November 2017

The performance of Multilevel Structural Equation Modeling (MSEM) in comparison to Multilevel Modeling (MLM) in multilevel mediation analysis with non-normal data

Thanh Vinh Pham
University of South Florida, tvpham2@mail.usf.edu

The performance of Multilevel Structural Equation Modeling (MSEM) in comparison to Multilevel Modeling (MLM) in multilevel mediation analysis with non-normal data

by

Thanh Vinh Pham

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Educational Measurement and Research
College of Education
University of South Florida

Major Professor: Eun Sook Kim, Ph.D.
John M. Ferron, Ph.D.
Yi-hsin Chen, Ph.D.
Tony Tan, Ed.D.

Date of Approval:
October 16, 2017

Keywords: Hierarchical, multilevel, mediation, structural equation, non-normal

ACKNOWLEDGMENTS

I am so grateful for all the support from my wonderful doctoral committee. First of all, I would like to express me deep appreciation to my mentor, Dr. Jeffrey Kromrey. He had provided advice and support through most of my doctoral program years until his retirement. His broad knowledge and patient guidance had not only helped me to complete my dissertation but also inspired me to delve into statistics and SAS programming as well as extend my research interest.

In the same manner, I would like to express my heartfelt thanks to Dr. Eun Sook Kim, my major advisor. Her insightful comments and strong support were very helpful for me to overcome all the challenges during this journey. I am grateful that I got a lot of positive influence from her passion and dedication to research and education.

I would like to thank Dr. John Ferron so much for the constructive comments and suggestions for my dissertation as well as my other studies during my years in the program. His expertise and very detailed explanation have cleared many obstructs on my way of doing research. I also would like to thank Dr. Yi-Hsin Chen and Dr. Tony Tan for their time, encouragement, and valuable comments on my dissertation.

It was my fortunate to be a part of the Educational Measurement and Evaluation program in the University of South Florida. The program provides me with a truly friendly, supportive and scholar learning environment. I would like to thank Dr. Robert Dedrick and Dr. Liliana Rodriguez-Campos for their caring, understanding and unconditional support.

This dissertation would not be possible without the support from my family. Thanks to my parents for the encouragement and support that they have provided throughout my life, for their sacrifice over years so that I can complete my study here. Last but not least, my special gratitude goes to my wonderful wife, Diep Nguyen, for her true belief in me, for her tireless support and her endurance with me over years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The mediation analysis has been used to test if the effect of one variable on another variable is mediated by the third variable. The mediation analysis answers a question of how a predictor influences an outcome variable. Such information helps to gain understanding of mechanism underlying the variation of the outcome. When the mediation analysis is conducted on hierarchical data, the structure of data needs to be taken into account. Krull and MacKinnon (1999) recommended using Multilevel Modeling (MLM) with nested data and showed that the MLM approach has more power and flexibility over the standard Ordinary Least Squares (OLS) approach in multilevel data. However the MLM mediation model still has some limitations such as incapability of analyzing outcome variables measured at the upper level. Preacher, Zyphur, and Zhang (2010) proposed that the Multilevel Structural Equation Modeling (MSEM) will overcome the limitation of MLM approach in multilevel mediation analysis. The purpose of this study was to examine the performance of the MSEM approach on non-normal hierarchical data. This study also aimed to compare the MSEM method with the MLM method proposed by MacKinnon (2008) and Zhang, Zyphur, and Preacher (2009). The study focused on the null hypothesis testing which were presented by Type I error, statistical power, and convergence rate. Using Monte Carlo method, this study systematically investigates the effect of several factors on the performance of the MSEM and MLM methods. Designed factors considered were: the magnitude of the population indirect effect, the population distribution shape, sample size at level 1 and level 2, and the intra-class correlation (ICC) level. The results of this study showed no significant effect of the degree of non-normality on any performance criteria of either MSEM

or MLM models. While the Type I error rates of the MLM model reached the expected alpha level as the group number was 300 or higher, the MSEM model showed very conservative performance in term of controlling for the Type I error with the rejection rates of null conditions were zero or closed to zero across all conditions. It was evident that the MLM model outperformed the MSEM model in term of power for most simulated conditions. Among the simulation factors examined in this dissertation, the mediation effect size emerged as the most important one since it is highly associated with each of the considered performance criteria. This study also supported the finding of previous studies (Preacher, Zhang, & Zyphur, 2011; Zhang, 2005) about the relationship between sample size, especially the number of group, and the performance of either the MLM or MSEM models. The accuracy and precision of the MLM and MSEM methods were also investigated partially in this study in term of relative bias and confidence interval (CI) width. The MSEM model outperformed the MLM model in term of relative bias while the MLM model had better CI width than the MSEM model. Sample size, effect size, and ICC value were the factors that significantly associate with the performance of these methods in term of relative bias and CI width.

