IMPACT OF NOT FULLY ADDRESSING CROSS-CLASSIFIED MULTILEVEL STRUCTURE IN TESTING MEASUREMENT INVARIANCE AND CONDUCTING MULTILEVEL MIXTURE MODELING WITHIN STRUCTURAL EQUATION MODELING FRAMEWORK

A Dissertation

by

MYUNG HEE IM

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

DOCTOR OF PHILOSOPHY

Chair of Committee,	Oi-Man Kwok
Co-Chair of Committee,	Myeongsun Yoon
Committee Members,	Victor L. Willson
	Eun Sook Kim
Head of Department,	Victor L. Willson

August 2014

Major Subject: Educational Psychology

ABSTRACT

In educational settings, researchers are likely to encounter multilevel data without strictly nested or hierarchical but cross-classified multilevel structure. However, due to the lack of familiarity and limitations of statistical software with cross-classified model, most substantive researchers adopt then the less optimal approaches to analyze cross-classified multilevel data. Two separate Monte Carlo studies were conducted to evaluate the impacts of misspecifying cross-classified structure data as hierarchical structure data in two different analytical settings under the structural equation modeling (SEM) framework.

Study 1 evaluated the performance of conventional multilevel confirmatory factor analysis (MCFA) which assumes hierarchical multilevel data in testing measurement invariance, especially when the noninvariance exists at the between-level groups. We considered two design factors, intra-class correlation (ICC) and magnitude of factor loading differences. This simulation study showed low empirical power in detecting noninvariance under low ICC conditions. Furthermore, the low power was plausibly related to the underestimated ICC and the underestimated factor loading differences due to the redistribution of the variance component from the crossed factor ignored in the analysis.

Study 2 examined the performance of conventional multilevel mixture models (MMMs), which assume hierarchical multilevel data, on the classification accuracy of class enumeration and individuals' class assignment when the latent class variable is at the between (cluster)-level. We considered a set of study conditions, including cluster size, degree of partial cross-classification, and mixing proportion of subgroups. From the results of the study, ignoring a crossed factor caused overestimation of the variance component of the remaining crossed factor at the between-level which was redistributed from the ignored crossed factor in the analysis. Moreover, no SEM statistical program can conduct MMM and take into account of the cross-classified data structure simultaneously. Hence, a researcher should acknowledge this limitation and be cautioned when conventional MMM is utilized with cross-classified multilevel data given the inflated variance component associated with the remaining crossed factor. Implications of the findings and limitations for each study are discussed.

DEDICATION

I dedicated this dissertation to my father who showed me wisdom of life and supported and encouraged me to go through many difficulties in life without giving up and to my mother who loved me a lot through her short life.

ACKNOWLEDGEMENTS

I would like to thank my committee chair and co-chair, Dr. Oi-Man Kwok and Dr. Myeongsun Yoon for their guidance and support throughout the course of the research. My committee member Dr. Victor Willson who have inspired me with intriguing questions and thought-evoking instructions. Especially I thank my committee member, Dr. Eun Sook Kim for her rigorous review for my study throughout my doctoral life as a committee member and her kind encouragement and help whenever I faced in difficulty as a sincere friend.

I would like to express my sincere appreciation to Dr. Jan Hughes. I learned the importance of scientifically-based research and passion as a researcher. The lessons I learned from her are invaluable not only for my study but also throughout my career as an educational researcher. Most importantly, Dr. Hughes had been a mentor with love and warmth for me in both career development and my personal life.

My special thanks go to my husband, Sung Won Lee. The accomplishment I made with this dissertation and my doctoral life could not be possible without his support and care. I appreciate his patience and endurance whenever I went through difficulties in my study and life. I also appreciate him being a great father of my children by taking a good care of them whenever I was not there. Thanks to my sweet and lovely daughter, Seojin and my energetic and curious son, Yoonho. I could enjoy every moment here in Texas A&M University because of my children. I love you, Seojin and Yoonho, forever!

I would like to thank all my family members in Korea who pray for me and make my dream come true. I also appreciate my friends in New Somang Church. With their care and concerns about me and my family, we could enjoy every minute here in College Station. I would like to say thanks to my friends in RMS program who supported with sincere advice and encouragement each other through questions and discussions.

TABLE OF CONTENTS

Page

ABSTRACT ii

DEDICATION iv

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vii

LIST OF FIGURES viii

LIST OF TABLES ix

CHAPTER I INTRODUCTION: IMPORTANCE OF STUDY 1

CHAPTER II STUDY 1: IMPACT OF NOT FULLY ADDRESSING CROSS-CLASSIFIED MULTILEVEL STRUCTURE IN TESTING MEASUREMENT INVARIANCE: A MONTE CARLO STUDY 4

- Theoretical Framework 9
- Method 22
- Results 28
- Discussion 34

CHAPTER III STUDY 2: IMPACT OF NOT FULLY ADDRESSING CROSS-CLASSIFIED MULTILEVEL STRUCTURE IN MULTILEVEL MIXTURE MODELING: A MONTE CARLO STUDY 40

- Theoretical Framework 43
- Method 51
- Results 59
- Discussion 66

CHAPTER IV CONCLUSIONS 74

REFERENCES 76

LIST OF FIGURES

Page

Figure 1. Comparison between Two-level Conventional Multilevel Confirmatory Factor Analysis (MCFA) and Two-level Cross-Classified MCFA. 12

Figure 2. Fitted Two-level Mixture Model with Between-level Categorical Latent Class Variable (CB). 56

LIST OF TABLES

Page

Table 1. Summary of Similarities and Differences between Cross-Classified MCFA and Conventional MCFA 13

Table 2. Summary of Empirical Power Rate of Chi-Square Difference Test and Δ Goodness-of-fit Indices in Factorial Invariance Testing Using Multiple-group MCFA (Misspecified Model) for Non-invariant Condition 29

Table 3. Relative Bias in Factor Loading and Factor Variance in Configural Invariance in Misspecified Model (Conventional MCFA) 32

Table 4. Classification Accuracy Rates of the Conventional Multilevel Mixture Modeling (Misspecified Models) with Between-level Latent Variable 61

Table 5. Summary of Eta-squared (η2) in Analysis of Variance (ANOVA) Results of the Conventional Multilevel Mixture Modeling (Misspecified Models) with Between-level Latent Variable 62

Table 6. Relative Bias of Parameter Estimates of the Conventional Multilevel Mixture Modeling (Misspecified Models) with Between-level Latent Variable 64

CHAPTER I

INTRODUCTION: IMPORTANCE OF STUDY

In educational and other social science research, multilevel data are commonly encountered. Although studies have investigated a variety of multilevel modeling methodological issues, such studies primarily have been limited to hierarchical linear models (Bell, Owens, Kromrey, & Ferron, 2009). Hierarchical linear models (HLMs) assume that in multilevel data, the levels are strictly nested or hierarchical, meaning that a lower-level observation belongs to one and only one higher-level cluster. For example, in education settings, a student may belong to only one class, and a class belongs to only one school. However, in educational settings, researchers are likely to encounter multilevel data without strictly hierarchical but cross-classified structure. For example, students attending the same schools can come from different neighborhoods. Hence, students are nested within schools and neighborhoods while schools and neighborhoods are more likely to be crossed rather than nested with each other. This type of non-strictly hierarchical data structure is named *cross-classified* multilevel structure.

With increased understanding of the importance of proper analytic approach for cross-classified multilevel data (Goldstein, 1986, 1995; Rasbash & Goldstein, 1994; Raudenbush & Bryk, 2002), many major multilevel modeling textbooks have introduced techniques for handling cross-classified multilevel data such as cross-classified random effect modeling (CCREM) in various multilevel modeling computer programs (e.g., HLM, SAS, MLwiN, and R). However, due to the lack of familiarity with this type of model, most substantive researchers adopt a less optimal approach to analyzing this type of data. In other words, instead of taking the full cross-classified data structure into account for their analysis, researchers treat the data as strictly hierarchical by ignoring one of the crossed factor and use the traditional multilevel model for the analysis. For instance, in the above students/schools/neighborhoods example, researchers may ignore the neighborhood information and analyze the data as students only nested within schools. Previous studies have shown that ignoring one of the crossed factors/levels (the neighborhood level in our example) can result in biased estimation of the random effect variances, which in turn, can lead to biased estimation of the standard errors of the fixed effects (or regression coefficients) in the model (Luo & Kwok, 2009).

In addition, due to the limitations of the structural equation modeling (or latent variable modeling) statistical software, researchers may still not be able to completely take the cross-classified data structure into account in their analysis even though they understand the importance of fully addressing the cross-classified structure in their analysis. Research examining the impact of misspecifying cross-classified multilevel data as strictly hierarchical multilevel data in different analytical settings such as structural equation modeling (SEM) framework has been limited. Hence, more controlled empirical investigations of cross-classified data structure in different analytical settings within SEM framework are warranted.

Two separate Monte Carlo studies were conducted to evaluate the impacts of inappropriately analyzing the cross-classified data in two different analytical settings: (a) testing measurement invariance, and (b) conducting multilevel mixture models. Study 1 evaluates the performance of conventional multiple-group multilevel confirmatory factor analysis (MCFA) in testing measurement invariance, especially when the non-invariance is present at the between-level groups. Testing measurement invariance is a very important step before one can meaningfully compare the (mean) difference on a latent construct or the corresponding composite score between groups. However, the conventional multiple-group MCFA available in computer packages is used for only hierarchical or strictly nested multilevel data. Study 2 examines the performance of conventional multilevel mixture models (MMM) on the classification accuracy of identifying correct selection of best solution and of identifying correct individual group/membership assignment when the categorical latent class occurs at the organizational level. Mixture modeling is a relatively new exploratory analytical approach and has been gaining more attention recently in educational research. Mixture models, which are sometimes viewed as a more general form of the traditional cluster analysis, can be used for uncovering the latent groups/unobserved classes based on a specific model. MMM are now used for analyzing only hierarchical or strictly nested multilevel data under the mixture modeling framework. In current computer packages, up to now, there is no statistical program which can conduct these two types of analyses and take into account the cross-classified data structure simultaneously. Hence, it is important to examine the potential impact of ignoring the cross-classified data structure in these two commonly used analytical approaches.

CHAPTER II

STUDY 1: IMPACT OF NOT FULLY ADDRESSING CROSS-CLASSIFIED MULTILEVEL STRUCTURE IN TESTING MEASUREMENT INVARIANCE: A MONTE CARLO STUDY

In educational and other social science research, multilevel data are commonly encountered. Although studies have investigated a variety of methodological issues related to a multilevel modeling, such studies primarily have been limited to hierarchical linear models (Bell et al., 2009). Hierarchical linear models (HLMs) assume that in multilevel data, the levels are strictly nested or hierarchical, meaning that a lower-level observation belongs to one and only one higher-level cluster. For example, in education settings, a student belongs to only a particular classroom while that classroom belongs to only a particular school. However, multilevel data may not always have a strictly nested or hierarchical structure, especially in education settings. For example, students are more likely to be nested within the schools they attend and the neighborhoods where they live at the same time, while schools and neighborhoods are not nested within each other. Instead, schools and neighborhoods are cross-classified with each other at the same level. This type of non-hierarchical multilevel structured data is also called *cross-classified* multilevel structured data.

Testing measurement invariance has become increasingly common in social science research when a measure is used across subgroups of a population and different time points of repeated measures. Measurement invariance refers to the equivalent probability of an observed score in a test given identical ability, regardless of group membership of an observation (Meredith & Millsap, 1992). In other words, measurement invariance holds when persons of the same ability on a latent construct have the identical probability of obtaining the observed score regardless of the group membership. Testing measurement invariance is a very important step before one can meaningfully compare the (mean) difference on a latent construct or the corresponding composite score between groups. The use of a measure with measurement bias (i.e., noninvariance) might lead to invalid comparison. In other words, when measurement invariance is violated, observed changes in latent construct or composite scores from items between subgroups or across time are ambiguous and difficult to interpret (Meredith & Teresi, 2006). Therefore, it is important to confirm that the scale we use measures the same latent construct (or has exactly the same meaning) to the groups we intend to compare. Although many researchers have discussed the importance of establishing measurement invariance and the practical impact of the measurement bias (Borsboom, 2006; Fan & Sivo, 2009; Meredith & Teresi, 2006; Widaman & Reise, 1997; Yoon & Millsap, 2007), there is very limited research on measurement invariance in multilevel data with non-hierarchical structure.

For measurement invariance testing in hierarchical multilevel data, multilevel confirmatory factor analysis (MCFA, Mehta & Neale, 2005; Kim, Kwok, & Yoon, 2012a) is widely used because MCFA assumes that multilevel data have hierarchical structure. However, in reality, multilevel data may not always have a strictly hierarchical structure, particularly in research situations where lower-level observations are nested within multiple higher-level clusters (e.g., schools and neighborhoods or items and raters) that are cross-classified simultaneously at the same level. For example, a researcher might be interested in examining factors that influence individuals' achievement scores while taking into account the contextual effects of their environments. In this case, students at a given school may not all come from the same neighborhood, and instead belong to various combinations of neighborhoods and schools. Thus, students are nested within schools and neighborhoods at the same time, while schools and neighborhoods are not nested within each other but crossed with each other (Raudenbush & Bryk, 2002). When levels of multilevel data are strictly hierarchical or nested, conventional multilevel modeling can be used to model the clustering effects. However, conventional multilevel modeling can only handle multilevel data that have a strictly nested or hierarchical structure. Conventional multilevel modeling cannot account for the effects of multiple cluster factors simultaneously when these multiple cluster unit factors are crossed at the same level. In the example above, conventional multilevel modeling treats cross-classified multilevel data as strictly nested multilevel data by ignoring one of the crossed factors (e.g., either schools or neighborhoods; items or raters) in the analysis.

A few methodological investigations have been conducted to examine the implications of misspecifying cross-classified multilevel data as strictly hierarchical multilevel data by ignoring one of the crossed factors in regression analysis (Berkhof & Kampen, 2004; Meyers & Beretvas, 2006; Luo & Kwok, 2009). In general, these studies have found that not taking a fully cross-classified multilevel data structure into account (i.e., treating the cross-classified data as strictly hierarchical by ignoring a crossed factor) can cause bias in variance component estimates, which results in biased estimation of the standard errors of parameter estimates. Ultimately, this may lead to incorrect statistical conclusion. Especially, Luo and Kwok's (2009) simulation study found that under the situation in which the crossed factors were completely cross-classified, all variance components from the ignored crossed factor at the higher level were redistributed and added to the variance component at the lower level (i.e., overestimated variance component) while the variance component of the remaining crossed factor was underestimated. Given this redistribution of variance component issue, previous studies have emphasized the importance of properly analyzing cross-classified multilevel data in which the effects of multiple cluster units should be taken into account when multiple cross-classified factors exist at the same level.

With increased understanding of the importance of proper analytic approach for cross-classified multilevel data (Goldstein, 1986, 1995; Rasbash & Goldstein, 1994; Raudenbush & Bryk, 2002), many major multilevel modeling textbooks have introduced techniques for handling cross-classified multilevel data such as cross-classified random effect modeling (CCREM) in various multilevel modeling computer programs (e.g., HLM, SAS, MLwiN, and R). However, research examining the impact of misspecifying cross-classified multilevel data as strictly hierarchical multilevel data in different analytical settings such as structural equation modeling (SEM) framework has been scarce.

In addition, the current capacity of the SEM software does not permit multiple-group comparison along with multilevel confirmatory factor analysis (MCFA) in cross-classified multilevel data, which is a critical feature for testing measurement invariance. Although many researchers understand the importance of establishing measurement invariance for a measure and conducting a correct analysis for cross-classified multilevel structured data, they still generally treat the cross-classified multilevel data as hierarchical multilevel data by ignoring crossed factor(s) in measurement invariance testing under some circumstances such as the limitations/restrictions of current SEM software. Hence, it is important to examine the potential impact of ignoring the cross-classified data structure in conventional multilevel CFA. This study is the first to test measurement invariance with cross-classified multilevel data within multilevel SEM framework.

