Exploring the Test of Covariate Moderation Effects in Multilevel MIMIC Models

Educational and Psychological
Measurement

2019, Vol. 79(3) 512–544

© The Author(s) 2018

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0013164418793490

journals.sagepub.com/home/epm

Chunhua Cao¹, Eun Sook Kim², Yi-Hsin Chen²,
John Ferron² and Stephen Stark²

Abstract

In multilevel multiple-indicator multiple-cause (MIMIC) models, covariates can interact at the within level, at the between level, or across levels. This study examines the performance of multilevel MIMIC models in estimating and detecting the interaction effect of two covariates through a simulation and provides an empirical demonstration of modeling the interaction in multilevel MIMIC models. The design factors include the location of the interaction effect (i.e., between, within, or across levels), cluster number, cluster size, intraclass correlation (ICC) level, magnitude of the interaction effect, and cross-level measurement invariance status. Type I error, power, relative bias, and root mean square of error of the interaction effects are examined. The results showed that multilevel MIMIC models performed well in detecting the interaction effect at the within or across levels. However, when the interaction effect was at the between level, the performance of multilevel MIMIC models depended on the magnitude of the interaction effect, ICC, and sample size, especially cluster number. Overall, cross-level measurement noninvariance did not make a notable impact on the estimation of interaction in the structural part of multilevel MIMIC models when factor loadings were allowed to be different across levels.

Keywords

multilevel, MIMIC, covariates interaction

¹Polk County School District, Bartow, FL, USA

²University of South Florida, Tampa, FL, USA

Corresponding Author:

Chunhua Cao, University of South Florida, 4202 East Fowler Ave, Tampa, FL 33620, USA.

Email: chunhuacao@mail.usf.edu

Introduction

The typical feature of the multiple-indicator multiple-cause (MIMIC) model is that the latent factor is measured by the observed indicators, which is the measurement model part, and is regressed on at least one observed variable, which is the structural model part (Jöreskog & Goldberger, 1975). The MIMIC model is flexible in modeling multiple covariates, which may be continuous or discrete, observed or latent, with at least one observed covariate. 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 been utilized extensively by applied researchers, especially for testing latent factor means or the regression effect from a grouping variable on the latent factor (Condon, 2010; Ogg, McMahan, Dedrick, & Mendez, 2013; Thompson & Green, 2006). When modeling the effects of multiple covariates on the latent factor in the MIMIC model, the interactions between the covariates should be investigated in addition to the main effects of the covariates. This interaction effect is also referred to as the 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).

Hierarchical or nested data structure is very common in educational and psychological research; for example, students are nested within classrooms that are nested within schools. Multilevel modeling accounts for the fact that individuals belonging to the same group share more similarities with one another than with individuals from other groups. The primary problem of employing single-level statistical models using nested data is the violation of the independence assumption of individuals in the same group (Raudenbush & Bryk, 2002). Previous studies showed that single-level models that did not account for the multilevel nature of hierarchically structured data yielded negatively biased standard error estimates, which consequently resulted in an inflation of the Type I error rates (Finch & French, 2011; Kim, Kwok, & Yoon, 2012; Snijders & Bosker, 1999). Multilevel MIMIC models have been gaining popularity in educational and psychological research (Davidov, Dülmer, Schlüter, Schmidt, & Meuleman, 2012; Jak, Oort, & Dolan, 2014; Kim & Cao, 2015; Kim, Yoon, Wen, Luo, & Kwok, 2015). Given the prevalence of structural equation modeling (SEM) in social and educational sciences, as well as the frequency of the presence of hierarchical data structure, it is critical for the researchers to employ the appropriate statistical methods for modeling hierarchical data in the context of latent variable models. Finch and French (2011) conducted one study to compare 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. Although this study indicated that it is important to take into account the nested structure of data in multilevel MIMIC models, their study was limited to just one covariate at the between or the within level. The advantage of the multilevel MIMIC model is its flexibility in modeling covariates from both the between and within levels and their interaction effects.

To our knowledge, there have been no studies examining the performance of MIMIC models in estimating the effects of multiple covariates and their interaction on the latent factor in a multilevel SEM context. The covariates interaction (referred to as interaction hereinafter) effect can be formed by covariates of the between level only, covariates of the within level only, and covariates across levels. The between-level interaction occurs when the effect of one between-level covariate on the between-level factor depends on another between-level covariate. Similarly, the within-level interaction occurs when the effect of one within-level covariate on the within-level factor is moderated by another within-level covariate. As for the cross-level interaction effect, the effect of the within-level covariate on the within-level factor differs as a function of the between-level covariate. The cross-level interaction has been used extensively by applied researchers in multilevel regression analysis, but not so much in multilevel SEM. The interaction effect in general and more specifically the cross-level interaction has not been given much attention in latent variable modeling. Therefore, the purpose of this study is twofold: (a) to examine the performance of multilevel MIMIC models in estimating the interaction effect under a variety of conditions through a Monte Carlo simulation and (b) to demonstrate the detailed modeling of the interaction effect in multilevel MIMIC models with an empirical example.

Theoretical Framework

Multiple-Indicator Multiple-Cause Models

The MIMIC method is very flexible in modeling multiple covariates and their interactions (Cheng et al., 2016; Fleishman, Spector, & Altman, 2002). The MIMIC model describing the linear relation between observed variables and latent factors consists of two parts: (a) a measurement model and (b) a structural regression model:

$$y_i = \nu + \Lambda \eta_i + \varepsilon_i, \quad (1)$$

$$\eta_i = \Gamma X_i + \zeta_i, \quad (2)$$

$$\varepsilon \sim N(0, \Theta_\varepsilon), \quad (3)$$

$$\zeta \sim N(0, \Psi), \quad (4)$$

where, for individual i , y is a vector of observed indicator variables, ν denotes a vector of intercepts, Λ indicates 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. The covariates, X , can be continuous or categorical, observed or even latent variables. When the covariates are observed, the model is

specifically called a MIMIC model. When a covariate is a dichotomous variable signifying membership in one of two comparison groups, the corresponding element of $\Gamma(\gamma)$ indicates the latent factor mean difference of the two groups, adjusting for the other covariates in X .

Because residuals are assumed to be normally distributed with a mean of 0 and uncorrelated with latent factors: $E(\varepsilon) = 0$, $\text{Cov}(\eta, \varepsilon) = 0$, the population variance–covariance matrix Σ is written as follows:

$$\Sigma = \Lambda(\Gamma\Phi\Gamma' + \Psi)\Lambda' + \Theta_{\varepsilon}, \quad (5)$$

where Φ and Θ_{ε} are variance–covariance matrices for common factors and residuals, respectively. As a matter of nomenclature in this study, the indicators (y) and factor loadings (Λ) refer to the observed effect variables of and their relations to the latent factor (η), respectively, in the measurement part; the covariates (X) and regression coefficients (Γ) refer to the observed cause variables of and their effects on the latent factor, respectively, in the structural regression part.

Measurement Invariance Assumption in MIMIC Models

In multiple group confirmatory factor analysis (CFA), a separate model for each group is specified, and the invariance of measurement parameters (e.g., factor loadings, item intercepts, and item residual variances) is tested across groups using equality constraints. Unlike multiple group CFA, the MIMIC modeling estimates just one model for the full (combined) sample of respondents. The MIMIC model implicitly assumes that the same measurement model holds in all the groups. More specifically, factor loadings, item intercepts, residual variances, and factor variances are invariant for all the groups. That is, all sources of covariation among observed variables are assumed to be equal across groups (Hancock, Lawrence, & Nevitt, 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 differences. When the assumption of homogeneous factor disturbance variances or residual variances was violated across the two groups, the performance of multilevel MIMIC models in testing latent mean differences across groups was comparable to that of multiple group multilevel CFA that does not require stringent measurement invariance assumptions. However, strong measurement invariance, that is, identical factor loadings and intercepts across groups, is a prerequisite to compare latent factor means across groups (Meredith, 1993).

