) of omitting covariates interaction effect is encouraged to compare the degree of model misspecifications of different studies. Thus, a more profound understanding of the sensitivity of SEM based fit indices is possible.

References

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, *19*, 716-723.

Aiken, L.S. & West, S.G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: Sage.

Allua, S., Stapleton, L. M., & Beretvas, S. N. (2008). Testing latent mean differences between observed and unobserved groups using multilevel factor mixture models. *Educational and Psychological Measurement, 68*, 357-378.

Bauer, D. J., & Curran, P. J. (2005). Probing Interactions in Fixed and Multilevel Regression: Inferential and Graphical Techniques. *Multivariate Behavioral Research*, *40*, 373-400.

Bentler, P. M., & Chou, C, P. (1986). Statistics for Parameter Expansion and Contraction in Structural Models. Presented at the annual meeting of the American Educational Research Association, San Francisco, CA: Sage.

Bentler, P. M., & Chou, C. P. (1993). Some new covariance structure model improvement statistics, In K. A. Bollen & J. S. Long (Eds), *Testing structural equation models* (pp. 235-255). Newburry Park, CA: Sage.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.

Booen, K. A., & Long, J. S. (1993). Introduction. In K. A. Bollen & J. S. Long (Eds), *Testing structural equation models* (pp. 1-9). Newbury Park. CA: Sage.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds), *Testing structural equation models* (pp. 136-162). Newbury Park. CA: Sage.

Chen, C. (1981). The EM Approach to the Multiple Indicators and Multiple Causes Model via the Estimation of the Latent Variable. *Journal of the American Statistical Association, 76*. 704-708.

Cheng, Y., Shao, C., & Lathrop, Q. N. (2016). The Mediated MIMIC Model for Understanding the Underlying Mechanism of DIF. *Educational And Psychological Measurement*, *76,* 43-63.

Chun, S., Stark, S., Kim, E. S., & Chernyshenko, O. S. (2016). MIMIC methods for detecting DIF among multiple groups: Exploring a new sequential-free baseline procedure. *Applied Psychological Measurement*, *40*, 486-499

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement, 33*, 107-112,

Condon, P. (2010). A multiple indicator, multiple cause method for representing social capital with an application to psychological distress. *Journal of Geographical Systems, 12,* 1-23.

Davidov, E., Dulmer, H., Schluter, E., Schmidt, P., & Meuleman, B. (2012). Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance. *Journal of Cross-Cultural Psychology*, *43*, 558-575.

Enders, C. K., & Tofighi, D. (2008). The impact of misspecifying class-specific residual variances in growth mixture models. *Structural Equation Modeling: A Multidisciplinary* Journal, *15*, 75-95.

Fan, X., & Sivo, S. A. (2005). Sensitivity of Fit Indexes to misspecified structural or measurement model component: Rationale of two-index strategy *Structural Equation Modeling: A Multidisciplinary* Journal, *12*, 343-367.

Fan, X., & Sivo, S. A. (2007). Sensitivity of Fit Indices to Model Misspecification and Model Types. *Multivariate Behavioral Research*, *42*, 509-529.

Fan, X., Thompson, B., & Wang, L. (1999). The effects of sample size, estimation methods, and model specification on SEM fit indices. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 56–83.

Fan, X., & Wang, L. (1998). Effects of potential confounding factors on fit indices and parameter estimates for true and misspecified SEM models. *Educational and Psychological Measurement*, *58*, 699–733.

Farley, J. U., & Reddy, S. K. (1987). A factorial evaluation of effects of model specification and error on parameter estimation in a structural equation model. *Multivariate Behavioral Research*, *22*, 71-90.

Ferron, J., Dailey, R., & Yi, Q. (2002). Effects of misspecifying the first-level error structure in two-level models of change. *Multivariate Behavioral Research*, *37*, 379-403.

Finch, H. (2005). The MIMIC model as a method for detecting DIF: Comparing with Mantel-Haenszel, SIBTEST, and IRT likelihood ratio. *Applied Psychological Measurement, 29*, 278-295.

Finch, W. H., & French, B. F. (2011). Estimation of MIMIC model parameters with multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*, 229-252.

Fleishman, J., Spector, W., & Altman, B. (2002). Impact of differential item functioning on age and gender differences in functional disability. *Psychological Sciences and Social Sciences, 57*, 275–284.