The primary purpose of the present study is to investigate the performance of the conventional MCFA in testing measurement invariance with cross-classified multilevel data. In MCFA, measurement invariance testing can be conducted at different levels in a multilevel model (Mehta & Neale, 2005). In the present study, we focused on examining measurement invariance testing when non-invariance is present among between-level groups. A Monte Carlo study was conducted to achieve the study purpose. This study also examined the statistical power of testing measurement invariance at the between level and factors that impact the statistical power. Below we first briefly review the conventional MCFA and cross-classified MCFA for measurement invariance testing, followed by the research design and simulation study conditions.

Theoretical Framework

Measurement Invariance Testing

Measurement invariance (MI) refers to the equivalent probability of an observed score in a test given identical ability, regardless of group membership of an interest (Meredith & Millsap, 1992) as shown in the following equation:

$$P(X = x \mid \eta,\ M) = P\ (X = x \mid \eta) \tag{1}$$

Given the latent factor score of η , the conditional probability of the observed score of a particular test is independent of group membership of M.

Linear confirmatory factor analysis (CFA) is a dominant methodological approach for testing measurement invariance (Meredith, 1993; Widaman & Reise, 1997; Yoon, 2008). In linear confirmatory factory analysis, a researcher tests a set of factor models with different levels of invariance constraints. Measurement invariance within a factor model is called factorial invariance. Factorial invariance (FI) can be represented within a linear factor model with mean and covariance structures whereas measurement invariance is a broad term that includes both linear and nonlinear relationships between observed variables and latent factors considering the entire score distribution (Yoon, 2008). Under the linear CFA framework, FI is defined as the equivalence of parameters specified in the model across groups. Thus, researchers check different levels of factorial invariance sequentially depending on the equivalent parameters in the testing of invariance. The levels of FI are discussed later.

Measurement Invariance Testing in Cross-Classified Multilevel Data

One primary issue with assessing the measurement invariance in multilevel data (Curran, 2003; Jones-Farmer, 2010; Kim et al., 2012a; Mehta & Neale, 2005; Selig, Card, & Little, 2008; Zyphur, Kaplan, & Christian, 2008) is data dependency. This is of concern because with multilevel data the measurement bias (i.e., measurement noninvariance) across group membership can be found at different levels, including the individual level, the cluster level (group or organizational level), or both. In other words, measurement invariance can be examined at different levels in multilevel data (Mehta & Neale, 2005), depending on the interest of measurement invariance. Multilevel SEM such as MCFA enables measurement invariance to be tested at different levels.

Like FI testing with conventional multilevel data, the differences across group membership can be found and tested at the individual-level, the organizational-level (cluster or between), or both with cross-classified multilevel data. Greater complexity might arise with FI testing of cross-classified multilevel data due to multiple higher level clusters (e.g., schools and neighbors; items and raters). Specifically, for the multiple higher level clusters, noninvariance can exist at each organizational-level cluster or both. No study has discussed measurement invariance issues with cross-classified multilevel data under the SEM framework. As an extension of conventional MCFA, measurement invariance testing in cross-classified multilevel data can also be conducted within multilevel SEM framework at different levels. In this study, we examine the impact of misspecifying the cross-classified MCFA as a conventional MCFA in FI testing by ignoring one of the crossed factors at the between-level. The next section provides a discussion of the differences and similarities between conventional MCFA and cross-classified MCFA. The simplest two-level conventional MCFA and two-level cross-classified MCFA of a single factor with four observed variables used for the simulation study are illustrated in Figure 1 panel A and panel B, respectively. For the study, we outline two MCFAs for continuous variables.

Comparison between Conventional MCFA and Cross-Classified MCFA

Table 1 provides a summary of the differences and similarities between conventional MCFA and cross-classified MCFA. Consider the example of a two- level conventional MCFA in which students (*i*th) are strictly nested within schools (*j*th) and a two-level cross-classified MCFA in which students (*i*th) are cross-classified by schools (j_1th) and neighborhoods (j_2th). Adopting the notation of Rasbash and Goldstein (1994), $X_{i(j_1 j_2)}$ refers to student outcome, i indexes a within-level unit, j_1 indexes a cluster of crossed factor of school (FB1), and j_2 indexes a cluster of crossed factor of neighbor (FB2). The relation between the observed variables and the latent factors for MCFA and cross-classified MCFA can be expressed as Equations 2a and 2b, respectively.

$$X_{ij} = \tau + \lambda\eta_{ij} + \varepsilon_{\boldsymbol{ij}} \tag{2.a}$$

$$X_{i(j_1 j_2)} = \tau + \lambda\eta_{i(j_1 j_2)} + \varepsilon_{\boldsymbol{i}(j_1 j_2)} \tag{2.b}$$

where X is a matrix of observed scores , τ is a vector of intercepts, λ is a matrix of factor loadings, η is a matrix of latent or common factor scores, and ε is a matrix of

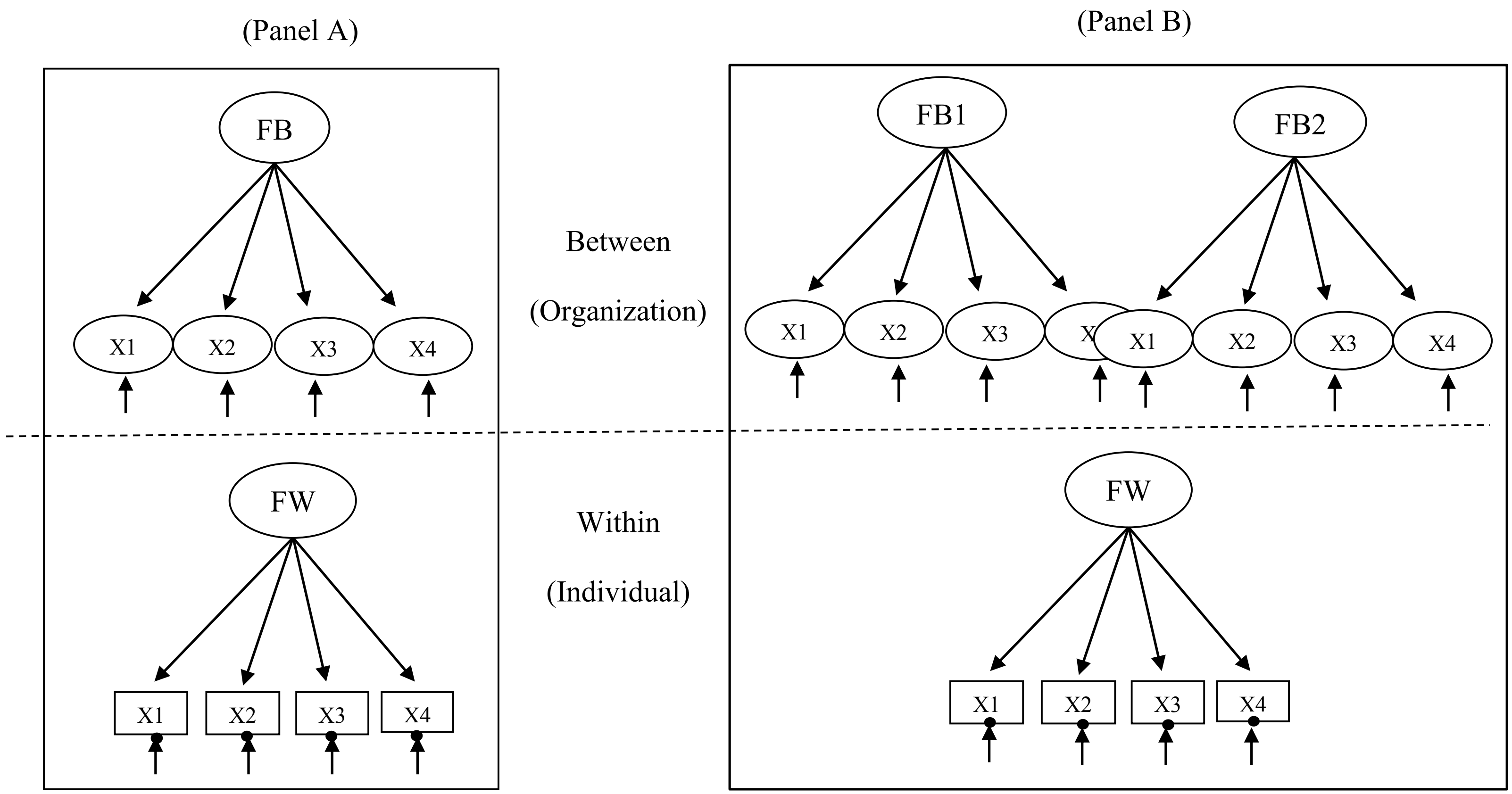

Figure 1. Comparison between Two-level Conventional Multilevel Confirmatory Factor Analysis (MCFA) and Two-level Cross-Classified MCFA.

Note: conventional MCFA depicted in Panel A and cross-classified MCFA depicted in Panel B. FW is within-level latent factor; FB is between-level latent factor; FB1 and FB2 are the two crossed factors1 and 2, respectively, at the between-level. In the within part of the model, X1 - X4 are the continuous observed variables, and the random intercept is shown as a filled circle at the end of the arrow pointing to each observed variable.

Table 1. Summary of Similarities and Differences between Cross-Classified MCFA and Conventional MCFA

Model	Cross-Classified MCFA	Conventional MCFA
Latent factor means: η	$\alpha_{Bj_1} + \alpha_{Bj_1} + \eta_{wij} + \eta_{Bj_1} + \eta_{Bj_2}$	$\alpha + \eta_{wij} + \eta_{Bj}$
Variance of the factor: $V\ (\eta)$	$\Psi_T = \Psi_W + \Psi_{Bj_1} + \Psi_{Bj_2}$	$\Psi_T = \Psi_W + \Psi_B$
Variance of observed scores: $V\ (y)$	$\Sigma_T = \Sigma_w + \Sigma_{Bj_1} + \Sigma_{Bj_2}$	$\Sigma_T = \Sigma_w + \Sigma_B$
Variance of unique factor: $V\ (\varepsilon)$	$\Theta_W + \Theta_{Bj_1} + \Theta_{Bj_2}$	$\Theta_W + \Theta_B$
Measurement model: X	$\tau_{Bj_1} + \tau_{Bj_2}$ $+ \Lambda_{Bj_1}\eta_{Bj_1} + \Lambda_{Bj_1}\eta_{Bj_2} + \Lambda_W\eta_{Wij}$ $+ \varepsilon_{Bj_1} + \varepsilon_{Bj_2} + \varepsilon_{wij}$	τ_B $+ \Lambda_W\eta_{Wij} + \Lambda_B\eta_{Bj}$ $+ \varepsilon_{wij} + \varepsilon_{Bj}$
Covariance structure of measurement model: Σ	$\Sigma_{Bj_1} = \Lambda_{Bj_1}\Psi_{Bj_1}\Lambda_{Bj_1}^{'} + \Theta_{Bj_2},$ $\Sigma_{Bj_2} = \Lambda_{Bj_2}\Psi_{Bj_2}\Lambda_{Bj_2}^{'} + \Theta_{Bj_2},$ $\Sigma_W = \Lambda_W\,\Psi_W\,\Lambda_W^{'} + \Theta_W$	$\Sigma_B = \Lambda_B\Psi_B\Lambda_B^{'} + \Theta_B$ $\Sigma_W\ \Lambda_W\,\Psi_W\,\Lambda_W^{'} + \Theta_W$

unique factor scores or residuals (Kaplan, 2009). It is assumed that the observed variables are multivariate normally distributed. In multilevel CFA, the assumption that observations are independent and identically distributed (Muthén, 1994) should be relaxed with the multilevel data where lower observations nested within higher-level cluster are dependent/correlated each other. No interaction between school and neighborhood is assumed.

By allowing random effects to vary across clusters, the latent factor scores (η_{ij}) for MCFA can be decomposed into two parts (Equation 3.a) whereas the latent factor scores ($\eta_{i(j_1 j_2)}$) for cross-classified MCFA can be decomposed into three parts (Equation 3.b) as follows:

$$\eta_{ij} = \alpha + \eta_{wij} + \eta_{Bj} \tag{3.a}$$

$$\eta_{i(j_1 j_2)} = \alpha_{j_1} + \alpha_{j_2} + \eta_{wij} + \eta_{Bj_1} + \eta_{Bj_2} \tag{3.b}$$

where α is the expected value or grand mean of X_{ij} and α_{j_1} and α_{j_2} are the expected values or mean of each crossed-factor of FB1 and FB2 at the between-level, respectively. η_{wij} is the individual effects for both models, whereas η_{Bj} is for one cluster effect for MCFA and η_{Bj_1} and η_{Bj_2} represent cluster effects of the crossed-factors FB1 and FB2, respectively for cross-classified MCFA. In the same way, observed variable of X_{ij} and $X_{i(j_1 j_2)}$ can also be re-expressed into two parts for MCFA (Equation 4.a), that is, within-level and between- level components and three parts for cross-classified MCFA(Equation 4.b), that is, one within-level and two between-level components as,

$$X_{ij} = \tau_B + \Lambda_W \eta_{Wij} + \Lambda_B \eta_{Bj} + \varepsilon_{wij} + \varepsilon_{Bj} \tag{4.a}$$

$$X_{i(j_1 j_2)} = \tau_{Bj_1} + \tau_{Bj_2} + \Lambda_W \eta_{Wij} + \Lambda_{Bj_1} \eta_{Bj_1} + \Lambda_{Bj_2} \eta_{Bj_2} + \varepsilon_{Bj_1} + \varepsilon_{Bj_2} + \varepsilon_{wij} \tag{4.b}$$

For multilevel CFA, the intercept of an observed variable is only expressed with the intercept (τ_B for MCFA and τ_{Bj_1} and τ_{Bj_2} for cross-classified MCFA) at the between-level. This is because an individual score is the combination of the group mean and its deviation from the group means (Heck & Thomas, 2009). It should be noted that unlike the multilevel model for strictly hierarchical data, in which only one intercept (τ_B) is estimated at the between-level for MCFA, two intercepts (τ_{Bj_1} and τ_{Bj_2} for FB1 and FB2, respectively) are estimated for cross-classified MCFA due to the two crossed factors of FB1 and FB2 at the between-level of an observed variable ($X_{i(j_1 j_2)}$).