Multilevel MIMIC Models

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 models, a subscript j is included to indicate that a certain

parameter of interest can vary across clusters. Equation (1) is decomposed into the between-level and within-level models and rewritten in Equation (6):

$$y_{ij} = \nu_B + \Lambda_B \eta_{Bj} + \Lambda_W \eta_{Wij} + \varepsilon_{Bj} + \varepsilon_{Wij}, \quad (6)$$

$$\varepsilon_{Wij} \sim N(0, \Theta_{\varepsilon W}), \quad (7)$$

$$\varepsilon_{Bj} \sim N(0, \Theta_{\varepsilon B}), \quad (8)$$

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; η_{Bj} is the between-level factor; Λ_W is the within-level factor loading, and it is not random; η_{Wij} is the within-level latent factor; ε_{Bj} and ε_{Wij} 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 are covariates only at the between level, the between-level covariates (X_{Bj}) are included in the regression model with direct effects on the latent factor η_{Bj} as shown in Equation (9):

$$\eta_{Bj} = \Gamma_B X_{Bj} + \zeta_{Bj}, \quad (9)$$

$$\zeta_{Bj} \sim N(0, \Psi_B), \quad (10)$$

where Γ_B represents the effects of the covariates on the latent factors. If one of the between-level covariates is a dichotomous grouping variable, the corresponding coefficient in Γ_B is the between-level group difference in the latent factor means, adjusted for the other covariates in X_{Bj} . Similarly, the total covariance matrix in Equation (5) is divided into the within-level and the between-level components:

$$\Sigma_T = \Sigma_B + \Sigma_W, \quad (11)$$

where subscripts T, W, and B denote the total, within, and 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 there are covariates only at the between level, Σ_W and Σ_B can be written as follows, respectively:

$$\Sigma_W = \Lambda_W \Phi_W \Lambda'_W + \Theta_{\varepsilon W}, \quad (12)$$

$$\Sigma_B = \Lambda_B (\Gamma_B \Phi_B \Gamma'_B + \Psi_B) \Lambda'_B + \Theta_{\varepsilon B}, \quad (13)$$

On the other hand, when there are covariates only at the within level, the latent variable at the within level (η_{Wij}) is explained by the within-level covariates:

$$\eta_{Wij} = \Gamma_W X_{Wij} + \zeta_{Wij}, \quad (14)$$

$$\Sigma_W = \Lambda_W (\Gamma_W \Phi_W \Gamma'_W + \Psi_W) \Lambda'_W + \Theta_{\varepsilon_W}, \quad (15)$$

$$\Sigma_B = \Lambda_B \Phi_B \Lambda'_B + \Theta_{\varepsilon_B}, \quad (16)$$

$$\zeta_{Wij} \sim N(0, \Psi_W). \quad (17)$$

If a within-level covariate is a dichotomous grouping variable, the corresponding element of Γ_W represents the latent mean difference between groups at the within level, adjusted for the other covariates in X_{Wij} , and it is a fixed effect.

In the two scenarios described above, the covariates are 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 covariates X_{Bj} on the between-level latent factor and the effect of the within-level covariates X_{Wij} on the within-level latent factor are specified in the model, that is, Equation (9) and Equation (14), correspondingly. The between-level covariance matrix and the within-level covariance matrix are listed in Equation (13) and Equation (15), respectively. In addition to its flexibility in modeling different types of covariates, for example, categorical, or continuous, the multilevel MIMIC model has the advantage of incorporating covariates at different levels simultaneously. This feature is not possible with multiple group analysis.

The Covariates Interaction in Multilevel MIMIC Models

In the situation of MIMIC models with a moderating covariate, the effect of one covariate on the latent factor depends on the value or the level of another covariate as presented in Figures 1 and 2. In Figure 1, the two covariates are at the between level, and the effect of covariate 2, X_{2B} , on the latent factor depends on the value of covariate 1, X_{1B} . Note that the six indicators are represented by circles instead of squares, because the indicators at the between level are considered latent random intercepts. Research scenarios with two covariates at the between level include research on the impact of some contextual variables on latent factors. For example, a researcher wants to know whether the impact of school control or treatment status on a latent factor 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. When there are two dichotomous covariates (X_{1Bj} and X_{2Bj}) with their interaction effect at the between level, the structural regression model in multilevel MIMIC models with paths from the covariates to the latent factor in Equation (9) can be written as

$$\eta_{Bj} = \gamma_{1B} X_{1Bj} + \gamma_{2B} X_{2Bj} + \gamma_{3B} X_{1Bj} * X_{2Bj} + \zeta_{Bj}, \quad (18)$$

where γ_{1B} is the main effect of covariate X_{1Bj} , γ_{2B} is the main effect of covariate X_{2Bj} , and γ_{3B} is the interaction effect of the two covariates above. The parameter of interest is the interaction effect, γ_{3B} .

Figure 1. The between-level multiple-indicator multiple-cause (MIMIC) model with two categorical covariates and the moderation effect.

In Figure 2, the two covariates are at the within level, and similarly, the effect of covariate 2, $X2_W$, on the within-level latent factor depends on the value of covariate 1, $X1_W$. For example, a latent factor could be regressed on two within-level covariates (e.g., gender and free lunch status), and there may be an interaction effect between the two covariates. When there are two dichotomous covariates ($X1_{Wij}$ and $X2_{Wij}$) at the within level, the latent variable at the within level (η_{Wij}) can be explained by the within-level grouping covariates and their interaction. Equation (14) can be written as

$$\eta_{Wij} = \gamma_{1W}X1_{Wij} + \gamma_{2W}X2_{Wij} + \gamma_{3W}X1_{Wij} * X2_{Wij} + \zeta_{Wij}. \tag{19}$$

In this case, γ_{1W} is the main effect of covariate $X1_{Wij}$, γ_{2W} is the main effect of covariate $X2_{Wij}$, and γ_{3W} is the interaction effect of the two covariates above.

In Figure 3, one of the covariates is at the between level and the other one is at the within level. The regression effect of the within-level covariate X_W on the latent factor is random and varies across the clusters at the between level, and this random coefficient can be explained by the between-level covariate X_B . If the effect of X_B on the random coefficient is significant, there exists a cross-level interaction, indicating that the effect of the within-level covariate on the latent factor depends on the value of the between-level covariate. As shown in Figure 3, the random coefficient is represented by a solid dot on the regression coefficient and it is written as

$$\eta_{Wij} = \gamma_{Wj}X_{Wij} + \zeta_{Wij}, \tag{20}$$

Figure 2. The within-level multiple-indicator multiple-cause (MIMIC) model with two categorical covariates and the moderation effect.