Gallini, J. K. (1983). Misspecifications that can result in path analysis structures. *Applied Psychological Measurement, 7*, 125–37.

Gerbing, D. W. & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K. A. Bollen & J. S. Long, (Eds.), *Testing structural equation models* (pp. 40–59). Newbury Park, CA: SAGE.

Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika, 66*, 373-388.

Hancock, G. R., Lawrence, F. R., & Nevitt, J. (2000). Type I error and power of latent mean methods and MANOVA in factorially invariant and noninvariant latent variable systems. *Structural Equation Modeling: A Multidisciplinary Journal*, *7*, 534-556.

Harring, J. R., Weiss, B. A., & Li, M. (2015). Assessing spurious interaction effects in structural equation modeling: A cautionary note. *Educational and Psychological Measurement*, *75,* 721-738.

Heck R, & Thomas S. (2000). *An introduction to multilevel modeling techniques*. Mahwah, NJ: Erlbaum.

Heck R, &Thomas S. (2015). *An introduction to multilevel modeling techniques*. Routledge, NY: New York.

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

Hox, J. J. (1998). Multilevel modeling: When and why. In I. Balderjahn, R. Mathar & M. Schader (Eds.). *Classification, data analysis, and data highways* (pp. 147-154). New York: Springer Verlag.

Hox, J. J. (2010). *Multilevel Analysis: Techniques and Applications*. New York (2nd ed.): Routledge.

Hox, J. J., & Mass, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling: A Multidisciplinary Journal, 8,* 157-174.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424–453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6,* 1–55.

Hsu, H., Kwok, O., Lin, J. H., & Acosta, S. (2015). Detecting misspecified multilevel structural equation models with common fit indices: A Monte Carlo study. *Multivariate Behavioral Research, 50*, 197-215.

Jak, S., Oort, F. J., & Dolan, C. V. (2014). Measurement bias in multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 31-39

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 70,* 631-639.

Kaplan, D. (1988). The impact of specification error on the estimation, testing, and improvement of structural equation models. *Multivariate Behavioral Research*, *23*, 69–86.

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, *96*, 201–210.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*, 333–351.

Kim, E. S., & Cao, C. (2015). Testing group mean differences of latent variables in multilevel data using multiple-group multilevel CFA and multilevel MIMIC modeling. *Multivariate Behavioral Research, 50,* 436-456.

Kim, E. S., Kwok, O., & Yoon, M. (2012). Testing Factorial Invariance in Multilevel Data: A Monte Carlo Study. *Structural Equation Modeling: A Multidisciplinary Journal*, *19,* 250-267.

Kim, E. S., Yoon, M., Wen, Y., Luo, W., & Kwok, O. (2015). Within-Level Group Factorial Invariance with Multilevel Data: Multilevel Factor Mixture and Multilevel MIMIC Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*, 603-616.

Kwok, O., West, S. G., & Green, S. (2007). The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A Monte Carlo study. *Multivariate Behavioral Research, 42*, 557-592.

Lukočienė, O, & Vermunt, J. K. (2010). Determining the Number of Components in Mixture Models for Hierarchical Data, in A. F. L. Berthold, W. Seidel, & A. Ultsch, *Advances in Data Analysis, Data Handling and Business Intelligence* (pp. 241–49). Ultsch. Berlin-Heidelberg: Springer.

Maas, C. J. M., & Hox, J.J. (2005). Sufficient sample size for multilevel modeling. *Methodology, 1*, 86-92.

Marsh, H. W., Balla, J. R., & Hau, K.-T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques*. Mahwah, NJ: Erlbaum.

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*, 275-300.

Marsh, H. W., Wen, A., & Hau, K.T. (2006). Structural equation models of latent interaction and quadratic effects. In G. Hancock & R. O. Mueller (Eds.). *Structural equation modeling: A second course* (pp. 225-265). Greenwich, CT: Information Age.

Mathieu, J. E., Aguinis, H., Culpepper, S. A., & Chen, G. (2012). Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology, 5*, 951.

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods, 10*, 259–284.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525-543.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54*, 557-585.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22,* 376-398.