Given that factor means vary across clusters as expressed in Equation 3, the variance of the factor can be partitioned into two components for conventional MCFA and three components for cross-classified MCFA as follows:

$$V\left(\eta_{ij}\right) = \Psi_T = \Psi_W + \Psi_B \tag{5.a}$$

$$V\left(\eta_{i(j_1 j_2)}\right) = \Psi_T = \Psi_W + \Psi_{Bj_1} + \Psi_{Bj_2} \tag{5.b}$$

where Ψ_B is the between-level factor variance for MCFA and Ψ_{Bj_1} and Ψ_{Bj_2} are the between-level factor variances for the two crossed factors of FB1 and FB2 in cross-classified MCFA. Ψ_W is the within-level factor variance for both models, and Ψ_T is the total factor variance which is the sum of between-level and within-level factor variances. The ratio of the between variance to the total variance which is called intra-class correlation (ICC) can be viewed as an indicator of data dependency (Raudenbush & Bryk, 2002; Snijder & Bosker, 1999). For the total variability between two crossed factors, two variance components (i.e., Ψ_{Bj_1} and Ψ_{Bj_2} for FB1 and FB2, respectively) are summed up as suggested by Meyer and Beretvas (2006). Under the completely cross-classified situation (i.e., units in a cluster of one crossed factor can affiliate with any clusters of the other crossed factor and vice versa), the two crossed factors (i.e., complete cross-classification situation) are independent from each other. The ICC in MCFA (Equation 6.a) and the two ICCs for each of the two crossed-factors in cross-classified MCFA can be estimated as follows:

$$ICC\eta_B = \frac{\Psi_B}{\Psi_W + \Psi_B} \tag{6.a}$$

$$ICC\eta_{Bj1} = \frac{\Psi_{Bj1}}{\Psi_W + \Psi_{Bj1} + \Psi_{Bj2}}; \qquad ICC\eta_{Bj2} = \frac{\Psi_{Bj2}}{\Psi_W + \Psi_{Bj1} + \Psi_{Bj2}} \tag{6.b}$$

where $ICC_{\eta B}$ refers to the ICC for the between latent factor FB in MCFA while $ICC_{\eta B_{j1}}$ and $ICC_{\eta B_{j2}}$ refer to the ICCs for the two crossed latent factors FB1 and FB2 in cross-classified MCFA, respectively. The variance of the unique factor or residual also comprises two elements for MCFA (Equation 7.a); that is, between-level component and within-level component. Similarly, three elements for cross-classified MCFA; that is, two between-level components and within-level component (Equations 7.b) as,

$$V\left(\varepsilon_{ij}\right) = \Theta_W + \Theta_B \tag{7.a}$$

$$V\left(\varepsilon_{i(j_1 j_2)}\right) = \Theta_W + \Theta_{Bj_1} + \Theta_{Bj_2} \tag{7.b}$$

Subsequently, the variance of observed scores, X_{ij} and $X_{i(j_1 j_2)}$, can be decomposed into the following two elements for MCFA (Equation 8.a) and three elements for cross-classified MCFA (Equation 8.b), respectively as well:

$$V\left(X_{ij}\right) = \Sigma_T = \Sigma_W + \Sigma_B \tag{8.a}$$

$$V\left(X_{i(j_1 j_2)}\right) = \Sigma_T = \Sigma_W + \Sigma_{Bj_1} + \Sigma_{Bj_2} \tag{8.b}$$

Σ_B is the between-level variance matrix for MCFA whereas Σ_{Bj_1} and Σ_{Bj_2} are the between-level variance matrix of crossed-factor FB1 and FB2, respectively for cross-classified MCFA. Σ_W is the within-level variance matrix for both MCFAs, and Σ_T is the total variance matrix, which is the sum of within-level and between-level variance matrices.

Finally, with the independent assumption between the common factor (η) and the unique factor (ε) as in a regular CFA ($Cov\ (\eta, \varepsilon)$ = 0), the covariance structure of the MCFA (Equation 9.a) and cross-classified MCFA (Equation 9.b) is defined as follows:

$$\Sigma_B = \Lambda_B \Psi_B \Lambda_B^{'} + \Theta_{B,}$$

$$\Sigma_W = \Lambda_W \Psi_W \Lambda_W^{'} + \Theta_W,$$

$$\Sigma_T = \Lambda_W \Psi_W \Lambda_W^{'} + \Lambda_B \Psi_B \Lambda_B^{'} + \Theta_W + \Theta_B \quad (9.a)$$

$$\Sigma_{Bj_1} = \Lambda_{Bj_1} \Psi_{Bj_1} \Lambda_{Bj_1}^{'} + \Theta_{Bj_1}; \quad \Sigma_{Bj_2} = \Lambda_{Bj_2} \Psi_{Bj_2} \Lambda_{Bj_2}^{'} + \Theta_{Bj_2},$$

$$\Sigma_W = \Lambda_W \Psi_W \Lambda_W^{'} + \Theta_W$$

$$\Sigma_T = \Lambda_W \Psi_W \Lambda_W^{'} + \Lambda_{Bj_1} \Psi_{Bj_1} \Lambda_{Bj_1}^{'} + \Lambda_{Bj_2} \Psi_{Bj_2} \Lambda_{Bj_2}^{'} + \Theta_W + \Theta_{Bj_1} + \Theta_{Bj_2} \quad (9.b)$$

Factorial Invariance (FI) in Cross-Classified Multilevel Data

In general, FI testing can be conducted using a series of null hypotheses, which impose identical parameters across groups. That is, the models that investigate the invariance of factor pattern (configural invariance), factor loadings (metric or weak invariance), latent intercepts (scalar or strong invariance), and unique or residual factor variances (strict invariance) are tested across groups in the sequential order. In FI testing with cross-classified MCFA, the above null hypotheses test the invariance of factor pattern for configural invariance, within- and between-level factor loadings for metric or weak invariance, between-level intercepts for scalar or strong invariance, and within- and between-level unique variances for strict invariance. As discussed before, because cross-classified MCFA has multiple cluster-level crossed factors (e.g., students nested within schools and neighborhoods; observations nested within items and raters) cross-classified MCFA has multiple cluster-level crossed factors (e.g., students nested within schools and neighborhoods; observations nested within items and raters) compared to only one group-level factor in conventional CFA, for each between-level group comparison the separate FI testing should be conducted to detect the violation of invariance at the different between-level models across the different between-level comparison clusters.

To illustrate FI in cross-classified multilevel data, suppose that a grouping variable exists at the organizational level such as a treatment administered at schools or neighborhoods at both crossed factors. With a grouping variable at the between-level, the two-level cross-classified MCFA (two crossed factors representing each cluster unit such as schools and neighbors at the same level as defined in Equation 2.b through 8.b) can be directly expanded to multiple-group cross-classified MCFA by incorporating a group indicator as such:

$$X_{i(j_1j_2)g} = \nu_{Bj_1g} + \nu_{Bj_1g} + \Lambda_{Bj_1g}\eta_{Bj_1g} + \Lambda_{B_2g}\eta_{Bj_2g} + \Lambda_{Wg}\eta_{Wijg} + \varepsilon_{Bj_1g} + \varepsilon_{Bj_2g} + \varepsilon_{wijg},$$

$$\Sigma_{Bj_1g} = \Lambda_{Bj_1g}\Psi_{Bj_1g}\Lambda'_{Bj_1g} + \Theta_{Bj_1g}, \quad \Sigma_{Bj_2g} = \Lambda_{Bj_2g}\Psi_{Bj_2g}\Lambda'_{Bj_2g} + \Theta_{Bj_2g},$$

$$\Sigma_{Wg} = \Lambda_{Wg}\,\Psi_{Wg}\,\Lambda'_{Wg} + \Theta_{wg} \tag{10}$$

where a subscript, g is a group indicator (1, 2, …, G) and others are as described above.

First, configural invariance evaluates whether the groups of interest have equivalent patterns for the within-level model and the two between-level models (e.g., number of factors in within and between models and number of indicators for each factor). Second, when testing metric/weak invariance, the null hypotheses of the invariance of factor loadings, $H_{0\Lambda}$, can be tested at both between-level crossed-factors FB1 and FB2 and at the within-level, respectively as such

$$H_{0\Lambda_{Bj_1}}: \Lambda_{Bj_1 1} = \Lambda_{Bj_1 2} = \cdots = \Lambda_{Bj_1 G}; H_{0\Lambda_{Bj_2}}: \Lambda_{Bj_2 1} = \Lambda_{Bj_2 2} = \cdots = \Lambda_{Bj_2 G};$$

$$H_{0\Lambda_W}: \Lambda_{W1} = \Lambda_{W2} = \cdots = \Lambda_{WG} \qquad (11)$$

Third, when testing scalar/strong invariance, the null hypotheses of the invariance of intercepts, $H_{0\tau}$, can be tested only at the between-level crossed-factors FB1 and FB2, respectively as such

$$H_{0\tau_{Bj_1}}: \tau_{Bj_1 1} = \tau_{Bj_1 2} = \cdots = \tau_{Bj_1 G}; H_{0\tau_{Bj_2}}: \tau_{Bj_2 1} = \tau_{Bj_2 2} = \cdots = \tau_{Bj_2 G} \qquad (12)$$

Fourth, when testing strict invariance, the null hypotheses of the invariance of unique variances, $H_{0\Theta}$, can be tested at both between-level crossed-factors FB1 and FB2, and at the within-level, respectively as such

$$H_{0\Theta_{Bj_1}}: \Theta_{Bj_1 1} = \Theta_{Bj_1 2} = \cdots = \Theta_{Bj_1 G}, H_{0\Theta_{Bj_2}}: \Theta_{Bj_2 1} = \Theta_{Bj_2 2} = \cdots = \Theta_{Bj_2 G},$$

$$H_{0\Theta_W}: \Theta_{W1} = \Theta_{W2} = \cdots = \Theta_{WG} \qquad (13)$$

In the current study, we investigated sequentially the models of factor pattern (configural) invariance and factor loading (metric/weak) invariance in FI testing since our main interest was on the impact of ignoring a crossed factor on detecting non-invariant factor loading (Λ) at the between-level across the between-level comparison groups.

Evaluations of Measurement Invariance Testing

Measurement invariance under the multilevel SEM framework is typically tested using the likelihood ratio test between a baseline model and sequentially more restricted invariance models. The likelihood ratio (LR) test is also called the chi-square difference ($\Delta \chi^2$) test. In the LR test, $\Delta \chi^2$ follows the chi-square distribution with the degrees of freedom difference (Δ *df*) if the data meet the assumptions, such as multivariate normality. The null hypothesis of the $\Delta \chi^2$ test is that the more restricted invariance model (e.g., same factor pattern and identical factor loadings) fits the data equally well as the less restricted invariance model (e.g., same factor pattern only). When the null hypothesis is failed to reject (non-significant at $\alpha = .05$), we conclude that the more restricted invariance model (e.g., weak invariance model) holds under study. Conversely, when the null hypothesis is rejected (significant at $\alpha = .05$), we conclude that the less restricted invariance model (e.g., configural invariance model) holds under study.

For multiple-group MCFA, the maximum likelihood estimation with robust standard errors (MLR) is employed as an estimator for continuous variables. The MLR yields a robust chi-square test (Kaplan, Kim, & Kim, 2009) by utilizing robust standard errors and a mean-adjusted chi-square statistic test. For comparing two competing models, the MLR requires the Satorra-Bentler scaled chi-square difference test (Brown, 2006; Heck & Thomas, 2009; Satorra & Bentler, 1994). However, it has been widely argued that the $\Delta \chi^2$ test can be too sensitive to the sample size like χ^2 statistic test. Thus, alternatively, the investigation of the changes (Δ) in goodness-of-fit-indices such as the standardized root mean square residual (SRMR) within and between; comparative fit index (CFI), the root mean square error of approximation (RMSEA); and information criteria (AIC and BIC) has been suggested to assess MI in addition to $\Delta\chi^2$.

Method

Data Generation

In the current Monte Carlo study, a two-level cross-classified MCFA pertinent to Equation 2.b was generated using M*plus* version 7 (Muthén & Muthén, 1998-2012). In M*plus* version 7, cross-classified analysis is available using a full SEM for each of the three levels (i.e., within level, between A-level, and between B- level). For the within and between models, we simulated a single factor with four factor indicators (DiStefano & Hess, 2005)of two groups while assuming an equivalent factor structure for all levels. For the population parameters, we referred to previous simulation studies on both measurement invariance and multilevel SEM (Hox & Maas, 2001; Maas & Hox, 2005; Yoon & Millsap, 2007; Kim, Yoon, & Lee, 2012b). In the present study, the factor loadings of the four items ranged from 0.7 through 0.9 at all three levels (within, FB1 and FB2). The factor mean was set to zero for group 1 and 1.0 for group 2 and the unique variances of the four observed variables were all set to 0.25 (Hox & Maas, 2001).

Two main design factors, namely, magnitude of the non-invariant factor loading and intra-class correlation (ICC), were considered. For the simulated invariance testing, the non-invariant target groups existed only at the between-level clusters. Thus, for the invariance (or 0 difference) condition, all parameters were set to be identical across groups. On the other hand, for the non-invariance condition, one of the between-level factor loadings was set to be different across groups for both crossed factors. Based on previous simulation studies of invariance testing (French & Finch, 2008; Meade & Lautenschlager, 2004; Stark, Chernyshenko, & Drasgow, 2006), the magnitude of difference in the target factor loading between the two groups at the between-level (e.g., the difference between λ_{Bj_1}, a particular factor loading at the FB1 crossed factor for group 1 and λ_{Bj_1}, the same factor loading at the FB1 crossed factor for group 2) was simulated at three levels: 0.15, 0.25, and 0.35 difference for small, medium and large, respectively. The group1 (G1) was modeled to have invariant factor structure (or factor loading) across two groups as a reference group while the group 2 (G2) was modeled to have non-invariant factor structure (or factor loading) across two groups as a focal group.

The intra-class correlation (ICC) was another design factor considered in this simulation study. Previous simulation studies have showed that the ICC could affect the statistical power for detecting non-invariance in MCFA, particularly at the between-level (Kim et al., 2012a). Based on previous simulation studies (Hox & Maas, 2001; Maas & Hox, 2005), three ICC conditions were examined in the study: small, medium, and large. The different levels of ICCs were simulated by varying the size of both Ψ_{Bj_1} and Ψ_{Bj_2}, between-level crossed factor variances of FB1 and FB2, respectively. Both crossed factor variances were set to be the same with three different levels: 0.10, 0.25, and 0.50 while the within-level factor variance (Ψ_w) was always fixed (1.00). Based on Equation 6.b, these combinations of variances resulted in three different levels of ICCs: 0.09, 0.17 and 0.25 for small, medium, and large ICC, respectively. These ICC levels represent common situations encountered in educational research with multilevel data.

We constructed a balanced design with two groups of equal size. This design reflects a common research situation in which two groups have similar sample sizes with an unknown direction of possible bias. Although the number of clusters has always been considered in the MCFA and related simulation studies (Hox & Maas, 2001; Maas & Hox, 2005), we only adopted a large number of clusters (i.e., 80 for FB1 and 100 for FB2) and a relatively small cluster size (2 observations per cell) for our data generation. Under complete cross-classified situation, a total of 16000 observations were generated for the two between-level comparison groups, which resulted in a grand sample size of 16000 (i.e., 80 × 100 × 2) for each simulated data set. With this large number of clusters, we could maintain a substantially high recovery rate for the population parameters when generating cross-classified data using M*plus version* 7. By combining the two study conditions (magnitude of non-invariant factor loading and different levels of ICC across levels), a total of 9 (3×3) scenarios were investigated in the study. We generated 1000 replications for each scenario.

Fitted Model

The generated data sets with cross-classified MCFA of two groups were then analyzed using the conventional multiple-group MCFA (i.e., the misspecified model) by ignoring one of the crossed factors (i.e., FB2) and treating it as hierarchical multilevel data. To explore the performance of the conventional multiple-group MCFA with cross-classified data, we used the Type=TWOLEVEL routine in M*plus,* recommended by Kim et al., (2012a) for the conventional multilevel data. When target groups are at the between-level, the TYPE=TWOLEVEL routine decomposes the variance and covariance matrix into within- and between- models for the analysis. Thus, the within- and between-level variance components can be separately investigated. All data analyses were conducted using M*plus* version 7.

Data Analytic Procedures

We examined the statistical power and Type I error rate of the chi-square difference test and goodness-of-fit indices and the relative bias in parameter estimates to explore the performance of multiple-group conventional MCFA (ignoring a crossed factor of cross-classified multilevel data and treating it as a hierarchical data) in detecting the between-level non-invariant factor loading.

Chi-square difference test ($\Delta\chi^2$)

In the present study, the $\Delta\chi^2$ test was used for comparing between the *configural* invariance model (one equality constraint for identification) and the *metric* invariance model (constraining all factor loadings equal across groups) to determine weak (or metric) invariance. If the chi-square difference between these two competing models was statistically significant, the configural invariance model would be rejected and the metric invariance model would be chosen, indicating the presence of a non-invariant factor loading. Under the non-invariant conditions, we expected the null hypothesis of the $\Delta\chi^2$ test to be rejected because one of the between-level factor loadings was simulated to be different (or non-invariant) across groups. The (empirical) statistical power rate was defined as the proportion of the replications in which the $\Delta\chi^2$ test correctly detected the non-invariance over the 1000 generated data sets by rejecting the null hypothesis of equal factor loadings across groups (metric invariance) in the Δ χ^2 tests.

For the invariant condition, Type I error rate was examined. Type I error referred to the proportion of the cases in which the Δ χ^2 test falsely detected invariance as non-invariance over 1000 replications. The invariance at factor loading across groups should lead to failing to the rejection of equal factor loadings across groups (metric invariance) in the Δ χ^2 tests of the misspecified model (multiple-group conventional MCFA). .