where γ_{Wj} with a subscript j is the random regression coefficient of the within-level covariate X_{Wij} predicting the within-level latent factor, η_{Wij} ; ζ_{Wij} refers to the residual of the latent factor, η_{Wij} . At the between level, the between-level latent factor is regressed on the between-level covariate, X_B . X_B also explains the random effect at the between level. The equations of the between-level latent factor and the random coefficient regressed on the between-level covariate, X_B , can be written as

$$\eta_{Bj} = \gamma_B X_{Bj} + \zeta_{Bj}, \quad (21)$$

$$\gamma_{Wj} = \gamma_{10} + \gamma_C X_{Bj} + \zeta_{\gamma Wj}, \quad (22)$$

where γ_B is the between-level regression coefficient representing the effect of the between-level covariate, X_{Bj} on the between-level latent factor η_{Bj} ; ζ_{Bj} refers to the residual of η_{Bj} that is not explained by X_{Bj} . γ_{10} and γ_C , respectively, represent the intercept and slope of the random coefficient γ_{Wj} , which is regressed on X_{Bj} . $\zeta_{\gamma Wj}$ is the residual, implying that the variance of γ_{Wj} is not fully explained by the between-level covariate, X_{Bj} . The combined model for the within-level latent variable, η_{Wij} , in Equation (20), can be written as follows by substituting γ_{Wj} with Equation (23):

$$\eta_{Wij} = \gamma_{10} X_{Wij} + \gamma_C X_{Wij} * X_{Bj} + \zeta_{\gamma Wj} * X_{Wij} + \zeta_{Wij}, \quad (23)$$

Figure 3. The cross-level covariates interaction in the multiple-indicator multiple-cause (MIMIC) model with one categorical covariate at the between level and one categorical covariate at the within level.

where γ_C is the cross-level interaction effect, which is one of the focuses of the study when the two covariates in multilevel MIMIC models are at the within level and the between level.

Purpose of the Study

There are two purposes of this study. The first one is to explore the performance of multilevel MIMIC models in estimating and testing the interaction effect (between-level interaction, within-level interaction, and cross-level interaction) using a simulation study. The performance will be examined through Type I error control, statistical power, relative bias of the parameter estimate of interaction effect, and root mean square error (RMSE). The second purpose is to provide applied researchers an empirical example of the detailed modeling process and results interpretation of interaction effect in multilevel MIMIC.

Simulation Study

Data Generation

In this simulation study, two-level data with two dichotomous covariates were generated using PROC IML in SAS/IML 9.4. The four groups were made by the crossing of the two dichotomous covariates. For both the between- and the within-level models, a single factor with six continuous indicators was generated. At the within level, factor loadings varied to create cross-level measurement invariance and

noninvariance conditions, which will be explained in the simulation conditions section. 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. 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 intraclass correlations (ICCs), respectively. The between-level residual variances were selected based on the following considerations. Satisfying scalar measurement invariance across clusters in two-level SEM models required to have between-level residual variances of zero based on the research on cluster bias in multilevel data conducted by Jak et al. (2014). In real-world data, residual variances are unlikely to be exactly zero, so values close to zero (.02, .01, and .004) but not zero were selected as the between-level residual variance. The between-level factor variance was varied to form different ICCs, which will be explained later in the section of simulation conditions.

When the two covariates were at the same level (i.e., both at the within or between level), the main regression 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 (Finch & French, 2011; Maas & Hox, 2005). 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 to be .40, and the between-level main effect was set to be .50. These numbers generally 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 regression coefficient ranging between $-.06$ and $.45$ and a between-level regression coefficient ranging between $-.23$ and $.35$. The population parameters are summarized in Table 1.

Simulation Conditions

To examine the performance of multilevel MIMIC models in detecting the interaction effect, a set of design factors were taken into account. Simulation conditions included the location of the interaction effect (i.e., at the between level, at the within level, and across levels), cluster number (CN) and cluster size (CS), ICC, the magnitude of the interaction effect, and cross-level measurement invariance status.

Location of the Interaction Effect. There were three locations of the interaction effect: at the between level, at the within level, and across levels. When the location of the interaction effect was different, some of the other design factors varied, too, which will be explained in the section of number of clusters and cluster size.

Number of Clusters and Cluster Size. CN was included because previous simulation studies on multilevel modeling found that CN had an impact on estimating

Table 1. Parameters in the Population Data Generation.

Parameter	Location of the interaction effect		
	Within level	Between level	Cross level
Number of indicators	6	6	6
Between-level factor loadings	.8	.8	.8
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; ICC = intraclass correlation.

parameters in multilevel models (Hox & Maas, 2001; Kim et al., 2012; Maas & Hox, 2005). It is expected that CN and CS are directly related to the power to detect the 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 (e.g., control vs. treatment schools and public vs. private schools), creating four groups. Four balanced groups were generated, with 10, 20, and 30 clusters in each group. CS was either 10 or 20. A CS of 10 and 20 was suggested by Hox (1998) as typical CSs for designing multilevel research. Also, previous simulation research used similar CS (e.g., Hox & Maas, 2001; Kim et al., 2012; Maas & Hox, 2005). When the two covariates were at the within level, the CN were set at 20, 40, and 60. CS was set to be 20 and 40, so that the four balanced groups had a cell size of 5 and 10, respectively. The CN for the within-level interaction conditions was only half of the clusters used in the between-level interaction conditions, which were 40, 80, and 120. On the other hand, the CS for the between-level interaction conditions was only half of that at the within level (i.e., 10 and 20 instead of 20 and 40). For the cross-level interaction conditions, CN and CS were the same as the between-level interaction scenario. Multiplying CN by CS, the minimum total sample size was 400 (CN = 40, CS = 10 for the between-level and cross-level interaction, and CN = 20, CS = 20 for the within-level interaction), and the maximum sample size was 2,400 (CN = 120, CS = 20 for the between-level and cross-level interaction, and CN = 60, CS = 40 for the within-level interaction). Thus, the total sample size was the same for the three different locations of the interaction effect. The major reason for adopting this method was because the interaction was at different levels.

Intraclass Correlation. ICC was also manipulated, as the previous studies (e.g., Kim et al., 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 interaction effect in multilevel MIMIC models. ICC is defined as the ratio of

the between-level factor variance over the total factor variance, which is the sum of the between-level factor variance and the within-level factor variance. As previously stated, the within-level factor variance was set to be 1.00, so varying ICC levels were obtained by differing the values of the between-level factor variance. In this study, the between-level factor variances were set to be .10, .25, and .50, resulting in three ICCs of .09, .20, and .33, respectively, as the small, medium, and large ICCs. 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 that were computed with within and between variances of observed variables were .07, .15, and .25 for all six items. The between-level residual variances of all six indicators were set to be .02, .01, and .004 for the large, medium, and small ICCs, respectively.

The Magnitude of the Interaction. Prior research showed that the power of detecting the cross-level interaction in multilevel multiple regression depended mainly on the effect size of the interaction (Mathieu et al., 2012). Thus, the magnitude of the interaction was manipulated in this research. When the two covariates were both at the between or within level, the interaction effect was set to be 0, .30, and .60. The value of 0 was used to assess Type I error rate when there was no interaction effect in the population. The values of .30 and .60 were used to assess power, with the former value representing a medium interaction effect and the latter representing a large interaction effect. The cross-level interaction effects were set to be at 0, .20, and .40, with .20 and .40 representing small and medium cross-level interaction effects, respectively, which were consistent with the range (−.06 to .45) reported in the literature review by Mathieu et al. (2012). In their simulation study, they manipulated the cross-level interaction effects, ranging from 0 to .75. In the current study, only small and moderate cross-level interactions were simulated to examine the performance of multilevel MIMIC models because those were usually found in the applied research and we expect that the performance will improve with a larger interaction effect.