Muthén, B. O. & Asparouhov, T. (2003). Modeling Interactions between Latent and Observed Continuous Variables Using Maximum-Likelihood Estimation in Mplus. Mplus Web Notes: No. 6.

Muthén, L. K., & Muthén, B. O. (1998–2012). Mplus user's guild (7th ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, B. O., & Satorra, A. (1995). Complex Sample Data in Structural Equation Modeling. *Sociological Methodology*, 25, 267-316.

Ogg, J., McMahan, M. M., Dedrick, R. F., & Mendez, L. R. (2013). Middle school students' willingness to engage in activities with peers with ADHD symptoms: A multiple indicators multiple causes (MIMIC) model. *Journal of School Psychology*, *51*, 407-420.

Pauler, D. K. (1998). The Schwarz Criterion and Related Methods for Normal Linear Models. *Biometrika 85*, 13-27.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods* (2nd Ed.). Thousand Oaks: CA: Sage.

Ryu, E., & West, S. G. (2009). Level-Specific Evaluation of Model Fit in Multilevel Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16,* 583-601.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 6, 461-464.

Sclove, S. L. (1987). Application of model-selection criterion to some problems in multivariate analysis. *Psychometrika*, 52, 333-343.

Smeden, M., & Hessen, D., J. (2013). Testing for two-way interactions in the multigroup common factor model. *Structural Equation Modeling: A Multidisciplinary Journal, 20,* 98-107.

Snijders T. A. B., & Bosker R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage Publications, London.

Sörbom, D. (1975). Detection of correlated errors in longitudinal data. *British Journal of Mathematical and Statistical Psychology, 24,* 138-151.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42*, 893-898.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds), *Testing structural equation models* (pp. 10-39). Newbury Park. CA: Sage.

Thompson, M. S., & Green, S. B. (2006). Evaluating between group differences in latent variable means. In G. Hancock & R. O. Mueller (Eds.). *Structural equation modeling: A second course* (pp. 225-265). Greenwich, CT: Information Age.

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with well fitting models. *Journal of Abnormal Psychology*, 112, 578–98.

Tucker, J. S., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1-10.

Wen, Z., Marsh, H. W., Hau, K., Wu, Y., Liu, H., & Morin, A. S. (2014). Interaction effects in latent growth models: Evaluation of alternative estimation approaches. *Structural Equation Modeling: A Multidisciplinary Journal, 21*, 361-374

Woods, C. M. (2009). Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. *Multivariate Behavioral Research*, *44,* 1-27.

Woods, C. M., & Grimm, K. J. (2011). Testing for nonuniform differential item functioning with multiple indicators multiple cause model. *Applied Psychological Measurement*, *35,* 339-361.

Yuan, K.-H., Marshall, L. L., & Bentler, P. M. (2003). Assessing the Effect of Model Misspecifications on Parameter Estimates in Structural Equation Models. *Sociological Methodology*, *33,* 241-265