Goodness-of-fit indices

Considering the sensitivity of sample size to the χ^2 test statistic, we have additionally examined the performance of the following difference (Δ) of the goodness-of-fit indices in comparing the two competing invariance models (configural versus metric): (a) IC (i.e., AIC and Δ BIC; (b) Δ SRMR between and within; (c) Δ CFI; and (d) Δ RMSEA. The recommended cutoff values for determining the goodness-of-fit indices for a configural invariance model over the metric invariance model are: both ΔAIC and ΔBIC ≤ 4 (Burnham & Anderson, 2002), ΔSRMR ≤ .01 (Chen, 2007), ΔCFI ≤ .01 (Cheung & Rensvold, 2002), and ΔRMSEA ≤ .015(Chen, 2007). Under the non-invariance conditions, if the Δ fit-index was smaller than the cutoff value, this indicated a miss/failure in detecting the non-invaraince. Conversely, when the obtained value was larger than the cutoff value, this indicated a hit/success in detecting the non-invariance.

Relative Bias

The relative bias of the non-invariant factor loading across two groups (i.e.,λ_W at the within-level; λ_{Bj_1} at the between-level) and the factor variance at each level (i.e., Ψ_W

at the within-level; Ψ_{Bj_1} at the between-level) in multiple-group conventional MCFA (misspecified Model). For the non-invariance condition, we set one of the between-level factor loadings to be different across groups for both crossed factors. The relative bias of the target between-level factor loading λ_{Bj_1} is the bias of the between-level factor loading of one group (group2 in the current study) which was set to be smaller (i.e., non-invariant) than the factor loading of the other group (group 1 in the current study). Correspondingly, for the relative bias of the within-level factor loading is the bias of the within-level factor loading of group 2. For the relative bias of the factor variance at the within- and between-level is the bias of within- and between-level factor variance of group 2. For estimating the relative bias, we used the group mean estimates across the replications from configural invariance model. The relative bias of the estimates was computed using the following equation:

$$B(\hat{\beta}) = \frac{\hat{\beta} - \beta}{\beta} \tag{14}$$

where $\hat{\beta}$ was the group mean estimates of non-invariant factor loading and factor variance across the valid replications in the misspecified model, and β was the true population value of the corresponding parameters. To evaluate the estimated relative bias, we applied cutoffs of 0.05 for the loading estimates and of 0.10 for the factor variance estimates which has been recommended by Hoogland and Boomsma (1998) as the acceptable magnitude of relative bias. Relative bias less than the corresponding recommended cutoff value indicates an unbiased estimate of the population parameter. A positive relative bias indicates an overestimation of the target parameter (i.e., factor loading and factor variance in this simulation study), whereas a negative relative bias indicates an underestimation of the target parameter.

Results

Results include information on (a) the empirical statistical power rate for the non-invariant models and Type I error rate for the invariant model and (b) the relative bias for factor loading estimates and factor variance estimates in conventional MCFA (misspecified model) when non-invariance was present at the between-level factor loading across two groups . All analyzed models were successfully converged.

Empirical Statistical Power for Detecting Non-invariance

For non-invariant model, the empirical statistical power was defined as the proportion of the cases in which the non-invariance at the between-level factor loading was correctly detected through the chi-square difference ($\Delta \chi^2$) test and the Δ goodness of fit indices when using the conventional MCFA for testing factorial invariance. Table 2 summarizes the empirical statistical power rate of the $\Delta \chi^2$ test and the Δ goodness-of-fit indices in FI testing using multiple-group conventional MCFA (misspecified Model) with a between-level grouping variable.

For the invariant model, Type I error rate was examined with the chi-square difference ($\Delta \chi^2$) test and the Δ goodness of fit indices. When the between-level factor loadings are invariant across groups, conventional MCFA performed very well in terms of Type I error because Type I error rates were almost zero for all ICC conditions even though one of the crossed factors at the between level was completely ignored and the

Table 2. Summary of Empirical Power Rate of Chi-Square Difference Test and Δ Goodness-of-fit Indices in Factorial Invariance Testing Using Multiple-group MCFA (Misspecified Model) for Non-invariant Condition

Intra-class Correlation (ICC)	Small			Medium			Large		
Difference in Factor Loading	Small	Medium	Large	Small	Medium	Large	Small	Medium	Large
$\Delta\ \chi^2$	0.00	0.00	0.03	0.00	0.97	1.00	0.99	1.00	1.00
p value	(.9411)	(.5664)	(.1570)	(.3744)	(.0105)	(.0000)	(.0046)	(.0000)	(.0000)
Goodness-of-fit indices									
Δ AIC	0.00	0.00	0.00	0.00	0.58	1.00	0.80	1.00	1.00
Δ BIC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.27	1.00
Δ SRMR between	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.02	0.46
Δ SRMR within	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Δ CFI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Δ RMSEA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.26

Note. For factorial invariance testing, we used the conventional multiple-group multilevel CFA with Type=TWOLEVEL routine in M*plus* and compared configural invariance model and metric invariance. For study conditions, the three different levels of Intra-class correlation (ICC) and the three different magnitudes of the non-invariant factor loading between two groups were simulated. The three different level of ICC, 0.09, 0.17, and 0.25 are corresponding to small, medium and large levels; the three different magnitudes of non-invariant factor loading, 0.15, 0.25, and 0.35 are corresponding to small, medium and large loading difference between two groups. $\Delta\ \chi^2$ is Satorra-Bentler scaled chi-square difference test. *p* value is the average *p* value of $\Delta\ \chi^2$ tests across 1000 replications. As for goodness-of-fit Indices, Akaike- and Bayesian-information criteria (AIC and BIC, respectively); the standardized root mean square residual (SRMR) within and between; the root mean square error of approximation (RMSEA); and comparative fit index (CFI) were used.

data were treated as strictly hierarchical.

Chi-square difference test ($\Delta\, \chi^2$)

For the non-invariant condition, chi-square difference testing showed a very high empirical power (0.97 to 1.00) when the magnitude of ICC was medium and large regardless of the magnitude of factor loading differences, except the small factor loading difference along with Medium ICC condition. However, under small ICC condition, the statistical power was extremely low (close to zero) regardless of the magnitude of factor loading differences between groups. The average *p* value of Satorra-Bentler chi-square difference testing across 1000 replications were reported. The cutoff values of $p \leq 0.05$ was used to determine the chi-square difference testing for a configural invariance model over the metric invariance model, indicating a success in detecting the non-invaraince. Conversely, when the *p* value was larger than the cutoff value, this indicated a failure in detecting the non-invariance.

For the invariant condition, the empirical Type I error rate of the chi-square difference test was very close to zero for all conditions which indicated that the invariant model (with invariant factor loadings) was generally not rejected even though the cross-classified data were analyzed with the conventional MCFA (misspecified model) by ignoring one of the crossed factors.

Δ Goodness-of-fit indices

For the non-invariant model, ΔAIC performed well (i.e., detecting the non-invariant factor loading) under large ICC condition regardless the magnitude of the factor loading difference (empirical power ranged from 0.80 to 1.00) while ΔBIC performed well only for large loading difference (empirical power = 1.00) but for small loading difference (empirical power close to zero). Additionally, only ΔAIC still performed well for large (empirical power = 1.00) loading difference under the medium ICC condition. Both Δ information criteria performed poorly under the low ICC condition regardless the magnitude of the loading difference.

For the other four Δ goodness of fit indices (i.e., ΔSRMR-within, ΔSRMR-between, ΔRMSEA, and ΔCFI), only ΔSRMR-between performed little better than any other fit indices under the large ICC condition with large factor loading difference (empirical power = 0.46). For all other conditions, all these Δ goodness of fit indices performed poorly on detecting the non-invariant factor loading (with empirical power ranged from 0.00 to 0.27).

For the invariant condition, the average differences (Δ) in all the goodness-of-fit indices across the 1000 replications were almost zero for all study conditions, which was consistent to the results in the chi-square difference testing. The findings indicated that the conventional MCFA performed well with respect to Type I error control when a crossed factor was omitted in testing measurement invariance.

Relative Bias of Parameter Estimates

Table 3 summarizes the relative bias for the target factor loading estimates and the factor variance estimates from the multiple-group conventional MCFA (misspecified model). For the relative bias calculation, the group mean estimates from the misspecified model were used while the true population parameters were known. We adopted the guidelines for acceptable magnitude of relative bias recommended by Hoogland and

Table 3. Relative Bias in Factor Loading and Factor Variance in Configural Invariance in Misspecified Model (Conventional MCFA)

Study Conditions			Factor Loading				Factor Variance			
Difference in λ_{Bj_1}	ICC	(Ψ_{Bj_2})	Within (λ_W)		Between (λ_{Bj_1})		Within (Ψ_W)		Between (Ψ_{Bj_1})	
			G1	G2	G1	G2	G1	G2	G1	G2
Small	Small	(0.10)	0.00	-0.01	0.00	0.10	0.10	0.10	-0.02	-0.07
	Medium	(0.25)	0.00	-0.03	0.01	0.10	0.25	0.25	-0.01	-0.05
	Large	(0.50)	0.00	-0.06	0.00	0.08	0.51	0.51	-0.03	-0.07
Medium	Small	(0.10)	0.00	-0.02	0.00	0.19	0.10	0.10	-0.02	-0.09
	Medium	(0.25)	0.00	-0.05	0.01	0.18	0.25	0.25	-0.01	-0.08
	Large	(0.50)	0.00	-0.09	0.00	0.15	0.51	0.51	-0.03	-0.09
Large	Small	(0.10)	0.00	-0.03	0.00	0.31	0.10	0.10	-0.02	-0.12
	Medium	(0.25)	0.00	-0.08	0.01	0.29	0.25	0.25	-0.01	-0.10
	Large	(0.50)	0.00	-0.13	0.00	0.24	0.51	0.51	-0.03	-0.11

Note. ICC is Intra-class Correlation. Difference in λ_{Bj_1} is difference in between-level target factor loading across two groups. G1 is group 1 with invariant factor structure (or factor loading) across two groups; G2 is group 2 with non-invariant factor structure where the factor loading of one item was set to be smaller than the factor loading of the G1 in the study. λ_{Bj_1} is the estimated parameter of the target between-level factor loading of one group. Correspondingly, λ_W is the estimated of the within-level factor loading. Ψ_W and Ψ_{Bj_1} is the factor variance at the within- and between-level (remaining), respectively. Ψ_{Bj_2} is the between-level factor variance omitted (FB2) in the analysis.

Boomsma (1998): cutoffs of 0.05 for the factor loading estimates and of 0.10 for factor variance estimates.

First, the relative bias in the factor variance at each level was examined. For group 2 (G2) where a non-invariant factor loading of one item was modeled, the relative bias of the within factor variance, $B(\widehat{\Psi_W})$, ranged from 0.10 to 0.51, while the relative bias of the between factor variance, $B(\widehat{\Psi_{BJ_1}})$, ranged from -0.05 and -0.11. For group 1(G1) where a invariant factor structure was modeled, the relative bias of the within factor variance,$B(\widehat{\Psi_W})$, ranged from 0.10 to 0.51, while the relative bias of the between factor variance, $B(\widehat{\Psi_{BJ_1}})$ was negligible with range from -0.03 and -0.01.

The relative bias of the factor variance at within level were almost identical across two groups for all study conditions. The within factor variance was generally overestimated and greater overestimation associated with larger ICC. On the other hand, the relative bias of the between factor variance in G1 was negatively biased, but mostly within the cutoff (< 0.10) which indicated an unbiased estimate for all study conditions whereas the relative bias of the between factor variance in G2 was negatively biased (i.e., underestimated by 10% to 12%) under the largest factor loading difference condition

Second, the relative bias of the target non-invariant factor loading estimate (G2 in the current study) at each level, namely, the within-level loading ($B(\widehat{\lambda_W})$) and the between-level loading ($B(\widehat{\lambda_{BJ_1}})$), was examined. Irrespective of study conditions, the estimates of λ_W and λ_{Bj_1} for G1 were almost identical to the population true parameter, which yielded almost zero $B(\widehat{\lambda_W})$) and $B(\widehat{\lambda_{BJ_1}})$, respectively. For G2, $B(\widehat{\lambda_W})$ was acceptable only under the small ICC condition, ranging from -0.01 to -0.03 whereas under other conditions, $B(\widehat{\lambda_W})$ was generally underestimated by 5% to 13%. On the other hand, $B(\widehat{\lambda_{BJ_1}})$ was unacceptable (or overestimated) under all the study conditions, ranging from 0.08 to 0.31. $B(\widehat{\lambda_{BJ_1}})$ exhibited larger bias as the magnitude of ICC decreased. Furthermore, the bias of the factor loadings of the invariant items (three items in the study) at both within- and between-level in both G1 and G2 were examined and found to be negligible (close to zero).

Discussion

Testing measurement invariance is a very important step before one can meaningfully compare the (mean) difference on a latent construct or the corresponding composite score between groups. Measurement invariance testing can be utilized to examine possible differences between groups at the organizational units of a particular measure. For between-level grouping comparison, more complexity arises with FI testing in cross-classified multilevel data due to the multiple crossed factors compared to FI testing in conventional multilevel data. Ideally, the multiple crossed factors should be taken into account when conducting FI test. However, up to date, there is no statistical program which can conduct FI testing and take into account of the cross-classified data structure simultaneously. For this reason, researchers often treat cross-classified data as a conventional multilevel data (i.e., as strictly nested or hierarchical) by ignoring one of the crossed factors. Hence, it is important to examine the potential impact of ignoring the cross-classified data structure in FI testing.

As shown in the simulation results, when the between-level factor loadings are invariant across the between-level comparison groups, conventional MCFA appears to perform well for cross-classified multilevel data even though a crossed factor was omitted in testing measurement invariance. On the other hand, for the non-invariant condition, even with a very large sample in the study, conventional MCFA (misspecified model) produced very low statistical power on rejecting the non-invariant model with the use of $\Delta\chi^2$ test and Δ goodness of fit indices, particularly, especially when ICC and difference in factor loading became smaller. That is, when testing measurement invariance without fully but only partially taking the cross-classified structure into account (i.e., treated the data as strictly hierarchical by ignoring one of the crossed factors and analyzed with conventional MCFA), the non-invariant model was far less likely rejected. The failure to detect the non-invariant factor loading difference resulted in concluding the non-invariant model as invariant between groups.

According to the findings from our simulation study, there are two potential sources leading to the low statistical power in multiple-group MCFA when misanalysing the cross-classified multilevel data: the underestimated ICC and the underestimated factor loading difference. ICC is computed by using the total variability (Ψ_T), the sum of factor variance components (Ψ_{Bj_1} at the between-level and Ψ_W at the within-level) as denominator and Ψ_{Bj_1} as numerator. Due to the redistribution of variance component mechanism (Luo & Kwok, 2009), the variance of the ignored crossed factor (Ψ_{Bj_2}) at the between-level redistributes to the lower level and results in substantial overestimation of the variance component at the lower level (Ψ_W, within-level) while the remaining between level factor variance is slightly underestimated. The combination of the substantially o*verestimated* (*inflated)* Ψ_W and the slightly *underestimated* Ψ_{Bj_1} results in an underestimated ICC in multiple-group MCFA. As shown by Kim and colleagues (2012a), ICC relates to the statistical power on detecting non-invariant factor loadings when testing factorial invariance in multilevel data, and lower ICC does typically link to lower statistical power even under large sample size conditions and a correctly specified MCFA model. Our current findings are consistent with what Kim and colleagues (2012a) have found. For the statistical power of detecting the non-invariant factor loading at the between-level in multiple-group MCFA, ICC plays a more important role than the number of cluster and the overall sample size.