Cross-Level Measurement Invariance Status. Multilevel MIMIC modeling assumed cross-level measurement invariance, that is, the equality of within-level and between-level factor loadings. However, in real-world applied research, it is common for researchers to observe noninvariant between-level and within-level factor loadings (Kim, Dedrick, Cao, & Ferron, 2016). It is important to examine the performance of multilevel MIMIC modeling when the between-level factor loadings are different from the within-level factor loadings. Thus, cross-level measurement invariance status was manipulated in this study. For cross-level measurement invariance conditions, the six within-level factor loadings were all set to be .80, equal to the corresponding between-level factor loadings. For cross-level measurement noninvariance conditions, the six within-level factor loadings were all set to be .60, .20 lower than the between-level factor loadings of .80. This is because usually the within-level factor

loadings are lower than the between-level factor loadings (Kim et al., 2016). Within-level item residual variances were .36 and .64 for cross-level invariance and noninvariance conditions, respectively.

In short, there were three locations of the interaction effect, three levels of CN, two different CSs, three varying ICC levels, three levels of the interaction effect size, and two levels of cross-level measurement invariance status. Thus, there were a total of 324 conditions ($3 \times 3 \times 2 \times 3 \times 3 \times 2$). For each condition, 1,000 replications were generated.

Analysis of Simulation Results and Outcome Variables

Each replication was analyzed using multilevel MIMIC models with correct specifications based on the population model and estimated with robust maximum likelihood estimation, which is the default estimation method of Mplus 7.1 (Muthén & Muthén, 1998-2012). For the two-level random effect model, the within-level random regression coefficient was specified to be predicted by the between-level covariate (cross-level interaction effect) as mentioned above. However, the random coefficient was not explained 100% by the between-level covariate, and the residual variance of the random coefficient γ_{Wj} , that is, $\zeta_{\gamma_{Wj}}$ in Equation (22) was specified to be .10. Pilot studies showed that setting the residual variance of the random coefficient to a reasonable value was important to obtain model convergence.

Outcome variables included admissible solution rates (ASR), Type I error, power, estimated standard errors, relative bias, and RMSE. For the correct statistical inferences of the interaction effects, Type I error and statistical power were examined along with associated standard errors. For the estimation accuracy and precision, relative bias and RMSE were examined. The replications were classified as inadmissible solutions if they had one of these following scenarios: (a) no output was produced, that is, no parameter estimates or standard errors were obtained; or (b) the estimated parameters were not possible values, such as negative variance. The ASR was defined as the proportion of replications that produced admissible solutions. Type I error was defined as the proportion of replications in which multilevel MIMIC models falsely flagged a statistically significant interaction when there was no interaction in the population model. Power was defined as the proportion of replications in which multilevel MIMIC models correctly detected the interaction effect as statistically significant at alpha .05. The standard error of the parameter estimate (SE) was extracted from the Mplus output and the average SE of more than 1,000 replications was computed and compared across simulation conditions.

Relative bias and RMSE were used to evaluate the accuracy and precision of a parameter estimate. The parameter of interest in this study was the interaction effect, γ_{3B} , γ_{3W} , and γ_C in Equations (18), (19), and (23), respectively. Relative bias (RB) was computed as the ratio of the raw bias (deviation of the parameter estimate from the population parameter (θ) across replications (R)) 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:

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

For each location of the interaction (at the between level, at the within level, across levels) the impact of the other design factors (i.e., CN, CS, the magnitude of interaction effect, cross-level measurement invariance status, and ICC level) on outcome variables (i.e., ASR, Type I error, power, SE, relative bias, and RMSE) was examined using factorial analysis of variance (ANOVA) with generalized eta-squared (η^2). 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 derived by dividing the Type III sum of squares of a particular design factor or the 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.

Simulation Results

The results of the between-, within-, and cross-levels interaction under cross-level measurement invariance conditions are presented in Tables 2, 4, and 6, respectively. Those under cross-level measurement noninvariance conditions are presented in Tables 3, 5, and 7, respectively. The results of cross-level measurement invariance and noninvariance conditions had similar patterns, and thus the following results focus on the cross-level measurement invariance conditions. The comparison of the results of cross-level measurement noninvariance conditions with those of invariance conditions is presented at the end of the results section. Type I error and relative bias did not vary much across different simulation conditions. The ASR varied in the between-level interaction conditions only. Thus, power, SE, and RMSE were used as dependent variables in the ANOVA analyses in all the three locations of the interaction, and ASR was used as the dependent variable only in the between-level interaction conditions.

Admissible Solution Rate. In multilevel MIMIC modeling with two covariates at the between level, ASR was related to ICC and the total sample size as presented in Tables 2 and 8. The ASRs were higher with the small ICC. Simulation conditions with the combination of large ICC and small total sample size produced relatively lower ASR. When the two covariates were at the within level or cross level, ASRs were all 100% across all simulation conditions except some conditions with smaller sample size, as presented in Tables 4 and 6.

Table 2. Power, Type I error, Relative Bias, Standard Error of Estimate, and RMSE Under Cross-level Measurement Invariance Conditions of the Between-Level Covariates Interaction Effects at .30 and .60.

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

Note. ICC = intraclass correlation; CN = number of clusters per group; CS = cluster size; ASR = admissible solution rate; Error = Type I error; SE = 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, Relative Bias, Standard Error of Estimate, and RMSE Under Cross-Level Measurement Noninvariance Conditions of the Between-Level Covariates Interaction Effects at .30 and .60.

Between-level covariates interaction												
ICC	CN/CS	Error	Covariates interaction of .30					Covariates interaction of .60				
			ASR	Power	SE	R Bias	RMSE	ASR	Power	SE	R Bias	RMSE
Small	40/10	.07	.96	.26	.21	-.05	.21	.98	.76	.22	-.04	.22
	40/20	.07	.98	.38	.18	-.03	.19	.97	.88	.18	-.01	.20
	80/10	.05	.99	.50	.15	-.04	.15	.99	.96	.15	-.03	.16
	80/20	.06	.98	.61	.13	-.06	.13	.97	1.00	.13	-.01	.14
	120/10	.04	.99	.67	.12	-.04	.13	.98	1.00	.12	-.05	.13
Medium	120/20	.06	.99	.78	.10	-.04	.11	.97	1.00	.11	-.01	.10
	40/10	.05	.94	.17	.28	-.11	.27	.90	.54	.28	-.04	.29
	40/20	.06	.91	.21	.25	-.02	.26	.85	.61	.26	-.03	.27
	80/10	.06	.93	.32	.19	-.04	.21	.89	.82	.20	-.06	.19
	80/20	.05	.89	.35	.18	-.04	.18	.93	.87	.19	-.03	.20
Large	120/10	.06	.96	.41	.16	-.06	.16	.90	.95	.16	-.04	.17
	120/20	.06	.91	.51	.15	-.01	.15	.97	.96	.15	-.05	.16
	40/10	.06	.87	.14	.36	-.03	.39	.78	.35	.37	-.03	.37
	40/20	.08	.77	.14	.35	-.03	.36	.88	.40	.35	-.00	.36
	80/10	.05	.83	.21	.26	-.04	.27	.88	.62	.26	-.03	.27
80/20	.06	.90	.24	.25	.05	.26	1.00	.64	.25	-.02	.26	
120/10	.06	.87	.23	.21	-.09	.21	.96	.75	.21	-.06	.22	
120/20	.06	.96	.30	.21	-.01	.20	1.00	.81	.21	-.02	.21	

Note. ICC = intraclass correlation; CN = number of clusters per group; CS = cluster size; ASR = admissible solution rate; Error = Type I error; SE = 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, Relative Bias, Standard Error of Estimate, and RMSE Under Cross-Level Measurement Invariance Conditions of the Within-Level Covariates Interaction Effects at .30 and .60.