Appendix 1: Data Generation SAS Code

Between-level covariates interaction effect of .30 with small ICC

```
/****************************************/
/*            Pattern Matrix            */
/****************************************/;
* specify matrices at the between level;
* step 1: obtain correlation matrix and SDs;
options noxwait xsync;
proc iml;
LY = {.8, .8, .8, .8, .8, .8};
PH = {.1};       **For medium and large ICC,PH is .25 and .50, respectively*;
TE = {.004 0 0 0 0 0,
        0 .004 0 0 0 0,
        0 0 .004 0 0 0,
        0 0 0 .004 0 0,
        0 0 0 0 .004 0,
        0 0 0 0 0 .004}; *TE is .01, and .02 for medium and large ICCs*;
COV = LY*PH*LY` + TE;
*print COV ;
      * obtain correlation matrix and SDs from COV;
S = sqrt(diag(cov));
R = (inv(S))*cov*(inv(S));
 create STDB from S [colname={y1 y2 y3 y4 y5 y6}];
       append from S;
create CR from R [colname={y1 y2 y3 y4 y5 y6}];
     append from R;
print S R;

* step 2: obtain between-level pattern matrix;
data A (type = corr);
_TYPE_ ='CORR';
input y1 y2 y3 y4 y5 y6;
cards;
        1 . . . . .
        0.9411765         1 . . . .
        0.9411765 0.9411765         1 . . .
        0.9411765 0.9411765 0.9411765         1 . .
        0.9411765 0.9411765 0.9411765 0.9411765         1 .
        0.9411765 0.9411765 0.9411765 0.9411765 0.9411765         1

;
*obtain factor pattern matrix for data generation;
proc factor n=6 outstat=facoutB noprint;
data patternB; set facoutB;
 if _TYPE_ = 'PATTERN';
 drop _TYPE_ _NAME_ ;
run;

* specify within-level model;
```

```
* step 1: obtain correlation matrix and SDs;
proc iml;
LYW = {.8, .8, .8, .8, .8, .8};
PHW = {1};
TEW = {.36 0 0 0 0 0,
        0 .36 0 0 0 0,
        0 0 .36 0 0 0,
        0 0 0 .36 0 0,
        0 0 0 0 .36 0,
        0 0 0 0 0 .36};
COVW = LYW*PHW*LYW` + TEW;
*print COVW ;
      * obtain correlation matrix and SDs from COV;
SW = sqrt(diag(covW));
RW = (inv(SW))*covW*(inv(SW));
 create STDW from SW [colname={y1 y2 y3 y4 y5 y6}];
       append from SW;
*print SW RW;

* step 2: obtain within-level pattern matrix;
data B (type = corr);
_TYPE_ ='CORR';
input y1 y2 y3 y4 y5 y6;
cards;
        1 . . . . .
        0.64   1 . . . .
        0.64 0.64   1 . . .
        0.64 0.64 0.64   1 . .
        0.64 0.64 0.64 0.64   1 .
        0.64 0.64 0.64 0.64 0.64   1
;
      *obtain factor pattern matrix for data generation;
proc factor n=6 outstat=facoutW noprint;
data patternW; set facoutW;
 if _TYPE_ = 'PATTERN';
 drop _TYPE_ _NAME_ ;
run;


/***************************************/
/*            Data Generation          */
/***************************************/;
options nonotes;
*libname ml 'U:\mlMIMIC\dataSAS';

%let mc = 1000; *number of replications;

%macro mlmimic;
      * do loop for cluster number;
%do cnloop = 1 %to 3;
 %if &cnloop = 1 %then %do; %let cn = 40; %end;
 %if &cnloop = 2 %then %do; %let cn = 80; %end;
 %if &cnloop = 3 %then %do; %let cn = 120; %end;
            * do loop for cluster size;
      %do csloop = 1 %to 2;
       %if &csloop = 1 %then %do; %let cs = 10; %end;
       %if &csloop = 2 %then %do; %let cs = 20; %end;
```

```
X mkdir "U:\dissertation simulation\between_3\S_icc\cn&cn.cs&cs\";
          * do loop to generate the given number of replications;
    %do i=1 %to &mc;
    proc iml;
    use patternB;
    read all var _NUM_ into F;
    F = F`;
    use stdb;
    read all var _NUM_ into STB;
                * create two groups with interaction (i.e., 4 cells);