In addition to ICC, the low statistical power may relate to the magnitude of non-invariance/difference in the target factor loadings between the two groups at the between-level (λ_{Bj_1}). Under the non-invariant condition, the target between-level factor loading in one of the groups (i.e., λ_{Bj_1} for G1) was set to 0.90 while the same factor loading for the other group (i.e., λ_{Bj_1} for G2) was set to be smaller: 0. 55, 0.65, or 0.75 to represent for large (0.35), medium (0.25), and small (0.15) difference, respectively. Nevertheless, given that no relative bias of λ_{Bj_1} for G1 was found, the positively biased λ_{Bj_1} for G2 led to narrowing the differences between the two groups, compared with the originally planned differences. Hence, given all other conditions were held at constant, the reduced difference between the non-invariant loadings would result in lower statistical power to detect such diminished effect and resulted in the failure in detecting the violation of metric invariance when the crossed factors were not fully taken into account in FI testing. In conclusion, the conventional multilevel CFA is not recommended for factorial invariance testing for cross-classified multilevel data given the considerably low empirical power.

Limitation and Directions for Future Research

Our current findings need to be interpreted given certain limitations. First, sample size was not considered in our simulation study. The reason of the use of our current sample size (i.e., 16000 observations nested within 8000 cells) was to warrant that all the replications could produce stable parameter estimates and the corresponding standard errors with good coverage. Although we did not examine the impact of the sample size, our findings showed that, even with a relatively large sample size, the statistical power for the detecting non-invariant model with misspecified MCFA is still low. Hence, in future study, larger sample size conditions may be considered even though these conditions may not be typical in educational and psychological research.

Second, for invariant model (all parameters are identical across groups), the results in the chi-square difference testing and the average differences (Δ) in all the goodness-of-fit indices were almost zero for all study conditions. For non-invariant model (between-level factor loading of one item only differ across groups while other parameters are identical), the study results found that (1) when invariant factor structure (G1 in the study) was ignored and not modeled, there was no impact on the factor loadings on the remaining factor structure in G1 (2) when non-invariant factor loading of one crossed factor (G2 in the study) was ignored and not modeled, the ignored non-invariant factor loading makes an impact on only the remained corresponding non-invariant factor loading, and the remaining factor loading of non-invariance was positively biased in G2. Regarding the impact of misspecification of invariance as well as the direction of the bias in the remaining factor loading of non-invariances, further investigation is required in conventional MCFA.

Third, in this study, the research scenario only focused on the invariance at the factor loading. In practice, non-invariance may exist at other parameters such as intercepts or both factor loadings and intercepts. In addition, non-invariance can occur at both within and between models simultaneously. Moreover, the structures of the within- and between-models may not always be identical. Overall, the performances of MCFA need to be studied under these more complex research settings with various sources of non-invariance.

Finally, up to date, there is no commercial statistical software program which can handle multiple-group analysis with cross-classified multilevel data within the SEM framework. Thus, we investigated the performance of a misspecified model (conventional multilevel CFA) as an alternative but not the correctly specified model (cross-classified multilevel CFA). Further development of software for such model (i.e., multiple group cross-classified SEM) or alternative option for analyzing this type of data is needed.

A completely cross-classified structure, in which units in a cluster of one crossed factor could affiliate with any clusters of the other crossed factor and vice versa (i.e., (i.e., no crossing/ overlap dependency between two crossed factors), was simulated in the study due to the current capacity of the software program (M*plus*) utilized in the present study. However, different levels of cross-classification such as partially cross-classified data structure would be possible in certain real situations. For example, in a situation where students are cross-classified by schools and neighborhoods, students living in certain neighborhoods only go to certain schools and students attending certain schools only live in certain neighborhoods. The performance of conventional multilevel CFA needs to be studied under these different structures of cross-classification (i.e., different degree of partial cross-classification) research settings.

CHAPTER III

STUDY 2: IMPACT OF NOT FULLY ADDRESSING CROSS-CLASSIFIED MULTILEVEL STRUCTURE IN MULTILEVEL MIXTURE MODELING: A MONTE CARLO STUDY

Mixture modeling is a relatively new exploratory analytical approach and has been gaining more attention recently in educational research. Mixture models, which are sometimes viewed as a more general form of the traditional cluster analysis, can be used for uncovering the latent groups/unobserved classes based on a specific model. Multilevel mixture modeling (MMM) is used for analyzing multilevel data under the mixture modeling framework (Asparouhov & Muthén, 2008). Under the traditional MMM, for identifying unknown heterogeneity across subpopulations, the relation between lower-level individuals and higher-level cluster units is assumed to be hierarchical or strictly nested in a cross-sectional study (Lüdtke, Marsh, Robitzsch, Trautwein, Asparouhov, & Muthén, 2008; Muthén & Asparouhov, 2009; Goldstein, 2003; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999; Van Horn et al., 2008). However, in reality, multilevel data may not always have a hierarchical (or strictly nested) structure.

Cross-classified multilevel modeling (e.g., cross-classified random effects model, CCREM) has been widely adopted for non-hierarchical multilevel data in social science research (Rashbash & Goldstein, 1994; Raudenbush, 1993). Compared to hierarchical linear models (HLMs) assuming the hierarchical structure of multilevel data, CCREM handles non-hierarchical multilevel data in which the relation between lower-level individuals and higher-level groups is not strictly nested or hierarchical. That is, HLM assumes that a lower-level observation belongs to one and only one higher-level cluster, whereas CCREM allows a lower observation to belong to multiple higher-level groups which are cross-classified simultaneously. For example, under the HLM situation, a student (e.g., level 1) belongs to only one cluster unit (e.g., level 2) such as either school or neighborhood. On the other hand, under the CCREM situation, students are nested within both the schools they attend and the neighborhoods where they live at the same time, that is, students (e.g., level 1) are cross-nested with schools (e.g., level 2) and neighborhoods (e.g., level 2) simultaneously (Raudenbush & Bryk, 2002). This type of non-hierarchical multilevel structured data is named *cross-classified* multilevel structured data. As an extension to HLM, CCREM enables one to take the multiple contextual effects into account in the analysis when researchers are interested in examining potential factors that influence individuals' outcomes (Meyers & Beretvas, 2006).

For analyzing multilevel data under the mixture modeling framework, when levels of multilevel data are not strictly nested or hierarchical, a conventional MMM using a current SEM statistical program is not an optimal approach to uncover the latent/unobserved heterogeneity of subgroups based on a specific multilevel model. It is because that up to date, the current statistical software for conventional MMM automatically assumes that multilevel data have a strictly nested or hierarchical structure, and there is no statistical program which can conduct mixture modeling and take into account the cross-classified (non-hierarchical) structured data simultaneously. In addition to the limitation of statistical programs, due to the lack of familiarity with this type of model or data structure, many substantive researchers adopt a less optimal approach to analyze this type of data. In other words, instead of taking the full cross-classified data structure into account for their analysis, researchers treat the data as strictly hierarchical by ignoring one of the crossed factors and use the traditional multilevel model for the analysis. For instance, in the above students/schools/neighborhoods example, researchers would ignore the school information and analyze the data as if students only nested within neighborhoods. Meyers and Beretvas (2006) and Luo and Kwok (2009) conducted methodological investigations to examine the impacts of misspecifying cross-classified multilevel data as strictly hierarchical multilevel data by ignoring one of the crossed factors in the regression analysis (HLM in these two studies). These previous studies showed that ignoring one of the crossed factors/levels (the school level in our example) can result in biased estimation of the random effect variances, which in turn, can lead to biased estimation of the standard errors of the fixed effects (or regression coefficients) in the missepcified models (HLMs in these studies). Ultimately, this may lead to erroneous statistical inferences. Therefore, a cross-classified structure should be modeled under the cross-classified MMM when researcher are interested in investigating multiple cluster (contextual) effects simultaneously such as school-level predictor(s) and neighborhood-level predictor(s) in conducting mixture modeling.

As an extension of the conventional MMM, the current study is the first to develop cross-classified MMM in the SEM framework. The primary purpose of the study is to investigate the performance of the conventional MMM in identifying optimal class enumeration (i.e., the number of latent classes) and in classifying individual group membership assignment (i.e., class identification) when the cross-classified multilevel structure is not fully considered and instead treated as strictly hierarchical multilevel data, ignoring a crossed factor in the analysis. The study focuses on the situation in which the unobserved subpopulation heterogeneity is captured by cluster (organization) level. A Monte Carlo study is conducted to achieve the study purpose. This study also examines the classification accuracy rate (i.e., statistical power rate) of conventional MMM (a misspecified model), the relative bias of parameter estimates, and factors that might impact the classification accuracy and the relative bias of parameter estimates. Below we first briefly review the conventional cross-classified MMM, followed by the research design and simulation study conditions.

Theoretical Framework

Multilevel Mixture Modeling (MMM)

Typically, conventional MMM allows researchers to classify the heterogeneity of subpopulation at the within (individual) level, between (organization) level, or both levels of units. In mixture modeling, the unobserved heterogeneity can be captured by categorical latent class variables, which represent the qualitatively different relationships across subpopulations. For the unobserved heterogeneity at different levels, the within-level variation can be expressed in terms of the variation among all subjects, whereas the between-level variation can be expressed in terms of the variation between clusters, that is, random intercept and(or) random slope (Muthén & Asparouhov, 2009). The model specification depends on where we assume and model unobserved heterogeneity.

In a MMM with a within-level latent class variable the categorical latent variable is a within-level variable. This means that the categorical latent classes are formed for individuals, and only the within-level variable, which is measured and modeled only at the within-level (e.g., individuals), is used to identify within-level categorical latent variables. For example, students can be classified into different subgroups within schools. In a MMM with a between-level latent class variable the categorical latent variable is a between-level variable, and only the cluster-level (e.g., organizations) is used to identify between-level categorical latent variables. Such models allow us to examine population heterogeneity that is caused by cluster level variables. In student/schools/ neighborhoods example, when heterogeneity in students' performance is caused by heterogeneity among students' schools, neighborhoods, or both schools and neighborhoods, the categorical latent class variables in the model should be a cluster level variable. A MMM with a within-between latent class variable incorporates both within-level and between-level latent class variables. For this model, the latent class variables can be measured and predicted by both within-level and between-level observed variables and the random effects can be measured and predicted by between-level variables. A between-level latent class variable can be considered as a special case of the within-between latent class variable.

Like the conventional MMM, each latent class can be either a within-level variable, a between-level variable, or a within-between level variable for cross-classified multilevel data. However, unlike the conventional MMM assuming the strict hierarchy, cross-classified multilevel data are non-hierarchical multilevel data in which multiple cross-classified cluster factors exist at the same level. Thus, the key purpose of using cross-classified MMM compared to conventional MMM is to account for these cross-classified factors in the analysis. Below we first briefly review a general formulation of the cross-classified MMM for heuristic purposes.

Cross-Classified Random Effect Modeling (CCREM)

For CCREM with a continuous outcome variable, adopting the notation of Rasbash and Browne (2001), the parenthesis was inserted around the pair of cross-classified factors that are represented by subscript *j*. $Y_{i(j_1 j_2)}$ is the observed score with individual-level covariate $X_{i(j_1 j_2)}$, neighborhood- level covariate W_{j_1}, and school-level covariate Z_{j_2} for individual *i* who belongs to neighborhood j_1 (cross-classified factor FB1) and attends school j_2 (cross-classified factor FB2) at the same-level. Two-level CCREM can be expressed in the following equation:

$$\text{Within-level:} \quad Y_{i(j_1 j_2)} = \beta_{0(j_1 j_2)} + \beta_1 X_{i(j_1 j_2)} + e_{i(j_1 j_2)}$$

$$\text{Between-level:} \quad \beta_{0(j_1 j_2)} = \gamma_{000} + \gamma_{010} W_{j_1} + \gamma_{020} Z_{j_2} + u_{0 j_1 0} + u_{00 j_2}$$

$$\beta_{1(j_1 j_2)} = \gamma_{100} \tag{15}$$

where $Y_{i(j_1j_2)}$ refers to the student (i) outcome that is cross-classified by both neighborhood j_1 and school j_2. The intercept coefficient $\beta_{0(j_1j_2)}$ was modeled to vary across clusters, indicating that its variability is explained by each of the W neighborhood-level and Z school-level predictors. The slope coefficient $\beta_{1(j_1j_2)}$ was modeled to be constant across both clusters (neighborhoods and schools). The neighborhood-level variable W_{j_1} varies by neighborhoods and the school-level variable Z_{j_2} varies by schools. In the fixed effect, γ_{000} refers to the overall intercept. γ_{100} refers to the slope parameter for within-level covariate $X_{i(j_1j_2)}$ and γ_{010} and γ_{020} refer to the slope parameters for neighborhood-level covariate W_{j_1} and school-level covariate Z_{j_2}, respectively. In the random effect, u_{0j_10} is the random effect associated with neighborhood , crossed factor FB1 ($u_{0j_10} \sim N$ (0, τ) at the between level, u_{00j_2} is the random effect associated with school, crossed factor of FB2 ($u_{00j_2} \sim N$ (0, ζ) at the between level, and $\varepsilon_{i(j_1j_2)}$ is residual variance ($\varepsilon_{i(j_1j_2)} \sim N$ (0, ψ) at the within level. Like HLM, random effects of cluster-level cross-classified factors and individual-level residuals are independent from each other (i.e., covariances between cluster-level random effects and within-level residuals are zero.) In addition, CCREM permits partitioning the remaining variability of $\beta_{0(j_1j_2)}$ into two components: a component between neighborhoods u_{0j_10} and a component between schools u_{00j_2}. The interaction $u_{0(j_1j_2)}$, between u_{0j_10} and u_{00j_2} can also be modeled, but it is typically not well estimated (Raudenbush & Bryk, 2002) and is not modeled here. It should be noted that with CCREM, the regression coefficient for neighborhood-level predictor γ_{010}

could also be modeled as randomly varying across schools and similarly the regression coefficient of school-level predictor γ_{020} could be modeled to vary across neighborhoods. In this study, however, these effects were each modeled as fixed across levels of the other cross-classified factor for the purpose of simplicity of demonstration.

Cross-Classified MMM with Between-level Latent Class Variable

As a simplest multilevel mixture model, suppose a two-level cross-classified regression mixture model with two latent classes. In general, multilevel analysis for binary outcome can be estimated using a logistic regression model (Henry & Muthén, 2010). Thus, for a binary outcome $C_{i(j_1j_2)}$, a logit link function can be applied in a two-level cross-classified logistic regression model with $P_{i(j_1j_2)}$ as the probability that $C_{i(j_1j_2)}$=1, and the log odds of $P_{i(j_1j_2)}$, logit $P_{i(j_1j_2)}$ as the natural log of $P_{i(j_1j_2)}$/ (1-$P_{i(j_1j_2)}$). The two-level cross-classified logistic random intercept regression model can be expressed as:

$$\text{logit } P_{i(j_1j_2)} = \beta_{0(j_1j_2)} + \beta_1 X_{i(j_1j_2)} + e_{i(j_1j_2)} \tag{16}$$

The corresponding between-level equations are

$$\begin{aligned} \beta_{0(j_1j_2)} &= \gamma_{000} + \gamma_{010} W_{j_1} + \gamma_{020} Z_{j_2} + u_{0j_10} + u_{00j_2} \\ \beta_{1(j_1j_2)} &= \gamma_{100} \end{aligned} \tag{17}$$

This implies that $P_{i(j_1 j_2)}$ can be expressed as the logistic function of a cross-classified random intercept model:

$$P(C_{i(j1j2)} = c \mid X_{i(j1j2)} W_{j1} Z_{j2}) = \frac{\exp(\gamma_{000} + \gamma_{100} X_{i(j1j2)} + \gamma_{010} W_{j1} + \gamma_{020} Z_{j2} + \mu_{0j10} + \mu_{00j2})}{\sum_C \exp(\gamma_{000} + \gamma_{100} X_{i(j1j2)} + \gamma_{010} W_{j1} + \gamma_{020} Z_{j2} + \mu_{0j10} + \mu_{00j2})} \quad (18)$$

where the random intercept is modeled to vary across clusters (neighborhood and school). For the residual variance ψ_c is often held class invariant for parsimony (Muthén & Asparouhov, 2009). Here, a single covariate for each level was used for simplicity of illustration, but further covariates can be added. At the between-level in Equation 3, the log odds of the outcome for a particular between-level unit j_1 and j_2 is defined as the population average of the log odds ($\gamma_{000} + \gamma_{010} W_{j_1} + \gamma_{020} Z_{j_2}$) plus the two random variations from group average for each group (u_{0j_10} and u_{00j_2}). These random variations are assumed to be normally distributed. The magnitude of u_{0j_10} and u_{00j_2} variance indicates the strength of the influence of the between-level units. That is, a larger variance indicates greater influence of the between-level units. For example, as shown in Equations from 1 to 4, the within-level predictor ($X_{i(j_1 j_2)}$) can be age, gender, or race of the log-odds of school dropout, and the cluster-level predictor(s) for neighborhood-level and school–level (W_{j_1} and Z_{j_2}, respectively) can be poverty rate of communities and the proportion of ethnic congruence of schools or participation in an intervention program for preventing students' dropout from school, respectively.