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

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

Table 5. Power, Type I error, Relative Bias, Standard Error of Estimate, and RMSE Under Cross-Level Measurement Noninvariance Conditions of the Within-Level Covariates Interaction Effects at .30 and .60.

Within-level covariates interaction												
ICC	CN/CS	Error	Covariates interaction of .30					Covariates interaction of .60				
			ASR	Power	SE	R Bias	RMSE	ASR	Power	SE	R Bias	RMSE
Small	20/20	.06	.99	.60	.13	.00	.14	1.00	.99	.14	.01	.14
	20/40	.06	1.00	.87	.09	.00	.10	1.00	1.00	.10	.00	.10
	40/20	.06	1.00	.88	.10	.00	.09	1.00	1.00	.10	.00	.10
	40/40	.06	1.00	1.00	.07	.01	.07	1.00	1.00	.07	.00	.07
	60/20	.05	1.00	.97	.08	.00	.08	1.00	1.00	.08	.00	.08
Medium	60/40	.07	1.00	1.00	.06	.00	.06	1.00	1.00	.06	.01	.06
	20/20	.06	1.00	.64	.13	.03	.14	1.00	.99	.14	.00	.14
	20/40	.05	1.00	.87	.09	.00	.10	1.00	1.00	.10	.00	.10
	40/20	.05	1.00	.88	.10	-.01	.10	1.00	1.00	.10	.00	.10
	40/40	.05	1.00	.99	.07	-.01	.07	1.00	1.00	.07	.00	.07
Large	60/20	.06	1.00	.96	.08	-.01	.08	1.00	1.00	.08	.00	.08
	60/40	.05	1.00	1.00	.06	-.00	.05	1.00	1.00	.06	.00	.06
	20/20	.07	1.00	.58	.13	-.02	.14	1.00	1.00	.14	.00	.14
	20/40	.08	1.00	.89	.09	.02	.10	1.00	1.00	.10	.00	.10
	40/20	.06	1.00	.88	.10	-.01	.09	1.00	1.00	.10	.00	.10
	40/40	.08	1.00	.99	.07	-.01	.07	1.00	1.00	.07	-.01	.07
	60/20	.06	1.00	.96	.08	.01	.08	1.00	1.00	.08	-.01	.08
	60/40	.06	1.00	.99	.06	-.00	.06	1.00	1.00	.06	.00	.06

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

Table 6. Power, Type I error, Relative Bias, Standard Error of Estimate, and RMSE Under Cross-Level Measurement Invariance Conditions of the Cross-Level Covariates Interaction Effects at .20 and .40.

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

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

Table 7. Power, Type I error, Relative Bias, Standard Error of Estimate, and RMSE Under Cross-Level Measurement Noninvariance Conditions of the Cross-Level Covariates Interaction Effects at .20 and .40.

Cross-level covariates interaction												
ICC	CN/CS	Error	Covariates interaction of .20				Covariates interaction of .40					
			ASR	Power	SE	R Bias	RMSE	ASR	Power	SE	R Bias	RMSE
Small	40/10	.06	.98	.32	.15	.09	.15	.98	.80	.15	.04	.15
	40/20	.06	1.00	.55	.10	.06	.10	1.00	.99	.10	.05	.10
	80/10	.06	1.00	.54	.10	.03	.10	1.00	.98	.10	.04	.10
	80/20	.07	1.00	.85	.07	.03	.07	1.00	1.00	.07	.02	.07
	120/10	.07	1.00	.73	.08	.04	.08	1.00	1.00	.08	.03	.08
Medium	120/20	.05	1.00	.96	.06	.03	.06	1.00	1.00	.06	.03	.06
	40/10	.05	1.00	.32	.15	.07	.15	1.00	.79	.15	.03	.15
	40/20	.05	1.00	.53	.10	.04	.10	1.00	.98	.10	.02	.10
	80/10	.04	1.00	.49	.10	.00	.10	1.00	.98	.10	.03	.11
	80/20	.04	1.00	.82	.07	.04	.07	1.00	1.00	.07	.02	.07
Large	120/10	.05	1.00	.70	.08	.02	.08	1.00	1.00	.08	.02	.08
	120/20	.06	1.00	.95	.06	.03	.06	1.00	1.00	.06	.02	.06
	40/10	.05	1.00	.27	.15	.01	.14	1.00	.81	.15	.04	.15
	40/20	.07	1.00	.52	.10	.02	.10	1.00	.98	.10	.03	.10
	80/10	.05	1.00	.54	.10	.06	.10	1.00	.99	.10	.03	.10
	80/20	.06	1.00	.83	.07	.03	.07	1.00	1.00	.07	.03	.07
	120/10	.06	1.00	.71	.08	.03	.08	1.00	1.00	.08	.02	.08
	120/20	.06	1.00	.95	.06	.03	.06	1.00	1.00	.06	.02	.06

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

Table 8. Eta-squared (η^2) (%) of Different Sources for Power, Standard Error, and RMSE When the Covariate Interaction Effect Was at the Between Level, Within Level, and Cross Level Under Cross-Level Measurement Invariance Conditions.

Sources	Between-level covariates interaction				Within-level covariates interaction				Cross-level covariates interaction			
	Power	SE	RMSE	ASR	Power	SE	RMSE	ASR	Power	SE	RMSE	ASR
Overall η^2	96.42	99.82	99.48	86.02	97.16	99.53	99.54		97.33	99.28	99.48	
Magnitude	54.22	0.00	0.03	1.41	31.65	0.00	0.02		51.11	0.01	0.06	
ICC	21.16	48.51	47.20	27.62	0.01	0.02	0.00		0.04	0.10	0.00	
CN	18.96	46.50	47.45	27.37	20.23	49.74	57.69		22.68	54.41	55.84	
CS	1.14	1.59	1.14	1.45	10.81	37.33	29.89		12.00	27.68	29.79	
Invariance	0.11	0.58	0.49	1.41	3.34	9.04	8.87		4.26	11.03	9.58	
Magnitude * CN	0.44	0.01	0.01	2.84	16.26	0.17	0.01		3.61	0.03	0.04	
magnitude * ICC	0.39	0.01	0.01	2.81	0.01	0.01	0.01		0.01	0.03	0.03	
ICC * CN	0.16	2.09	2.88	5.55	0.02	0.01	0.04		0.01	0.09	0.05	
ICC * CS	0.14	0.14	0.15	2.90	0.02	0.01	0.01		0.01	0.03	0.02	
Magnitude * CS	0.04	0.00	0.01	1.40	8.69	0.01	0.00		2.22	0.01	0.00	
CN * CS	0.01	0.28	0.01	2.81	2.49	2.03	1.86		0.45	3.82	3.04	
Invariance * CN	0.01	0.05	0.08	2.84	0.60	0.30	0.55		0.18	1.32	0.65	
Invariance * CS	0.04	0.04	0.00	1.40	0.56	0.76	0.57		0.05	0.61	0.33	
Invariance * ICC	0.04	0.01	0.01	2.80	0.01	0.01	0.00		0.00	0.10	0.04	
Invariance * magnitude	0.00	0.01	0.01	1.41	2.46	0.09	0.02		0.70	0.01	0.01	

Note. ICC = intraclass correlation; CN = number of clusters per group; CS = cluster size; RMSE = root mean square error; SE = standard error; invariance = cross-level measurement invariance status. Bold values indicate that they are statistically significant.