     n = &cn/4;
     Y1 = rannor(J(n, 6, 0));
     Y1 = Y1`;
     z1 = F*Y1;
     z1 = STB*z1;
     z1 = z1`;
     Y2 = rannor(J(n, 6, 0));
     Y2 = Y2`;
     z2 = F*Y2;
     z2 = STB*z2 +.4;
     z2 = z2`;
     Y3 = rannor(J(n, 6, 0));
     Y3 = Y3`;
     z3 = F*Y3;
     z3 = STB*z3 +.3;
     z3 = z3`;
     Y4 = rannor(J(n, 6, 0));
     Y4 = Y4`;
     z4 = F*Y4;
     z4 = STB*z4 +1.0;  *1.3 for interaction effect of .60*;
     z4 = z4`;
    z = z1//z2//z3//z4;
    *print n z2 z21 z;
          /* z is a set of cluster means */
          %do j = 1 %to &cn;
          use patternW;
          read all var _NUM_ into FW;
          FW = FW`;
          use stdW;
          read all var _NUM_ into STW;
          YW = rannor(J(&cs, 6, 0));
          YW = YW`;
          zW = FW*YW;
          zW = STW*zW;
          zW = zW`;
          cmean = z[&j,1:6];
          wthn = zw + cmean;
          clid = J(&cs, 1, &j);
          W = wthn||clid;
          run;
          * create a SAS dataset per cluster;
          create cl&j from W [colname={y1 y2 y3 y4 y5 y6 clus}];
          append from W;
          *print YW zW z cmean W;
          %end; *end of cluster loop;
          * stack up clusters to create a final dataset;
          data rep&i ;
```

```
                        %if &cnloop = 1 %then %do;
                         set cl1-cl40;
                         if clus lt 11 then d1 = 0;
                         if clus lt 11 then d2 = 0;
                         if clus gt 10 and clus lt 21 then d1 = 0;
                         if clus gt 10 and clus lt 21 then d2 = 1;
                         if clus gt 20 and clus lt 31 then d1 = 1;
                         if clus gt 20 and clus lt 31 then d2 = 0;
                         if clus gt 30 then d1 = 1;
                         if clus gt 30 then d2 = 1;
                        %end;
                        %if &cnloop = 2 %then %do;
                         set cl1-cl80;
                         if clus lt 21 then d1 = 0;
                         if clus lt 21 then d2 = 0;
                         if clus gt 20 and clus lt 41 then d1 = 0;
                         if clus gt 20 and clus lt 41 then d2 = 1;
                         if clus gt 40 and clus lt 61 then d1 = 1;
                         if clus gt 40 and clus lt 61 then d2 = 0;
                         if clus gt 60 then d1 = 1;
                         if clus gt 60 then d2 = 1;
                        %end;
                        %if &cnloop = 3 %then %do;
                         set cl1-cl120;
                         if clus lt 31 then d1 = 0;
                         if clus lt 31 then d2 = 0;
                         if clus gt 30 and clus lt 61 then d1 = 0;
                         if clus gt 30 and clus lt 61 then d2 = 1;
                         if clus gt 60 and clus lt 91 then d1 = 1;
                         if clus gt 60 and clus lt 91 then d2 = 0;
                         if clus gt 90 then d1 = 1;
                         if clus gt 90 then d2 = 1;
                        %end;
                  run;
                  proc export data= rep&i
                   outfile = "U:\dissertation
simulation\between_3\S_icc\cn&cn.cs&cs\rep&i..dat"
                   dbms = dlm replace     ;
                   putnames = no;
                  run;
            %end; *end of replication loop;
      %end; *end of cluster size loop;
%end; *end of cluster number loop;
%mend mlmimic; * end of macro;
%mlmimic; run;
```