By applying this framework to cross-classified MMM, an observed variable $C_{i(j_1 j_2)}$ becomes a latent class variable $C_{i(j_1 j_2)}$ because it is inferred from the data. Here, we assess the log-odds of belonging to a reference group and allow the log-odds to vary across clusters. That is, the random effects, u_{0j_10} and u_{00j_2}, capture this variability in the log-odds. For example, the log-odds of being a student that drops out of school might differ depending on the level of the proportion of ethnic congruence within school a student attends and (or) on the level of poverty or unemployment of the neighborhood where a student belongs.

Then, because only between-level variables can be used as predictors for CB (Muthén &, Muthén, 2006), the corresponding between-level equations are written as functions of the between-level covariates W and Z with variation in coefficients across a between-level latent classes of CB,

$$CB_{i(j_1j_2)} = CB_{(j_1j_2)} \tag{19}$$

which represents the equality of CB within-cluster observations (*i*).

The corresponding between-level equations are written as functions of the between-level covariates W and Z with variation in coefficients across a between-level latent classes of CB,

$$\beta_{0(j_1j_2)|CB_{(j_1j_2)}} = \gamma_{000} + \gamma_{010}W_{j_1} + \gamma_{020}Z_{j_2} + u_{0j_10} + u_{00j_2} \tag{20}$$

Finally, the between-level latent class model can be expressed the probability of latent class membership for between-level latent class variable (CB) as the logistic regression with random intercept,

$$P(CB_{(j1j2)} \mid W_{j1}, Z_{j2}) = \frac{\exp(\gamma_{000} + \gamma_{010}W_{j1} + \gamma_{020}Z_{j2} + \mu_{0j10} + \mu_{00j2})}{\sum_{CB}\exp(\gamma_{000} + \gamma_{010}W_{j1} + \gamma_{020}Z_{j2} + \mu_{0j10} + \mu_{00j2})} \tag{21}$$

The model with two latent classes as a simplest version described here can be expanded to general situation where a model has any numbers of classes by expressing the probability of latent class membership for between level latent class variable as the multinomial logistic regression. The model can also be expanded to express the heterogeneity in the slopes for the covariates W and Z by allowing further random variation in these slopes across between-level clusters (i.e. u_{1j_10} and u_{10j_2} which were not included in the study).

Issues of Ignoring Cross-Classified Multilevel Structure

A conventional MMM can be used for uncovering the latent class or group membership based on a specific HLM. Given the bias caused by misspecifying CCREM to HLM in multilevel modeling framework (Luo & Kwok, 2009; Meyers & Beretvas, 2006), the potential bias would be expected in the conventional MMM when cross-classified structure is not fully considered and treated as hierarchical multilevel data. Especially, Luo and Kwok (2009) discussed that the distribution of variance components associated with the ignored crossed factor u_{00j_2} in the analysis resulted in biased estimation (either over- or underestimation) of the standard errors of fixed effects at a different level. For this reason, we focused on investigating the behaviors of the conventional MMM that assumes the strict hierarchy of multilevel data in identifying the correct number of latent classes and individual classification with cross-classified multilevel dataWe conducted Monte Carlo study, focusing on the MMM when the latent class variable was simulated in the organizational units (between-level). A set of design factors including the different levels of partial cross-classification (the number of FB1 and the number of FB2 to be cross-classified), cluster size (CS, the number of individuals per cluster), and mixing proportion of subpopulations were considered.

Method

Data Generation

The simulation study used two-level cross-classified multilevel data in which individuals at the within level were cross-classified by two crossed factors (e.g., schools or neighborhoods) at the between level with two known classes. Data with two known subpopulations under two-level cross-classified random effect model (CCREM) with random intercepts were first generated using SAS version 9.3 (SAS, 2011). For simplicity of the illustration, the regression slopes are fixed, which is commonly found in educational settings. In order to examine the effect of the variance component associated with the crossed factor 2 (FB2) which was omitted, we did not include the covariate of crossed factor FB2 in the analysis. The two-level CCREM with random intercept for data generation can be expressed as:

Within-level: $$Y_{i(j_1 j_2)} = \beta_{0(j_1 j_2)} + \beta_1 X_{i(j_1 j_2)} + e_{i(j_1 j_2)}$$

$$\text{Between-level:} \quad \beta_{0(j_1 j_2)} = \gamma_{000} + \gamma_{010} W_{j_1} + u_{0 j_1 0} + u_{00 j_2}$$

$$\beta_{1(j_1 j_2)} = \gamma_{100} \tag{22}$$

In this two-level CCREM, we simulated a two-level mixture model with between-level latent class variable. To create the condition where between-level variable contribute directly to between-level latent class formation and identification, γ_{010} differed across the two subpopulations whereas γ_{100} was equivalent for the two subpopulations. For the average regression coefficient, γ_{010} was set to .80 for positive effect group (group1) and - .80 for negative effect group (group2), at the between-level, so that they represented two "well-separated" classes following design by Nylund, Asparouhov, and Muthén (2007). γ_{100} was set to .50 for both subpopulations at the within-level. For the intra-class correlation (ICC), followed by Bell et al., (2009), three error variances were simulated to produce the target cross-classified ICC = .15 using the RANNOR random number generator in SAS version 9. 3 (SAS, 2011). The between-level intercept errors of $u_{0 j_1 0}$ (i.e., τ) and $u_{00 j_2,}$(i.e., ζ) were also generated from a normal distribution but with variance of .06 for neighborhoods and .12 for schools, respectively while the within-level errors $e_{i(j_1 j_2)}$ (i.e., ψ) were generated from a normal distribution with a variance of 1.00. The overall intercept γ_{000} was generated with the value of .10. For the population parameters, we referred to the previous simulation studies on both cross-classified multilevel regression and mixture modeling (Hox & Maas, 2001; Luo & Kwok, 2009; Maas & Hox, 2005; Meyers & Beretvas, 2006); students are cross-classified by schools and neighborhoods.

Simulation Conditions

Simulation conditions included the different levels of partial cross-classification (5, 25, or 50 for the number of feeders (neighborhoods in the example) × 2 or 8 for the number of receivers (schools in the example)), cluster size (10 or 20), and the mixing proportion of two groups (75%:25% or 50%:50%). The total 24 (3×2×2×2) conditions were included in the simulation.

Different degree of partial cross-classification

Suppose there are 50 neighborhoods (feeder, FB1) nested within 20 schools (receiver, FB2) and students are cross-classified by schools and neighborhoods as followed by Luo and Kwok's (2009) simulation study. In order to mimic real educational settings, we simulated the different levels of partial cross-classification by combining two conditions: (a) the number of feeder (neighborhoods) selected as cross-classified (100%, 50%, or 10%) and (b) the number of receivers (schools) assigned as cross-classified (10% or 40%). In a complete cross-classification condition, students from a specific school can live in any neighborhood and students from a specific neighborhood can go to any school. For example, all 50 neighborhoods (100%) are selected as cross-classified with all 20 schools (100%) in this study. In reality, however, a partially cross-classified data condition in which students living in certain neighborhoods only go to certain schools and students attending certain schools only live in certain neighborhoods is more likely to occur. For a different level of partial cross-classification, we created six different situations. For example, as the least partially cross-classified situation (i.e., 5 feeders and 2 receivers), we randomly selected 5 neighborhoods in the sample (i.e., 10% out of total 50 neighborhoods) and then randomly assigned half of the students from each of these 5 neighborhoods to the originally designated schools but the other half of the students to one of the 2 randomly selected non-designated schools (i.e., 10% out of 20 schools) in the sample. As the most partially cross-classified situation (i.e., 50 feeders and 8 receivers), we randomly selected 50 (100%) neighborhoods in the sample and then selected randomly 8 (40%) non-designated schools in the sample. In other words, we can create a more cross-classified data structure by increasing the selected number of feeder (neighborhoods, FB1) or receiver (schools, FB2).

Cluster size

Cluster size (CS or the number of individuals per cluster) varies at two levels (i.e., 10 and 20) which were suggested by Hox (1998) and also represented common cluster sizes found in multilevel research (Kim et al., 2012a). Thus, the total sample sizes were 500 (CS=10) and 1000 (CS=20), which covered a reasonable range of sample sizes commonly seen in educational and psychological studies.

Mixing proportion

The average regression models for the two subpopulations were specified as suggested by Nylund et al., (2007) so that the subpopulations represented two “well-separated” classes. The mixing proportions of the two subpopulations were set to be balanced or unbalanced (Chen, Kwok, Luo, & Willson, 2010). In the balanced situation, the mixing proportion was set to be 50% and 50% for the two subpopulations. In the unbalanced situation, the mixing proportion was set to be 25% for the positive effect group (group1) and 75% for the negative effect group (group 2).

Fitted Model

The hypothetical two-level HLM model with random intercept assuming hierarchical multilevel data structure (students nested within neighborhoods only; FB2) by omitting a crossed factor of schools (FB2) in conventional MMM (misspecified model) can be expressed as:

$$
\begin{aligned}
&\text{Within-level:} \quad && Y_{ij1} = \beta_{0j} + \beta_{1j}X_{ij1} + e_{ij1} \\
&\text{Between-level:} \quad && \beta_{0j1} = \gamma_{00} + \gamma_{01}W_{j1} + u_{0j1} \\
& && \beta_{1j1} = \gamma_{10}
\end{aligned}
\tag{23}
$$

From the comparison between Equation (22) for CCREM used for data generation and Equation (23) for HLM used for analysis, it is found that in Equation (23) for HLM the random effect u_{00j_2} of a crossed-factor ignored in the analysis was deleted from Equation (22).

For the two-level mixture model with two between-level latent classes, the categorical latent variable is a between-level variable. Again, only the between-level variable (W) measured and modeled only at the between level was used to identify between-level categorical latent classes. Figure 2 depicts the two-level mixture model with two between-level latent classes. In the within part of the model, the random intercept is shown in Figure 2 as a filled circle at the end of the arrow pointing to y. The filled circle refers to Y (or random intercept of y depicted as a circle) at the between level. In the between part of the model, the arrow from the between-level categorical latent class variable (CB) to y indicates that the intercept of Y varies across the classes of CB. In addition, the random intercept y and CB are regressed on a cluster-level covariate W. In this example, CB is a between-level variable with two latent classes. The multinomial logistic regression of CB on the cluster-level covariate was employed in the study. Maximum likelihood with robust standard errors using a numerical integration algorithm was used for the estimation of the fitted models.

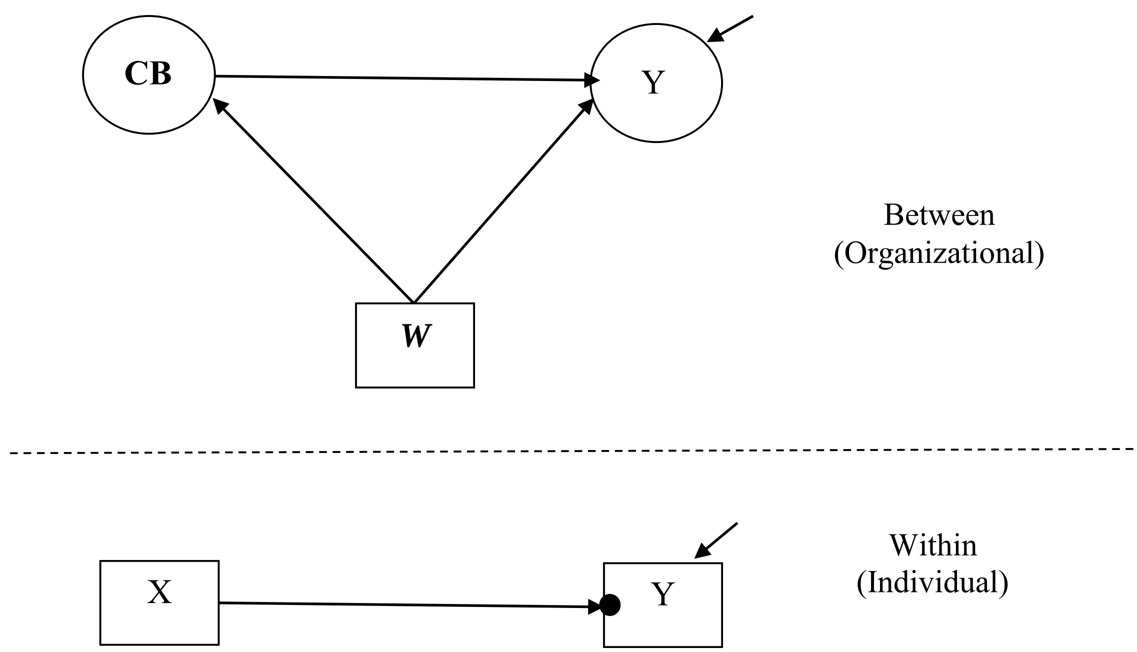

Figure 2. Fitted Two-level Mixture Model with Between-level Categorical Latent Class Variable (CB).

Data Analytic Procedure

To explore the performances of a conventional MMM, each data set was then fitted to a two-level MMM by ignoring a crossed factor using the M*plus* 7 with the Type=TWOLEVEL MIXTURE routine (Muthén & Muthén, 1998–2012). That is, the generated data sets with two-level cross-classified multilevel structure having two cross-classified factors were analyzed using the misspecified model (i.e., conventional two-level MMM). For each set of generated data, three different models with different numbers of classes (1, 2, and 3) were analyzed under the between-level latent class modeling in order to select the best solution model based on fit indices. The TYPE=TWOLEVEL routine decomposes the variance and covariance matrix into within- and between- models for the analysis. The BETWEEN option is also used to identify between-level categorical latent variables with TYPE=TWOLEVEL MIXTURE routine. First, we evaluated the empirical classification accuracy in two areas: 1) the proportion of the replications in which the correct number of class (i.e., two latent classes) was identified by the model selection index of conventional MMM (Class Enumeration Accuracy; CEA) and 2) the proportion of classifying correct group membership of individuals when the two-class model was selected as the best solution (Individual Classification Accuracy; ICA). For model evaluation, we used a fit index of sample size adjusted Bayesian information criterion (SABIC) which has been commonly used and recommended in many previous studies because it has been shown to have superior performance in simulation studies on mixture models (Nylund et al., 2007; Tofighi & Enders, 2008; Yang, 2006).

Second, the relative bias in the parameter estimates of the target regression coefficient at each level for two groups and residual (or error) variance at each level were also calculated and examined in the misspecified model. For between-level target regression coefficients $B\ (\widehat{\gamma_{01}})$, $B\ (\widehat{\gamma_{011}})$ and $B\ (\widehat{\gamma_{012}})$ refer to negative effect group (group1) and positive effect group (group2), respectively. For within-level regression coefficient, one $B\ (\widehat{\gamma_{10}})$ was evaluated. The relative bias in the target between-level error variance $B\ (\hat{\tau})$ and the within-level residual variance $B\ (\hat{\psi})$ and were calculated and evaluated in the misspecified model.