Type I Error. Across all simulation conditions of this study, Type I error rates were well controlled in multilevel MIMIC modeling. Type I error rates were around .06 ranging from .04 to .08 (see Tables 2, 4, and 6). The design factors appeared to have no impact on Type I error rates. 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 rates varied considerably as a result of the design factors with a minimum of .14 and a maximum of 1.00 as presented in Table 2. The ANOVA with eta-squared analysis indicated that interaction effect magnitude, ICC level, and CN were significantly associated with the power as shown in Table 8. When all other factors were kept constant, the power became larger with a larger interaction effect, larger CN, and smaller ICC. CN had a stronger impact on power than CS (or total sample size).

When the two covariates were at the within level or across levels, the power differed substantially depending on the interaction magnitude, CN, and CS. When the within-level interaction magnitude was .60, the power rates were 1.00 with a few exceptions (.95). On the other hand, when the interaction magnitude was .30, both CN and CS were positively associated with the power (.46-1.00). ICC appeared to have no impact on the power. For the cross-level interaction, the power ranged from .22 to .83 depending on the total sample size when the interaction effect was .20; from .60 to 1.00 when the interaction effect increased to .40.

With respect to the power, the location of the interaction effect was the most important factor. When the interaction occurred at the within or cross level, the power was much greater than that at the between level. This was expected because the sample size at the between level was based on the between-level units (i.e., the CN), whereas the total sample size was used at the within and cross levels.

Standard Error. In the between-level interaction conditions, SE ranged from .11 to .38 (see Table 2) and varied as a function of the ICC and CN as seen in Table 8. SE increased with the increasing level of ICC, and it decreased as CN increased. SEs were very similar in the two different interaction magnitudes of .30 and .60.

In the within-level interaction conditions, SE ranged from .07 to .17, smaller than that in the between-level interaction conditions, after controlling for all other simulation factors. As shown in Table 8, SE was influenced by CN and CS. ICC had no impact on the SE. In the cross-level conditions, SE was very similar to that in the within-level conditions.

Relative Bias. Relative bias was negligible across most conditions, except a few conditions with a small sample size in the between-level interaction conditions. In the between-level interaction conditions, relative bias ranged from $-.06$ to $.08$ when the interaction effect was .30, and it was smaller when the interaction effect was .60, ranging from $-.04$ to 0 (see Table 2).

Root Mean Square Error. When the interaction effect was at the between level, as shown in Table 2, RMSE ranged from .11 to .39. A smaller ICC level and larger total sample size, CN in particular, were associated with a smaller RMSE value. When the interaction effect was at the within level, the values of RMSE ranged from .07 to .17 (see Table 4), smaller than those in the between-level conditions. ICC and the magnitude of interaction effect seemed not to affect RMSE. Similar values and patterns of RMSE were observed in the cross-level interaction conditions as in the within-level conditions.

The Results of Between-Level, Within-Level, and Cross-Level Interaction Under Cross-Level Measurement Noninvariance Conditions

When the within-level factor loadings and the between-level factor loadings were not equal in multilevel MIMIC, that is, when cross-level measurement invariance was violated, the performance of multilevel MIMIC models in detecting the interaction seemed comparable to the cross-level measurement invariance conditions as presented in Tables 3, 5, and 7. In fact, the power of detecting interaction effect was slightly higher than in most of the cross-level measurement invariance conditions of between-level, within-level, and cross-level interaction, especially for cross-level interaction conditions. The average power for the smaller interaction magnitude (.3 for the between-level and within-level interaction, and .2 for the cross-level interaction) was .02, .09, and .15 higher than cross-level measurement invariance conditions for between-level, within-level, and cross-level interactions, respectively. Standard error and RMSE were a little lower (.03 or less) in cross-level measurement noninvariance conditions than in cross-level measurement invariance conditions as shown in Tables 3, 5, and 7. Type I error rates, ASR, and relative bias were similar for cross-level measurement invariance and noninvariance conditions.

Empirical Example

To illustrate the use of multilevel MIMIC modeling approach in estimating interaction effect, we present an example using seventh grade Longitudinal Study of American Youth (LSAY). LSAY aims to study the development of students' attitudes toward and achievement in science and mathematics (Miller, 2010). The latent factor used in this analysis was students' perception of the usefulness of mathematics (LSAY variables: GA32H, GA32I, GA32K, and GA32L). The four items have five response categories (strongly agree, agree, not sure, disagree, strongly disagree). All four items were reverse coded so that a higher value indicated a higher rating of the usefulness of mathematics (see Table 9 for a list of the items and their descriptive statistics). The skewness of all items fell between 0 and 1, and their kurtosis values were between -1 and 1. For demonstration purposes, the data were treated as continuous. Mplus 7.1 (Muthén & Muthén, 1998-2012) was used for data analysis.

Table 9. Descriptive Statistics of the Four Items in the Empirical Example ($n = 2,173$).

Item	Mean	SD	Skewness	Kurtosis
GA32H Math is useful in everyday problems	3.65	0.94	−.63	.10
GA32I Math helps logical thinking	3.70	0.90	−.76	.69
GA32K Need math for a good job	3.79	0.92	−.64	.22
GA32L Will use math often as an adult	3.77	0.96	−.61	.15

The data set included 2,173 students (50.35% female and 49.65% male) nested in 143 teachers (48.95% female and 51.05% male). Teachers with five or more students were included in this study so that the CS was similar to that in the simulation study. Teachers instead of schools were used as the between-level unit of analysis for demonstration purpose because the teacher questionnaire provided covariates information about teachers. The two within-level dichotomous covariates were students' gender and their parents' encouragement of work on mathematics (29.68% no and 70.32% yes). The two between-level covariates are teachers' gender and teachers' master's degree status (37.06% no, 46.15% yes, and 16.78% missing). We hypothesized that students' gender, their parents' encouragement, and the interaction of students' gender and their parents' encouragement had an impact on students' perception of the usefulness of mathematics. Similarly, teachers' gender, teachers' master's degree status, and the interaction of teachers' gender and teachers' master's degree status were hypothesized to affect the students' average perception of the usefulness of mathematics. Note that the four covariates were dichotomous variables, and for gender 0 and 1 represented female and male, respectively. For parents' encouragement and teachers' master's degree status, 0 and 1 represented no and yes, respectively. For demonstration purposes, we selected covariates based on the availability, but applied researchers are supposed to select covariates in the modeling on theoretical basis.

As a preliminary analysis, measurement invariance was conducted to ensure the equality of measurement model for the within groups (female vs. male students; students with vs. without their parents' encouragement), and for the between groups (female vs. male teachers; teachers with vs. without master's degree). Strong measurement invariance held for both within groups and between groups. Because measurement invariance testing using multilevel data is not the focus of the study, the detailed results are not reported, but they are available on request from the first author. After checking tha the measurement invariance held at both the within and between levels, we started to model the effects of the covariates and their interaction effect on the latent factor of students' perception of mathematics.

First, we examined the effects of the within-level covariates. We hypothesized that the impact of parents' encouragement on students' perception of mathematics differed for boys and girls. Students' gender, their parents' encouragement, and the interaction of students' gender and their parents' encouragement were specified to

predict the within-level factor. The interaction effect was specified by creating the product of students' gender and their parents' encouragement with the DEFINE statement in Mplus (see the appendix for the Mplus syntax about how to create the interaction effect). This new interaction variable in addition to students' gender and their parents' encouragement should be listed as within-level variables (WITHIN = under the VARIABLE statement).