Within-level covariates interaction effect of .30 with small ICC

```
/****************************************/
/*            Pattern Matrix            */
/****************************************/;
* specify matrices at the between level;
* step 1: obtain correlation matrix and SDs;
options nonotes nosource nosource2 noxwait xsync;
proc iml;
```

```
LY = {.8, .8, .8, .8, .8, .8};
PH = {.10};     **For medium and large ICC,PH is .25 and .50, respectively*;
TE = {.004 0 0 0 0 0,
        0 .004 0 0 0 0,
        0 0 .004 0 0 0,
        0 0 0 .004 0 0,
        0 0 0 0 .004 0,
        0 0 0 0 0 .004}; *TE is .01, and .02 for medium and large ICCs*;

COV = LY*PH*LY` + TE;
*print COV ;
      * obtain correlation matrix and SDs from COV;
S = sqrt(diag(cov));
R = (inv(S))*cov*(inv(S));
 create STDB from S [colname={y1 y2 y3 y4 y5 y6}];
       append from S;
create CR from R [colname={y1 y2 y3 y4 y5 y6}];
     append from R;
print S R;

* step 2: obtain between-level pattern matrix;
data A (type = corr);
_TYPE_ ='CORR';
input y1 y2 y3 y4 y5 y6;
cards;
        1 . . . . .
        0.9411765         1 . . . .
        0.9411765 0.9411765         1 . . .
        0.9411765 0.9411765 0.9411765         1 . .
        0.9411765 0.9411765 0.9411765 0.9411765         1 .
        0.9411765 0.9411765 0.9411765 0.9411765 0.9411765         1
;

*obtain factor pattern matrix for data generation;
proc factor n=6 outstat=facoutB noprint;
data patternB; set facoutB;
 if _TYPE_ = 'PATTERN';
 drop _TYPE_ _NAME_ ;
run;

* specify within-level model;
* step 1: obtain correlation matrix and SDs;
proc iml;
LYW = {.8, .8, .8, .8, .8, .8};
PHW = {1};
TEW = {.36 0 0 0 0 0,
        0 .36 0 0 0 0,
        0 0 .36 0 0 0,
        0 0 0 .36 0 0,
        0 0 0 0 .36 0,
        0 0 0 0 0 .36};
COVW = LYW*PHW*LYW` + TEW;
*print COVW ;
      * obtain correlation matrix and SDs from COV;
SW = sqrt(diag(covW));
RW = (inv(SW))*covW*(inv(SW));
 create STDW from SW [colname={y1 y2 y3 y4 y5 y6}];
```

```
       append from SW;
*print SW RW;

* step 2: obtain within-level pattern matrix;
data B (type = corr);
_TYPE_ ='CORR';
input y1 y2 y3 y4 y5 y6;
cards;
        1 . . . . .
        0.64   1 . . . .
        0.64 0.64   1 . . .
        0.64 0.64 0.64   1 . .
        0.64 0.64 0.64 0.64   1 .
        0.64 0.64 0.64 0.64 0.64   1
;
      *obtain factor pattern matrix for data generation;
proc factor n=6 outstat=facoutW noprint;
data patternW; set facoutW;
 if _TYPE_ = 'PATTERN';
 drop _TYPE_ _NAME_ ;
run;


/*****************************************/
/*           Data Generation            */
/*****************************************/;
options nonotes;
*libname ml 'U:\mlMIMIC\dataSAS';

%let mc = 1000; *number of replications;

%macro mlmimic;
      * do loop for cluster number;
%do cnloop = 1 %to 3;
 %if &cnloop = 1 %then %do; %let cn = 20; %end;
 %if &cnloop = 2 %then %do; %let cn = 40; %end;
 %if &cnloop = 3 %then %do; %let cn = 60; %end;
            * do loop for cluster size;
      %do csloop = 1 %to 2;
       %if &csloop = 1 %then %do; %let cs = 20; %end;
       %if &csloop = 2 %then %do; %let cs = 40; %end;
        X mkdir "U:\dissertation simulation\within_3\S_icc\cn&cn.cs&cs\";
                  * do loop to generate the given number of replications;
            %do i=1 %to &mc;
            proc iml;
            use patternB;
            read all var _NUM_ into F;
            F = F`;
            use stdb;
            read all var _NUM_ into STB;
            Y = rannor(J(&cn, 6, 0));
            Y = Y`;
            z = F*Y;
            z = STB*z;
            z = z`;
            /* z is a set of cluster means */
                  %do j = 1 %to &cn;
```

```
                  use patternW;
                  read all var _NUM_ into FW;
                  FW = FW`;
                  use stdW;
                  read all var _NUM_ into STW;
                  n= &cs/4;
                  YW = rannor(J(&cs, 6, 0));
                  YW = YW`;
                  zW = FW*YW;
                  zW = STW*zW;
                  zW = zW`;
                  cmean = z[&j,1:6];
                  wthn = zw + cmean;
                  wthn1 = wthn[1:n, ] - 0.2;
                  wthn2 = wthn[n+1:2*n, ] + 0.2;
                  wthn3 = wthn[2*n+1:3*n,] + 0.1;
                  wthn4 = wthn[3*n+1:4*n,] +0.8;
                  g1 = J(n, 2, 0);
                  g2 = J(n, 2, 0);
                  g2[,2] = 1;
                  g3 = J(n, 2, 0);
                  g3[,1] = 1;
                  g4 = J(n, 2, 1);
                  wthn = (wthn1||g1)//(wthn2||g2)//(wthn3||g3)//(wthn4||g4);
                  clid = J(&cs, 1, &j);
                  W = wthn||clid;
                  run;
                  * create a SAS dataset per cluster;
                  create cl&j from W[colname={y1 y2 y3 y4 y5 y6 d1 d2clus}];
                  append from W;
                  print YW zW z cmean W;
                  %end; *end of cluster loop;
                  * stack up clusters to create a final dataset;
                  data rep&i ;
                        %if &cnloop = 1 %then %do;
                         set cl1-cl20;
                        %end;
                        %if &cnloop = 2 %then %do;
                         set cl1-cl40;
                        %end;
                        %if &cnloop = 3 %then %do;
                         set cl1-cl60;
                        %end;
                  run;
                  proc export data= rep&i
                   outfile = "U:\dissertation
simulation\within_3\S_icc\cn&cn.cs&cs\rep&i..dat"
                   dbms = dlm replace     ;
                   putnames = no;
                  run;
            %end; *end of replication loop;
      %end; *end of cluster size loop;
%end; *end of cluster number loop;
%mend mlmimic; * end of macro;
%mlmimic; run;
```