The relative bias of the estimates was computed using the following equation:

$$B(\hat{\beta}) = \frac{\hat{\beta} - \beta}{\beta} \tag{24}$$

where $\hat{\beta}$ is the group mean estimates of the target regression coefficient and error variance across the valid replications in the misspecified model, and β is the true parameter value of the corresponding parameters. Relative bias equal to zero indicates an unbiased estimate of the population parameter. A positive bias indicates an overestimation of regression coefficient and error variance, whereas negative relative bias indicates an underestimation of such parameters. To evaluate the estimated relative bias, we applied cutoffs of 0.05 for regression coefficients and of 0.10 for residual variance as an acceptable magnitude of relative bias recommended by Hoogland and Boomsma (1998).

Third, analysis of variance (ANOVA) was conducted to determine the impacts of the study conditions (i.e., CS, partial cross-classification, and mixing proportion). Eta-squared effect sizes (η^2 = *Sum of Square* $_{Effect}$ / *Sum of Square* $_{Total}$) were computed and reported as an indicator of practical significance of study conditions. Eta-squared of the design factors on two classification accuracy rates and the relative bias of parameter estimates were computed. This analysis was conducted only when the classification accuracy rates and the relative bias of parameter estimates exhibited considerable variability across simulation conditions.

Results

Admissible Solutions

It is known that a potential drawback of mixture model is that there is no guarantee of model convergence (Duncan, Duncan, & Strycker, 2013). Because of the different inadmissible solution rates (see Table 4) for each study condition, the number of replications used in the study fell between 743 and 869 so that we could have the same number of valid replications (i.e., 500) used for evaluating the results from each of three latent classes (i.e., 1, 2, and 3 latent class size) for all study conditions. Since all cases of the 1 class solution model were successfully analyzed with admissible solutions across conditions, the reports of inadmissible solutions were limited to the cases of the 2 latent class and 3 latent class solutions. Even though the cases with inadmissible solutions provided parameter estimates and standard errors, we did not include such cases in the analysis because parameter estimates and standard errors might not be within a plausible range (e.g., negative variance). As shown in Table 4, as CS increased, the percentage of inadmissible solution increased. For example, in the case of CS=10, from 33% to 38% of the replications yielded inadmissible estimates whereas from 36% to 42% did in case of CS=20.

Classification Accuracy Rates

The empirical classification accuracy rate of conventional MMM when a categorical latent variable is a between-level variable, and only the cluster-level (e.g., organizations) is used to identify between-level categorical latent variables are presented in Table 4. The summary of eta-squared (η^2) of the ANOVA results are presented in Table 5. When the cross-classified data structure was ignored and MMM was utilized instead of the cross-classified MMM approach, the class enumeration accuracy (CEA) rate of identifying the correct number of latent classes (i.e., two latent classes) by the model selection index (i.e., SABIC) fell between 75% and 85%. We then evaluated the individual's classification accuracy (ICA) rate if the two-class model was selected as the best solution. The accuracy rate of correct individual group membership assignment ranged from 79% to 86%.

For the classification accuracy rate of identifying the correct number of classes (CEA), ANOVA results showed that CS (η^2 =25%) had a substantial impact on the CEA while cross-classification (feeders, receivers, and interaction between feeders and receivers, η^2 =5%, 1% and 11%, respectively) and mixing proportion (η^2 = 4%) had a very small impact on the CEA. It is shown that as CS increased, the CEA decreased.

For the classification accuracy of individuals in each data set when the two-class model was selected as a best solution (ICA), ANOVA results showed that the mixing

Table 4. Classification Accuracy Rates of the Conventional Multilevel Mixture Modeling (Misspecified Models) with Between-level Latent Variable

Mixing	Cross-classification		CEA		ICA		IADS Rate	
(Proportion)	Feeder	Receiver	CS=10	CS=20	CS=10	CS=20	CS=10	CS=20
Unbalanced (25%:75%)	5	2	0.80	0.76	0.84	0.86	0.36	0.40
		8	0.76	0.74	0.85	0.85	0.37	0.41
	25	2	0.81	0.79	0.85	0.86	0.33	0.39
		8	0.77	0.80	0.85	0.86	0.36	0.42
	50	2	0.77	0.80	0.85	0.86	0.38	0.37
		8	0.81	0.82	0.86	0.86	0.37	0.36
Balanced (50%:50%)	5	2	0.85	0.75	0.79	0.81	0.35	0.40
		8	0.83	0.77	0.79	0.80	0.37	0.41
	25	2	0.84	0.79	0.79	0.81	0.33	0.42
		8	0.80	0.75	0.80	0.81	0.34	0.40
	50	2	0.80	0.79	0.80	0.82	0.36	0.42
		8	0.85	0.76	0.81	0.83	0.35	0.41

Note: CEA refers to Class Enumeration Accuracy rate which is the proportion of identifying the correct model (two classes) as a best solution. ICA refers to Individual Classification Accuracy which is the proportion of accuracy in individual classification when the two-class model was selected as a best solution. IADS Rate is the proportion of inadmissible solution.

Table 5. Summary of Eta-squared (η2) in Analysis of Variance (ANOVA) Results of the Conventional Multilevel Mixture Modeling (Misspecified Models) with Between-level Latent Variable

Outcomes	CS	MixProp.	Cross-classification		
			feeder	receiver	feeder × receiver
Classification Accuracy					
Class Enumeration	25%	4%	5%	1%	11%
Individual Group Assignment	6%	87%	4%	0%	1%
IADS Rate	64%	0%	2%	1%	4%
Relative Bias					
Regression Coefficient					
$B(\widehat{\gamma_{10}})$	1%	1%	15%	1%	12%
$B(\widehat{\gamma_{011}})$	5%	34%	7%	1%	11%
$B(\widehat{\gamma_{012}})$	30%	16%	13%	1%	5%
Residual Variance					
$B(\hat{\psi})$	0%	0%	80%	11%	7%
$B(\hat{\tau})$	8%	2%	67%	13%	8%

Note. Class Enumeration is the proportion of accuracy in identifying the correct model (two classes) as a best solution. Individual Group Assignment is the proportion of accuracy in individual classification when the two-class model was selected as a best solution. IADS Rate is the proportion of inadmissible solution. MixProp is mixing proportion condition. CS is cluster size. For Regression Coefficient, γ_{10} refers to regression coefficient at the within-level-; γ_{011} and γ_{012} refer to regression coefficient at the between-level for group1 and group 2, respectively. For Residual Variance, ψ and γ refer to residual variance at the within-level and at the between-level, respectively.

proportion (η^2 =87%) had a substantial impact on the ICA. On the other hand, CS (η^2 =6%) had a small impact on the ICA. As the mixing proportion changed from unbalanced to balanced condition, the ICA became lower. As CS became larger, the ICA increased.

In terms of the impact of study factors, the inconsistent patterns over two classification accuracy rates (CEA and ICA) were observed. For instance, mixing proportion had a trivial impact on the CEA whereas mixing proportion had the largest impact on the ICA among the study factors in the study. For CS, the CEA decreased as CS became larger whereas the ICA increased as CS became larger.

Relative Bias of Parameter Estimates

The summary of relative bias of the estimates of the target residual (or error) variance and the regression coefficient at each level is presented in Table 6. First, the relative bias in the target residual variance at between-level $B\,(\hat{\tau})$ and within-level $B\,(\hat{\psi})$ was evaluated. Irrespective of study conditions, the large positive $B\,(\hat{\tau})$ was found and ranged from 0.549 to 1.391. ANOVA results indicated that the partial cross-classification factors (feeders, receivers, and interaction between feeders and receivers, η^2 =67%, 13% and 8%, respectively) had substantial impacts on $B\,(\hat{\tau})$ while CS (η^2 =8%) and mixing proportion of subgroups (η^2 =2%) had small impacts on $B\,(\hat{\tau})$. As the data structure became more partially cross-classified, the positive $B\,(\hat{\tau})$ became smaller. As CS increased and mixing proportion changed from unbalanced to balanced, the positive $B\,(\hat{\tau})$ became larger. $B\,(\hat{\psi})$ was acceptable except under the conditions in which the data structure was the most partially cross-classified (i.e., 50 feeders and 8 receivers)

Table 6. Relative Bias of Parameter Estimates of the Conventional Multilevel Mixture Modeling (Misspecified Models) with Between-level Latent Variable

		Cross-classification		Regression Coefficients			Residual Variance	
CS	MixProp.	Feeder	Receiver	γ_{10}	γ_{011}	γ_{012}	ψ	τ
10	25:75	5	2	0.000	-0.069	0.013	0.006	1.174
		5	8	0.007	-0.030	0.018	-0.001	1.239
		25	2	-0.006	0.006	0.019	0.014	1.066
		25	8	-0.001	-0.001	0.019	0.028	0.885
		50	2	-0.011	-0.038	0.000	0.029	0.851
		50	8	0.009	-0.034	0.018	0.050	0.549
	50:50	5	2	0.002	0.011	0.023	0.004	1.363
		5	8	-0.008	0.013	0.027	0.004	1.294
		25	2	-0.001	0.027	0.016	0.013	1.103
		25	8	0.003	0.014	0.003	0.030	0.982
		50	2	-0.010	0.019	0.014	0.029	1.019
		50	8	-0.003	0.033	0.013	0.054	0.624
20	25:75	5	2	0.002	-0.065	0.023	0.006	1.378
		5	8	0.002	0.007	0.024	0.004	1.391
		25	2	0.001	0.001	0.020	0.009	1.223
		25	8	-0.001	-0.033	0.016	0.026	1.016
		50	2	-0.004	-0.010	0.012	0.032	1.035
		50	8	-0.004	-0.034	0.018	0.046	0.730
	50:50	5	2	0.006	-0.010	0.030	0.002	1.383
		5	8	-0.001	-0.006	0.035	0.007	1.334
		25	2	0.002	-0.023	0.041	0.015	1.275
		25	8	-0.005	-0.012	0.040	0.023	1.115
		50	2	0.000	0.006	0.026	0.031	1.119
		50	8	-0.004	-0.006	0.025	0.052	0.782

Note: For MixProp. 25:75 refers to unbalanced mixing proportion condition between two groups; 50:50 refers to balanced mixing proportion condition between groups. CS is cluster size. For Regression Coefficients, γ_{10} refers to regression coefficient at the within-level-; γ_{011} and γ_{012} refer to regression coefficient at the between-level for group1 and group 2, respectively. For Residual Variance, ψ and τ refer to residual variance at the within-level and at the between-level, respectively.

in the current study. Similar to the findings in $B\ (\hat{\psi})$, ANOVA results showed that only the partial cross-classification factors (feeders, receivers, and interaction between feeders and receivers, η^2 =80%, 11% and 7%, respectively) had substantial impacts on $B\ (\hat{\psi})$. Furthermore, the large proportion of FB2 variance component (ζ) was added (or redistributed) to the variance component of the remaining factor of FB1 (τ) whereas a very small proportion of ignored FB2 variance component (ζ) was added (or redistributed) to the variance component at the within-level (ψ).

Second, the relative bias in the parameter estimates of the target regression coefficients ($B\ (\widehat{\gamma_{10}})$ for within-level and , $B\ (\widehat{\gamma_{011}})$ and $B\ (\widehat{\gamma_{012}})$ for between-level, respectively) was examined for group 1 and group 2, respectively. Irrespective of simulation conditions, the estimates of γ_{10} were very close to the population true parameter, which yielded negligible $B(\widehat{\gamma_{10}})$ (less than .01 in most conditions). Given that the simulated two group parameters (γ_{011} and γ_{012}) were 0.8 and -0.8 for group 1 and 2, respectively, the estimates, on average, fell between 0.745 and 0.826 and between -0.833 and -0.800 for group1 and group 2, respectively. $B\ (\widehat{\gamma_{011}})$ was from -0.069 to 0.034 for group 1 and $B\ (\widehat{\gamma_{012}})$ was from 0.000 to .041 for group 2. All relative biases were within ± 0.05 or close to zero except when the smallest cross-classification condition (5 feeders and 2 receivers) was combined with unbalanced mixing proportion for two subpopulations. In such conditions, $B\ (\widehat{\gamma_{012}})$ was about -.069 and -.065 for CS=10 and CS=20, respectively. Although most of the $B\ (\widehat{\gamma_{01}})$s in this study were acceptable, the between-level regression coefficient ($\widehat{\gamma_{01}}$) showed generally larger bias than the within-level regression coefficient ($\widehat{\gamma_{10}}$). ANOVA results showed that none of the study conditions had a substantial effect on the relative bias in the parameter estimates of regression coefficients at both the within- and the between-level.

Discussion

Multilevel mixture modeling (MMM) is used for analyzing multilevel data, which is assumed to be hierarchical or strictly nested, under the mixture modeling framework (Asparouhov & Muthén, 2008). Cross-classified multilevel modeling (e.g., cross-classified random effects model, CCREM) has been widely adopted for non-hierarchical multilevel data in social science research (Rashbash & Goldstein, 1994; Raudenbush, 1993). Ideally, cross-classified multilevel structure should be taken into account when conducting MMM. However, up to date, there is no statistical program which can conduct MMM and take into account of the cross-classified data structure simultaneously. For this reason, researchers often treat cross-classified data as a conventional multilevel data (i.e., as strictly nested or hierarchical) by ignoring one of the crossed factors. Hence, it is important to examine the potential impact of ignoring the cross-classified data structure in conducting MMM.

As shown in the simulation results, regardless of study conditions we found the *overestimatio*n in error variances at both within- and between-levels, which was redistributed from the ignored variances of crossed factor in conventional MMM with cross-classified multilevel data. Due to the redistribution of ζ ignored in the analysis to τ and ψ, when the remaining crossed factor (FB1) is almost nested within the ignored crossed factor (FB2), the large proportion of FB2 variance component (ζ) was added to the variance component of the remaining factor of FB1 (τ) whereas a very small proportion of ignored FB2 variance component (ζ) was added to the variance component at the within-level (ψ).Our current findings are consistent with what Luo and Kwok (2009) have found.

The overestimation of between-level error variances is plausibly related to the accuracy of the regression coefficients (fixed effects) in the conventional MMM with between-level latent class variable, which in turn, may result in less accurate classification regarding optimal class solution as well as the individual class membership/classification when only part of the cross-classified data structure was considered. In other words, the correct number of latent class was less likely to be identified and even though the correct number of classes was selected, individuals were likely to be classified in the wrong group. The findings might imply that when the overestimation becomes much larger the classification accuracy would get lower in conducting MMM without fully but only partially considering the cross-classified data structure.

The classification accuracy may relate to the cluster size. Irrespective of study conditions, the large positive $B\,(\hat{\tau})$ was found and became more positively biased as CS increased. We reasoned that this can be explained by the design effect in a cluster sampling (Kalton, 1983). Design effect is formally the ratio of the actual variance of a statistic to the variance of the statistic computed under the assumption of simple random sampling, presuming the same sample size. For a cluster sampling with m observations in each cluster and intra-class correlation (ICC) of ρ, the design effect, D_{effect}, can be expressed as

$$D_{\text{effect}} = 1 + (m-1)\,\rho \tag{25}$$

where the magnitude of the design effect is a function of the number of observations in each cluster (CS or *m*) and ICC. Given the fixed ICC (i.e., 0.15) across all study conditions in our study, the larger CS, the larger the design effect, and which in turn, resulted in the more biased parameter estimates when the homogeneity within cluster was not taken into account (Meyers & Beretvas, 2006).

Furthermore, the magnitude of overestimation in residual variance at within- and between- levels is related to the level of partial cross-classification. As the data structure became more partially cross-classified (the increasing number of selected feeders and receivers to be selected as cross-classified), the positive bias (overestimation) in between-level error variance became smaller. Thus, with more partial cross-classification the higher accuracy of regression coefficients at the between level is expected because the error variance at the between level becomes smaller (i.e., standard error (SE) will be smaller).

On the other hand, the mixing proportion did not have much impact on the classification accuracy rate in identifying the correct number of the latent classes whereas the mixing proportion had substantial impact on the classification accuracy rate in correctly classifying individual's class assignment: the classification accuracy rate became lower as the mixing proportion changed from unbalanced to balanced condition. The result is consistent with the previous studies from Chen et al., (2010) in that the accuracy of individual's classification became larger when a larger discrepancy between classes on the mixing proportion existed. The mechanism of the unbalanced mixing proportion on the increasing accuracy of the individual classification is, however, not fully explained and future investigation is still needed.