The results showed good model fit (CFI [comparative fit index] = .95, RMSEA [root mean square error of approximation] = .07, SRMR_{within} [standardized root mean square residual] = .02, and SRMR_{between} = .05). The interaction between students' gender and their parents' encouragement had a statistically significant effect ($\gamma = .15$, $p = .03$, $d = .24$) on students' perception of mathematics. Note that the regression effect of the covariate and its effect size are denoted as γ and d , respectively. Thus, the interpretation of the effects of the two covariates should be made with caution. The main effect of students' gender on the perception of mathematics was $-.12$ ($p = .06$, $d = -.18$). The main effect parents' encouragement was $.04$ ($p = .45$, $d = .06$). The statistically significant interaction effect implied that the effect of parents' encouragement on students' perception of mathematics depended on students' gender. For female students, the perception of mathematics was $.04$ higher for those who were encouraged by their parents than those who were not encouraged. For male students, the perception of mathematics for those who received their parents' encouragement was $.19$ higher than those who did not receive it.

Second, the two between-level covariates (teachers' gender, and their master's degree status) and their interaction effect (product of teachers' gender and their master's degree status) were specified to have an effect on the between-level factor (see the appendix for the Mplus syntax). Similar to the within-level interaction, the new between-level interaction variable should be specified as a between-level variable along with teachers' gender and their masters' degree status (BETWEEN = under the VARIABLE statement in Mplus).

The results showed inadequate model fit (CFI = .88, RMSEA = .11, SRMR_{within} = .09, and SRMR_{between} = .08). None of the main effects of teachers' gender and teachers' master's degree and their interaction effect was significant (the effect of teachers' gender on the between-level latent factor was $-.03$ ($p = .90$, $d = -.39$); the effect of teachers' master's degree status on the latent factor was $.05$ ($p = .88$, $d = .65$); and the interaction effect was $-.08$ ($p = .86$, $d = -1.03$). Based on the simulation results of between-level interaction effect conditions, it is not surprising that an interaction effect of this magnitude was not statistically significant. However, the estimated effect size of this between-level interaction was quite big.

Third, the cross-level interaction was specified to test if the between-level covariate moderated the within-level effect of students' gender on their perception of mathematics. The between-level covariate of teachers' master's degree status was chosen because it had a relatively larger effect on the between-level latent factor than teachers' gender. In the cross-level interaction model, we hypothesized that the effect of students' gender on the perception of mathematics differed as a function of

teachers' master's degree status. It should be kept in mind that the two variables, students' gender and teachers' master's degree status were specified as WITHIN and BETWEEN variables, respectively, in Mplus. The random effect of students' gender on the within-level perception was defined in the within-level modeling (see the appendix for the Mplus syntax). This effect was allowed to vary across teachers as a random effect and its variability was modeled at the between level. In other words, in the between-level modeling, this random within-level effect of students' gender on their perception was specified to be regressed on the between-level covariate of teachers' master's degree status. The effect of teachers' master's degree status on the random effect of students' gender on their perception of mathematics ("s on master" in the Mplus syntax) represented the cross-level interaction. The results suggested that the cross-level interaction effect was statistically significant ($\gamma = .14, p = .02$). Again, the effect of students' gender on their perception of mathematics should be interpreted with caution because of the significant interaction effect. The estimate of the main effect of the students' gender on their perception of mathematics was $-.09$ ($p = .07$). Thus, for the students whose teachers had no master's degree, the effect of students' gender on their perception was $-.09$; however, for the students whose teacher had a master's degree, the effect of students' gender on their perception was $.05$ (.14 higher). Thus, in the classrooms with teachers without a master's degree, boys' perception of mathematics was $.09$ lower than that of girls. On the other hand, in the classrooms with teachers having a master's degree, boys' perception of mathematics was $.05$ higher than girls' perception.

Note that model fit indices such as CFI, RMSEA, and SRMR were not available in cross-level interaction effect model because numerical integration was entailed. To evaluate the fit of the cross-level interaction effect model, the models with and without the interaction effect were compared using Akaike information criterion (AIC), Bayesian information criterion (BIC), and sample size-adjusted BIC. Lower indices indicates better fitting models. All the indices supported the model with the interaction as a better fitting model (AIC was 18,075 and 18,081, BIC was 18,186 and 18,187, and sample size-adjusted BIC was 18,122 and 18,126, respectively, for the models with and without the cross-level interaction effect).

In sum, there was a statistically significant interaction effect between students' gender and their parents' encouragement on their perception of mathematics. In addition, the cross-level interaction effect of teachers' master's degree status on the effect of students' gender on their perception was statistically significant. However, the between-level interaction effect between teachers' gender and their master's degree status was not statistically significant. The nonsignificant between-level interaction may be because there was no effect or may be because of limited power in testing interaction at the between level. Recall in the simulation study, for interactions of similar sizes, the power was substantially higher for within- and cross-level interactions than between-level interactions.

Discussion and Conclusion

This study purported to examine the efficacy of multilevel MIMIC models in detecting the covariates interaction effect, and to demonstrate the detailed modeling of interaction effect using an empirical example. Prior research has demonstrated the importance of using multilevel MIMIC modeling rather than traditional single-level MIMIC modeling with multilevel data when there was one within-level or one between-level covariate related to the latent variable (Finch & French, 2011). This study extended the prior research by examining the performance of multilevel MIMIC models in estimating the interaction effect when the covariates were at the within level, at the between level, or at both levels.

The performance of multilevel MIMIC modeling was more impressive with the within-level interaction effect than with the between-level interaction effect. The power of the within-level interaction effect was considerably higher than that of the between-level interaction effect, and thus, the power at the within level was less affected by other simulation design factors. This finding is also consistent with the literature (Finch & French, 2011; Konstantopoulos, 2006). Of note is that the behaviors of multilevel MIMIC models with the cross-level interaction effect were very similar to those with the within-level interaction effect. Thus, the substantively meaningful cross-level interaction can be modeled flexibly and be estimated accurately in multilevel MIMIC models with a relatively large sample size. With the increasing prevalence of examining cross-level interaction effect in applied psychology (Mathieu et al., 2012), the decent performance of multilevel MIMIC models in estimating cross-level interaction allows the applied researchers to apply it more comfortably.

Given the low power of multilevel MIMIC models in detecting the between-level interaction, especially when the interaction effect was .30, an alternative modeling method was explored for higher power. That is, we conducted one-level MIMIC modeling by aggregating data in each cluster for two selected conditions (an interaction effect of .30, medium ICC, 80 clusters, and CS of 10 and 20). The power increased slightly from .27 to .28, and from .31 to .34, respectively, for the two conditions. Aggregating data did not result in an obvious improvement of the performance. Thus, when the CN is not big enough (e.g., less than 100), ICC level is high, and the interaction effect is expected to be small, researchers should be aware of the low power of the between-level interaction effect. On the other hand, the parameter estimates were accurate with multilevel MIMIC modeling, and thus, replications of studies and meta-analytic models could lead to more precise estimates.

Regarding the impact of simulation factors on the performance of multilevel MIMIC models, ICC is worthy of note first. ICC was highly associated with the SE, RMSE, and power for the between-level interaction, but not for the within-level or cross-level interaction conditions. In this simulation study, ICC was manipulated by varying the between-level factor variance holding the within-level factor variance at one. When the effect of the interaction on the latent factor is tested for statistical significance, the estimated standard error is the function of the latent factor variance and associated sample size. Thus, it is not surprising that ICC had no impact on the

estimated standard errors at the within level because the within-level factor variance was constant regardless of ICC levels. Subsequently, ICC had no impact on power. Only the impact of sample size was observed. On the other hand, as the between-level factor variance increased with higher ICC, the corresponding standard error became larger given the same sample size. With larger standard errors, the power to detect the between-level interaction effect became lower. The negative impact of ICC on power at the between level has been well documented in the multilevel modeling literature (Finch & French, 2011; Hedges & Hedberg, 2007; Spybrook, 2008).