1.1 replaces .80 for interaction effect of .60;

Cross-level covariates interaction effect of .20 with small ICC

```
/****************************************/
/*            Pattern Matrix            */
/****************************************/;
* specify matrices at the between level;
* step 1: obtain correlation matrix and SDs;
title ' ';
proc iml;
LY = {.8, .8, .8, .8, .8, .8};
PH = {.1};  **For medium and large ICC,PH is .25 and .50, respectively*;
TE = {.004 0 0 0 0 0,
        0 .004 0 0 0 0,
        0 0 .004 0 0 0,
        0 0 0 .004 0 0,
        0 0 0 0 .004 0,
        0 0 0 0 0 .004};*TE is .01, and .02 for medium and large ICCs*;
COV = LY*PH*LY` + TE;
*print COV ;
      * obtain correlation matrix and SDs from COV;
S = sqrt(diag(cov));
R = (inv(S))*cov*(inv(S));
 create STDB from S [colname={y1 y2 y3 y4 y5 y6}];
       append from S;
 /*create CR from R [colname={y1 y2 y3 y4 y5 y6}];
     append from R;*/
*print S R;

* step 2: obtain between-level pattern matrix;
data A (type = corr);
_TYPE_ ='CORR';
input y1 y2 y3 y4 y5 y6;
cards;
        1 . . . . .
        0.9411765         1 . . . .
        0.9411765 0.9411765         1 . . .
        0.9411765 0.9411765 0.9411765         1 . .
        0.9411765 0.9411765 0.9411765 0.9411765         1 .
        0.9411765 0.9411765 0.9411765 0.9411765 0.9411765         1
;
*obtain factor pattern matrix for data generation;
proc factor n=6 outstat=facoutB noprint;
data patternB; set facoutB;
 if _TYPE_ = 'PATTERN';
 drop _TYPE_ _NAME_ ;
run;

* specify within-level model;
* step 1: obtain correlation matrix and SDs;
proc iml;
LYW = {.8, .8, .8, .8, .8, .8};
```

```
PHW = {1};
TEW = {.36 0 0 0 0 0,
        0 .36 0 0 0 0,
        0 0 .36 0 0 0,
        0 0 0 .36 0 0,
        0 0 0 0 .36 0,
        0 0 0 0 0 .36};
COVW = LYW*PHW*LYW` + TEW;
*print COVW ;
      * obtain correlation matrix and SDs from COV;
SW = sqrt(diag(covW));
RW = (inv(SW))*covW*(inv(SW));
 create STDW from SW [colname={y1 y2 y3 y4 y5 y6}];
       append from SW;
*print SW RW;

* step 2: obtain within-level pattern matrix;
data B (type = corr);
_TYPE_ ='CORR';
input y1 y2 y3 y4 y5 y6;
cards;
        1 . . . . .
        0.64   1 . . . .
        0.64 0.64   1 . . .
        0.64 0.64 0.64   1 . .
        0.64 0.64 0.64 0.64   1 .
        0.64 0.64 0.64 0.64 0.64   1
;
      *obtain factor pattern matrix for data generation;
proc factor n=6 outstat=facoutW noprint;
data patternW; set facoutW;
 if _TYPE_ = 'PATTERN';
 drop _TYPE_ _NAME_ ;
run;


/*****************************************/
/*           Data Generation            */
/*****************************************/;
options nonotes;
options noxwait xsync;
*libname ml 'U:\mlMIMIC\dataSAS';

%let mc = 1000; *number of replications;

%macro mlmimic;
      * do loop for cluster number;
%do cnloop = 1 %to 3;
 %if &cnloop = 1 %then %do; %let cn = 40; %end;
 %if &cnloop = 2 %then %do; %let cn = 80; %end;
 %if &cnloop = 3 %then %do; %let cn = 120; %end;
            * do loop for cluster size;
      %do csloop = 1 %to 2;
       %if &csloop = 1 %then %do; %let cs = 10; %end;
       %if &csloop = 2 %then %do; %let cs = 20; %end;
        X mkdir "U:\dissertation simulation\cross_2\S_ICC\cn&cn.cs&cs\";
                  * do loop to generate the given number of replications;