In summary, results from the study using the design factors (i.e., cluster size, degree of cross-classification, and mixing proportion of subgroups) for simulation indicate that model misspecification resulted in an positive bias (i.e., overestimation) in the variance at the remaining level (between) added from the ignored factor at the same between-level. According to the findings from our simulation study, the variance associated with the remaining factor (e.g., neighborhood in the study) was consistently overestimated regardless of study conditions. In applied settings, researchers might find the more spuriously inflated variance component associated with neighborhood crossed factor when the only neighborhood cluster factor is taken into account (while omitting school cluster factor) in the model than when cross-classified structure between schools and neighborhoods is considered simultaneously. Up to date, there is no statistical SEM (or latent variable modeling) program which can conduct multilevel mixture modeling and take into account of the cross-classified data structure concurrently. In other words, the capacity of current SEM statistical software for handling the complexity of cross-classified MMM limits the use of an optimal analytic approach for cross-classified multilevel data. For example, although many researchers understand the importance of conducting an optimal analysis for cross-classified multilevel structured data, researchers should treat the cross-classified multilevel data structure as hierarchical multilevel data by ignoring crossed factor(s) in conducting conventional MMM. Hence, a researcher should acknowledge this limitation and be cautioned when conventional MMM is utilized with cross-classified multilevel data given spuriously inflated variance component associated with remaining crossed factor.

Limitations and Directions for Future Research

Study findings need to be interpreted given certain limitations. First, this study only provides a preliminary investigation of the impact of various factors on parameter estimates and statistical inferences that result from inappropriate modeling in cross-classified multilevel data. In the study, the cross-classification condition was generated and analyzed such that 50 neighbors (feeder, FB1) are nested within 20 schools (receiver, FB2) while students are cross-classified by schools and neighborhoods, and FB2 was ignored in the misspecified model. On the other hand, the research scenario where FB2 is nested with FB1, and FB2 is ignored in the analysis is also possible. The previous simulation study conducted by Luo and Kwok (2009) tested the two situations; 1) where the remaining crossed factor nested within the ignored crossed factor (simulated situation in this study) and 2) the ignored crossed factor nested within the remaining crossed factor. Our interest is of between-level and to investigate the impact of ignored crossed factor on the variance component of ignored crossed factor on the remaining level. Given the small bias on the remaining factor FB1 at the between-level found from the similar previous studies from the latter condition, we simulated only the former condition where the proportion of the variance component at the remaining factor redistributed from the ignored crossed factor was much larger than the proportion of variance component at the lower level.

Second, the ICC values were generated to be the same for all study conditions (followed by Bell et al., 2009) in the current study given the weak empirical evidence supporting ICC as a main factor related to the behaviors of cross-classified multilevel modeling. For example, there is no previous study regarding how to compute ICC in cross-classified random effect modeling (CCREM), especially under the partial cross-classification condition. In Meyer and Beretvas (2006) study, two variance components of two crossed factors (FB1 and FB2 in our study) were summed up as a part of the total variability (i.e., denominator in the computation of ICC). It is appropriate when no correlation between two crossed factors (i.e., complete cross-classification situation) is assumed under complete cross-classification situation (i.e., units in a cluster of one crossed factor could affiliate with any clusters of the other crossed factor and vice versa). ICC is related to power in conventional multilevel modeling (Hox & Maas, 2001). Further studies should investigate the impacts of ICC in CCREM.

Third, the research scenario in the study only focused on a between-level latent class modeling and provides a preliminary investigation of impacts of various factors on estimates and inferences that result from inappropriate modeling of cross-classified multilevel data in multilevel mixture modeling. The current study focused on modeling the between-level latent class where the between-level data contribute directly to the class formation and identification. Asparouhov and Muthén (2008) discussed the issue of relative small sample size on the between level (i.e., number of clusters). In many practical research situations, between-level sample size of 100 clusters or less is common in multilevel data. Even though simple Expectation Maximization (EM) estimation approach (Muthén & Shedden, 1999) can be used when the class variable is on the between level, we need further investigation about issues of biased parameter estimation due to limited sample size. Hence, for more accurate parameter estimates, it is recommended to construct the between-level latent class modeling with within-level observed variable(s) that contributes directly to the class formation and identification and sufficient number of parameters that differ across different subgroups (Asparouhov & Muthén, 2008). The further exploration of between-level latent class modeling including within-level data for unbiased classification is needed. Furthermore, to investigate impacts of not fully addressing cross-classified multilevel structure in various multilevel mixture modeling situations, we can extend the study to within-level latent class modeling (where latent class can be measured and predicted only by within-level observed variables) and within-between latent class modeling (where latent class can be measured and predicted by within level observed variables, while the random effects can be measured and predicted by between observed variables).

Finally, for model selection in the context of latent class mixture modeling, many studies were conducted to examine the performance of fit statistics (Celeux & Soromenho, 1996; Henson, Reise, & Kim, 2007; Nylund et al., 2007; Tofighi & Enders, 2008). These studies revealed that SABIC among information based criteria and Lo-Mendell-Rubin adjusted likelihood ratio test (LMR) and the bootstrap likelihood ratio test (BLRT) among nested model likelihood ratio tests are commonly used and recommended (Tofighi & Enders, 2008; Vermunt & Magidson, 2002; Henson et al., 2007) for model selection. In the study, we used only SABIC due to intensive computation procedure (i.e., requiring long computation time) of nested model likelihood ratio tests, especially BLRT. Through preliminary investigation of the performance of fit statistics in identifying the best solution, we observed that (1) SABIC performed much better than nested model likelihood ratio tests and (2) the performances of LMR and BLRT were found to be similar in terms of the magnitude of classification accuracy rate on most of study conditions. Although we chose SABIC for selecting a best model selection in the study, a combination of criteria has been recommended to guide applied researchers in selecting the optimal number of classes. One available recommendation is to use SABIC to narrow the solutions to a few plausible models first and then request BLRT for these models to select the best model. Nevertheless, further investigation is still worthy of searching for an optimal selection as a correct model with cross-classified multilevel data.

CHAPTER IV

CONCLUSIONS

In study 1, the results of the study suggest that the inappropriate modeling of cross-classified multilevel data is problematic particularly if researchers are interested in testing invariance to investigate the differences of a particular scale between groups at the organizational level. When partially taking the cross-classified multilevel data structure into account and using the conventional multilevel CFA in testing factorial invariance, low empirical power in detecting non-invariance was observed. In other words, the non-invariant models (with non-invariant loadings at the between- or organizational-level) are likely considered to be as invariant models and lead to an incorrect conclusion of factorial invariance (specifically, metric invariance in this study) when one of crossed factors was ignored in the analysis. Thus, with the non-hierarchical multilevel data in which the relation between lower level individuals and higher level clusters is not strictly nested or hierarchical in cross-sectional study conditions, an optimal approach which takes the non-hierarchical multilevel data structure into account in the analysis is required. This simulation study demonstrated possible incorrect statistical inferences by using the conventional multilevel CFA in the available SEM program (i.e., M*plus*) which assumes multilevel data as strictly nested or hierarchical structure, for testing FI with cross-classified multilevel data (non-hierarchical structure). Due to the limitations/restrictions of the available structural equation modeling software for the multiple group multilevel modeling analysis, researchers should be aware of the potential impact of not fully addressing cross-classified multilevel structure in testing measurement invariance.

In study 2, when evaluating latent class effect at the between-level, according to the findings from our simulation study, we found the overestimated variance component at the remaining level (between-level in the study) which was redistributed from the variance component from the ignored crossed factor in conducting conventional MMM without taking fully but only partially the cross-classified data structure into account. It implies that researchers might find the more spuriously inflated variance component associated with remaining crossed factor when the only part of crossed cluster factors is considered and instead treated as hierarchical multilevel data, compared to the variance component when the cross-classified structure is fully considered. Furthermore, to date, there is no statistical SEM program which can conduct multilevel mixture modeling and take into account of the cross-classified data structure concurrently. Hence, a researcher should acknowledge this limitation and be cautioned when conventional MMM is utilized with cross-classified multilevel data in conducing mixture modeling with between-level latent class variables given an expected inflated variance component associated with the remaining crossed factor.

REFERENCES

Asparouhov, T., & Muthen, B. (2008). Multilevel mixture models. In G. R. Hancock and K. M. Samuelson (Eds.), *Advances in latent variable mixture models*, (pp. 7-51). Charlotte, NC: Information Age Publishing.

Bell, B.A., Owens, C.M., Kromrey, J.D., & Ferron, J.M. (2009). Parsimony vs. complexity: A Monte Carlo investigation of hierarchical and cross-classified modeling Using SAS® PROC MIXED. Proceedings of the Annual SAS Global Forum.

Berkhof, J., & Kampen, J. K. (2004). Asymptotic effect of misspecification in the random part of the multilevel model. *Journal of Educational and Behavioral Statistics, 29,* 201–218.

Borsboom, D. (2006). When does measurement invariance matter? Commentary. [Editorial Material]. *Medical Care, 44*(11), S176-S181.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.).New York, NY: Springer-Verlag.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification, 13*, 195-212.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464-504. doi:http://dx.doi.org/10.1080/10705510701301834

Chen, Q., Kwok, O., Luo, W., & Willson, V. L. (2010). The impact of ignoring a level of nesting structure in multilevel growth mixture models: A monte carlo study. *Structural Equation Modeling, 17*(4), 570-589. doi:http://dx.doi.org/10.1080/10705511.2010.510046

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255. doi:http://dx.doi.org/10.1207/S15328007SEM0902_5

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research, 38*(4), 529-568. doi: 10.1207/s15327906mbr3804_5

DiStefano, C., & Hess, B. (2005). Using confirmatory factor analysis for construct validation: An empirical review. *Journal of Psychoeducational Assessment, 23*, 225-241.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2013). *An introduction to latent variable growth curve modeling: Concepts, issues, and application*. Routledge Academic.

Fan, X., & Sivo, S. A. (2009). Using ΔGoodness-of-Fit indexes in assessing mean structure invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(1), 54-69.

French, B. F., & Finch, W. H. (2008). Multiple group confirmatory factor analysis: Locating the invariant referent sets. *Structural Equation Modeling, 15,* 96–113.

Goldstein, H. (1986).Multilevelmixed linear model analysis using iterative generalized least squares. *Biometrika, 73*, 43–56.

Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.

Goldstein, H. (2003) *Multilevel Statistical Models*, 3rd edn. London: Arnold.

Heck, R. H., & Thomas, S. L. (2009). *An introduction to multilevel modeling techniques* (2nd ed.). New York, NY: Routledge.

Henry, K. L., & Muthén, B. (2010). Multilevel latent class analysis: An application of adolescent smoking typologies with individual and contextual predictors. *Structural Equation Modeling, 17*(2), 193-215. doi:http://dx.doi.org/10.1080/10705511003659342

Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relativemodel fit statistics. *Structural Equation Modeling, 14*(2), 202-226.

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., & Maas, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation*

Modeling: A Multidisciplinary Journal, 8(2), 157-174. doi:http://dx.doi.org/10.1207/S15328007SEM0802_1

Jones-Farmer, L. A. (2010). The effect of among-group dependence on the invariance likelihood ratio test. *Structural Equation Modeling: A Multidisciplinary Journal, 17*(3), 464 - 480. doi: 10.1080/10705511.2010.489020

Kalton, G. (1983). *Introduction to survey sampling.* Newbury Park, CA: Sage.

Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions* (2nd ed.), Thousands Oaks, CA: Sage.

Kaplan, D., Kim, J.-S., & Kim, S.-Y. (2009). Multilevel latent variable modeling: Current research and recent developments. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The sage handbook of quantitative methods in psychology* (pp. 592-612). Thousand Oaks, CA: Sage Publications.

Kim, E. S., Kwok, O., & Yoon, M. (2012a). Testing factorial invariance in multilevel data: A monte carlo study. *Structural Equation Modeling, 19*(2), 250-267. doi:http://dx.doi.org/10.1080/10705511.2012.659623

Kim, E. S., Yoon, M., & Lee, T. (2012b). Testing measurement invariance using MIMIC: Likelihood ratio test with a critical value adjustment. *Educational and Psychological Measurement, 72*(3), 469-492. doi: 10.1177/0013164411427395

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods, 13*(3), 203-229. doi:http://dx.doi.org/10.1037/a0012869

Luo, W., & Kwok, O. (2009). The impacts of ignoring a crossed factor in analyzing cross-classified data. *Multivariate Behavioral Research, 44,* 182-212.

Maas, C. J., & Hox, J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*(3), 86-92. doi: 10.1027/1614-1881.1.3.86

Meade, A. W., & Lautenschlager, G. J. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods, 7*(4), 361–388.

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods, 10*(3), 259-284. doi: 10.1037/1082-989x.10.3.259

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543.

Meredith, W., & Millsap, R. E. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika, 57*(2), 289-311.

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medic al Care, 44*(11), S69-S77.

Meyers, J., & Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research, 41*, 473–497.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research, 22*(3), 376-398.

Muthén, B., & Asparouhov, T. (2009). Multilevel regression mixture analysis. *Journal of the Royal Statistical Society Series A-Statistics in Society, 172,* 639-657

Muthén, B. O., & Muthén, L. K. (1998-2012). *Mplus User's Guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55,* 463-469.

Nylund, K. L., Asparauhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14,* 535–569.

Rasbash, J., & Browne, W. J. (2001). Modelling non-hierarchical structures. In A. H. Leyland & H.Goldstein (Eds.), *Multilevel modelling of health statistics* (pp. 93–105). Chichester, UK: Wiley.

Rasbash, J., & Goldstein, H. (1994). Efficient analysis of mixed hierarchical and crossed random structures using a multilevel model. *Journal of Educational and Behavioral Statistics, 19*, 337–350.

Raudenbush, S. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational Statistics, 18,* 321–349.

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

SAS Institute. (2011). *SAS/STAT 9. 3 User's Guide: Mixed Modeling (Book Excerpt)*. SAS Publishing (Ed.). SAS Institute.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent* *variable analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.

Selig, J. P., Card, N. A., & Little, T. D. (2008). Latent variable structural equation modeling in cross-cultural research: Multigroup and multilevel approaches. In F. J. R. van de Vijver, D. A. van Hemert & Y. H. Poortinga (Eds.), *Multilevel analysis of individuals and cultures*. New York: Lawrence Erlbaum Associates.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage Publications.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology, 91*(6), 1292-1306.

Tofighi, D., & Enders, C. K. (2008*). Identifying the correct number of classes in a growth mixture model.* In G. R.

Van Horn, M. L., Fagan, A. A., Jaki, T., Brown, E. C., Hawkins, J. D., Arthur, M. W., ...& Catalano, R. F. (2008). Using multilevel mixtures to evaluate intervention effects in group randomized trials. *Multivariate Behavioral Research*, *43*(2), 289-326.

Vermunt, J.K., & Magidson, J. (2002). *Latent class cluster analysis.* In J. Hagenaars & A.McCutcheon (Eds.), Applied latent class analysis (pp. 89-106). New York: Cambridge University Press.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J.

Bryant, M. Windle & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Yang, C. (2006). Evaluating latent class analyses in qualitative phenotype identification. *Computational Statistics & Data Analysis, 50,* 1090–1104.

Yoon, M. (2008). Statistical power in testing factorial invariance with ordinal measures. *Dissertation Abstracts International, 68*(11), 7705B. (UMI No. 3288042)

Yoon, M., & Millsap, R. E. (2007). Detecting violations of factorial invariance using data-based specification searches: A Monte Carlo study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 435-463.

Zyphur, M. J., Kaplan, S. A., & Christian, M. S. (2008). Assumptions of cross-level measurement and structural invariance in the analysis of multilevel data: Problems and solutions. *Group Dynamics-Theory Research and Practice, 12*(2), 127-140. doi: 10.1037/1089-2699.12.2.127