The impact of CN and CS on power in detecting the interaction effect varied depending on the location of the interaction effect in multilevel MIMIC models. To be more specific, CN was more important than CS in detecting the between-level interaction effect, while both CN and CS were important in estimating the within-level and cross-level interaction effect. Confirming well-known findings of sample size in the multilevel modeling literature, we reiterate common recommendations on the sampling strategies and allocation of resources for researchers. If the researcher is interested in estimating between-level interaction effect, more clusters should be sampled if resources allow (sampling more clusters may be more expensive than sampling more individuals from a smaller CN). If the researcher is interested in estimating within-level or cross-level interaction, more thorough sampling of individuals in clusters (which may be less costly than sampling more clusters) enables an enhanced power because the total sample size is an important factor at the within level.

When cross-level measurement invariance did not hold (the between-level factor loadings (.80) were higher than the within-level factor loadings (.60)), the performance of multilevel MIMIC models had similar patterns as those of the cross-level measurement invariance conditions with slightly higher power. However, a closer examination of the estimated model parameters of cross-level measurement invariance and noninvariance conditions showed that the within-level factor variance was lower in the noninvariance conditions than in the invariance conditions. A lower within-level factor variance was associated with a lower standard error of the estimate, which resulted in a higher power of detecting interaction effect in the cross-level measurement noninvariance conditions. Thus, the observed higher power in cross-level measurement noninvariance conditions should be interpreted with caution. Overall, it appears that cross-level measurement noninvariance did not make a notable impact on the estimation of interaction in the structural part of multilevel MIMIC models if factor loadings were allowed to be different across levels. To further explore the impact of cross-level measurement noninvariance on detecting interaction effect in multilevel MIMIC models, conditions with more different factor loadings were simulated and analyzed. That is, the between-level factor loadings were increased from .80 to .90, and the within-level factor loadings were kept the same at .60. The performance of multilevel MIMIC models in detecting and estimating an interaction effect was similar to that of the previous cross-level measurement noninvariance conditions with between-level factor loadings at .80 and the within-level factor loadings at .60. It seems that the measurement part in multilevel MIMIC

modeling did not affect its performance in detecting and estimating the structural part, or the interaction effect, as long as the measurement part is correctly specified (e.g., allows factor loadings to be different across levels). This was consistent with a previous study by Kim and Cao (2015) in which they indicated that multilevel MIMIC models showed robustness to factor variance noninvariance and residual variance noninvariance in estimating latent factor mean differences.

The generalization of the performance of multilevel MIMIC models beyond the design of this study should be made with caution. Specifically, there was only one factor with six continuous indicators. 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 interaction conditions. Even though similar behaviors of multilevel MIMIC models are expected in detecting and estimating interaction effects (i.e., structural parameters) if the measurement model is reasonable and correctly specified, the impact of changing the design factors mentioned above on the performance of multilevel MIMIC models is unknown. More comprehensive simulation studies varying the measurement and structural parts of the model (e.g., severely nonnormal or categorical data) are called for.

In addition, the low power seen in some conditions, particularly for the between-level interaction, calls into question the possibility that researchers who build models may miss the interaction effect and proceed to interpret a simpler, but misspecified, model that does not include the interaction. The degree to which such misspecification would create biases would be an important consideration in further research on multilevel MIMIC modeling.

In sum, when the within-level or cross-level interaction effect is of focal interest, unbiased estimates and decent power to detect the effects are expected if sample size is sufficiently large. If the expected size of the interaction effect is small (.3), a total sample size of about 800 is recommended for a power over .80. With an effect size .60, a total sample size of 400 is sufficient for a power near 1.00 and a bias close to 0. For the between-level interaction effect, applied researchers should be cognizant of the low power when the expected effect size is small and the ICC is substantially high. In such cases, it is recommended to focus on an interval estimate of the estimated effect size and recognize that more precision in the estimate can be obtained through replications of the study. It is recommended that when the total sample size is decent (e.g., 800), applied researcher can proceed to model within-level and cross-level interaction when the CS reaches 5 and above.

Appendix

Mplus Syntax for Within-Level Covariates Interaction in Multilevel MIMIC

title: within MLMIMIC

data: file is foranalysis.txt;

format is 1f4 1f5 8f1;

variable: names are id tchid tchsex master stsex parent v1-v4;

```

usevariables are v1 v2 v3 v4 stsex parent int;
missing are .;
cluster=tchid;
within=stsex parent int;
define: int=stsex*parent;
analysis: type=twolevel;
model:
  %within%
  fw1 by v1-v4;
  fw1 on stsex parent int;
  %between%
  fb1 by v1-v4;
  [v1@0];
  [fb1 *];
output: sampstat standardized;

```

Mplus Syntax for Between-Level Covariates Interaction in Multilevel MIMIC

```

title: between MLMIMIC
data: file is foranalysis.txt;
  format is 1f4 1f5 8f1;
variable: names are id tchid tchsex master stsex parent v1-v4;
  usevariables are v1 v2 v3 v4 tchsex master int;
  missing are .;
  cluster=tchid;
  between=tchsex master int;
define: int=tchsex*master;
analysis: type=twolevel;
model:
  %within%
  fw1 by v1-v4;
  fw1@1;
  %between%
  fb1 by v1-v4;
  fb1 on tchsex master int;
  [v1@0];
  [fb1 *];
output: sampstat standardized;

```

Mplus Syntax for Cross-Level Covariates Interaction in Multilevel MIMIC

```

title: cross-level interaction MLMIMIC
data: file is foranalysis.txt;

```

```

format is 1f4 1f5 8f1;
variable: names are id tchid tchsex master stsex parent v1-v4;
usevariables are v1 v2 v3 v4 master gender;
missing are .;
cluster=tchid;
within=stsex;
between=master;
analysis: type=twolevel random;
algorithm=integration;
model:
  %within%
  fw1 by v1-v4;
  s|fw1 on stsex; !the effect of gender on fw1 was defined as s;
  %between%
  fb1 by v1-v4;
  fb1 on master;
  s on master;
  s; !the effect of gender on fw1 is modeled as a random effect at the between level;
  fb1;
output: sampstat standardized;

```

Authors' Note

Chunhua Cao is now affiliated with University of South Florida, Tampa, FL, USA.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- 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.
- 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., Dülmer, H., Schlüter, 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.
- 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.
- 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.
- Heck, R., & Thomas, S. (2000). *An introduction to multilevel modeling techniques*. Mahwah, NJ: Erlbaum.
- Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, 29, 60-87.
- 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, NY: Springer Verlag.
- 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.
- 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.
- 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., Dedrick, R. F., Cao, C., & Ferron, J. M. (2016). Multilevel factor analysis: Reporting guidelines and a review of reporting practices. *Multivariate Behavioral Research*, 6, 881-898.
- 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.
- Konstantopoulos, S. (2006). Trends of school effects on student achievement: Evidence from NLS: 72, HSB: 82, and NELS: 92. *Teachers College Record*, 108, 2550-2581.
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample size for multilevel modeling. *Methodology*, 1, 86-92.

- 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*, 951-966.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525-543.
- Miller, J. D. (2010). *Longitudinal Study of American Youth, 1987-1994, and 2007: User Guide (ICPSR 30263)*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
- Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54*, 557-585.
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- 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.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, England: Sage.
- Spybrook, J. (2008). Are power analyses reported with adequate detail? Evidence from the first wave of group randomized trials funded by the Institute of Education Sciences. *Journal of Research on Educational Effectiveness, 1*, 215-235.
- 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.