```

```
%do i=1 %to &mc;
proc iml;
use patternB;
read all var _NUM_ into F;
F = F`;
use stdb;
read all var _NUM_ into STB;
            * create two level-2 group marginal means ;
 n = &cn/2;
 Y1 = rannor(J(n, 6, 0));
 Y1 = Y1`;
 z1 = F*Y1;
 z1 = STB*z1;
 z1 = z1`;
 Y2 = rannor(J(n, 6, 0));
 Y2 = Y2`;
 z2 = F*Y2;
 z2 = STB*z2 +.6;          **.70 replaces .60 for interaction
 z2 = z2`;                 effect of .40**;

z = z1//z2;
*print n z1 z2 z;
      /* z is a set of cluster means */
      %do j = 1 %to &cn/2; * cluster loop 1;
      use patternW;
      read all var _NUM_ into FW;
      FW = FW`;
      use stdW;
      read all var _NUM_ into STW;
      m= &cs/2;
      YW = rannor(J(&cs, 6, 0));
      YW = YW`;
      zW = FW*YW;
      zW = STW*zW;
      zW = zW`;
      cmean = z[&j,1:6];
      wthn = zw + cmean;
      wthn1 = wthn[1:m, ] - 0.2;
      wthn2 = wthn[m+1:2*m, ] + 0.2;
      g1 = J(m, 1, 0);
      g2 = J(m, 1, 1);
      wthn = (wthn1||g1)//(wthn2||g2);
      l2g = J(&cs, 1, 0);
      clid = J(&cs, 1, &j);
      W = wthn||l2g||clid;
      run;
      * create a SAS dataset per cluster;
create cl&j from W [colname={y1 y2 y3 y4 y5 y6 L1grp L2grp clus}];
      append from W;
      *print YW zW z cmean W;
      %end; *end of cluster loop 1;
      %do j = &cn/2 + 1 %to &cn; * cluster loop 2;
      use patternW;
      read all var _NUM_ into FW;
      FW = FW`;
      use stdW;
      read all var _NUM_ into STW;
```

```
                  m= &cs/2;
                  YW = rannor(J(&cs, 6, 0));
                  YW = YW`;
                  zW = FW*YW;
                  zW = STW*zW;
                  zW = zW`;
                  cmean = z[&j,1:6];
                  wthn = zw + cmean;
                  wthn1 = wthn[1:m, ] - 0.3;
                  wthn2 = wthn[m+1:2*m, ] + 0.3;
                  g1 = J(m, 1, 0);
                  g2 = J(m, 1, 1);
                  wthn = (wthn1||g1)//(wthn2||g2);
                  l2g = J(&cs, 1, 1);
                  clid = J(&cs, 1, &j);
                  W = wthn||l2g||clid;
                  run;
                  * create a SAS dataset per cluster;
       create cl&j from W [colname={y1 y2 y3 y4 y5 y6 L1grp L2grp clus}];
                  append from W;
                  *print YW zW z cmean W;
                  %end; *end of cluster loop2;
                  * stack up clusters to create a final dataset;
                  data rep&i ;
                        %if &cnloop = 1 %then %do;
                         set cl1-cl40;
                        %end;
                        %if &cnloop = 2 %then %do;
                         set cl1-cl80;
                        %end;
                        %if &cnloop = 3 %then %do;
                         set cl1-cl120;
                        %end;
                  run;
                  proc export data= rep&i
                   outfile = "U:\dissertation
simulation\cross_2\S_ICC\cn&cn.cs&cs\rep&i..dat"
                   dbms = dlm replace     ;
                   putnames = no;
                  run;
            %end; *end of replication loop;
      %end; *end of cluster size loop;
%end; *end of cluster number loop;
%mend mlmimic; * end of macro;
%mlmimic; run;

quit;
```

.40 replaces .30 in these two lines for interaction effect of .40;