CHAPTER ONE: INTRODUCTION TO THE STUDY

The idea of mediation and the indirect effect have a long and important history in social science research (Mackinnon, Warsi, & Dwyer, 1995). In 1986, Baron and Kenny discussed the distinction between mediation and moderation and introduced a definition of mediation. Their definition later was cited commonly among mediation studies. Baron and Kenny (1986) emphasized mediation as a causal process and defined a mediator as a variable that accounts for the relation between independent and dependent variables. A basic mediation model includes three variables where the independent variable has an effect on the mediator variable, which then affects the outcome variable. An example of the mediation effect is the study of Loukas, Suzuki, and Horton (2006). In this study, Loukas et al. examined whether the effect of positive perceived school climate on reducing emotional and behavioral problems was mediated by students' sense of connection to their school (school connectedness). In this example, students' perception might have an impact on their connection to school, which then affects their emotions and behaviors in school. Instead of analyzing the direct relationship between independent variable and dependent variable, a mediation analysis aims to determine whether the effect of an independent variable on the dependent variable is mediated by one or more mediator variables. In other words, a mediation analysis answers the question of how a predictor affects an outcome rather than whether it does so.

Mediation analysis provides evidence on how a predictor influences an outcome variable indirectly via a mediating variable. Such information increases understanding of the mechanism underlying variation of the outcome. Mediation analysis as an examination of a causal process also allows researchers to build and test a more general and complex theory. The research results supported by a complex and general theory are easier to apply to other settings or populations (Judd & Kenny, 1981; MacKinnon & Dwyer, 1993).

Mediation models have been shown to be useful tools in social science research. Since its publication in 1986, Baron and Kenny's work has been cited 20,380 times according to Psycnet as of June 2014. The application of mediation analysis has been seen in various areas from social psychology, psychology, and public health, in both experimental and non-experimental research (Krull & MacKinnon, 2001; MacKinnon, 2008). Mediation models, in the field of educational research, are typically used to describe effects of a context factor (predictor) on the outcome. For example, Quiroga, Janosz, Bisset, and Morin (2013) used a mediation model to determine how students' depression affected school dropout rate. Their model showed that the effect of depression on school dropout was mediated by students' self-perception of academic competence. Matsumura, Garnier, and Spybrook (2013) applied a multilevel mediation model to test the mediation effect of the Content-Focused Coaching (CFC) program on student reading achievement through the quality of classroom text discussions. The findings identified the effect of the CFC program and also supported the hypothesis of mediation effect.

The mediation model introduced by Baron and Kenny (1986) is designed for single level data (i.e., data with all observations in the same level without clustering by groups or repeated measures). As a causal process, a mediation effect in single level models is tested by applying the ordinary least squares (OLS) technique (Baron & Kenny, 1986). The OLS technique requires an important assumption of independence. This assumption, in simple term, means that all observations are independent. However, the independence assumption is not always satisfied in social science research because of the association with group setting. In the field of education, there are many studies with nested data where students are nested in their class or schools. Nested data are also frequently encountered in social and behavioral sciences research. For example, researchers may be interested in data in which people are grouped by their neighborhoods, patients are nested in hospitals or clinics, or workers are clustered in teams or companies (Krull & MacKinnon, 2001). Individuals within a group are likely to share the same group characteristics, to get the same within-group interaction, and to receive the same effects from a group environment. Therefore, they tend to be more similar to each other than they are to the individuals in different groups. The dependency of individuals on group membership is measured by the intra-class correlation (ICC). A large value of ICC index shows a violation of the independence assumption. In such situations, single mediation models using standard OLS techniques will result in inefficient estimates and biased standard errors (Krull & MacKinnon, 2001, 1999; Raudenbush & Bryk, 2002).

Nested data can be referred to as multilevel data or clustered data because the data may be collected at multiple levels. For example, with students nested in their classroom, the data collection might be at the student level (referred to as level-1 data) or at the classroom level (referred to as level-2 data). In order to avoid violating the assumption of independence in mediation analysis with multilevel data, the clustering needs to be taken into account when conducting a mediation analysis (Preacher et al., 2011). Multilevel Modeling (MLM), also called Hierarchical Linear Modeling (HLM), mixed modeling, or random-coefficient modeling is a statistical approach based on regression but defines different error terms for different data levels. This approach allows for the analyses of multilevel data with more accurate Type I error control than traditional regression methods (Raudenbush & Bryk, 2002).

Applying the MLM in mediation analysis with multilevel data was first suggested by Krull and MacKinnon (1999). The advantages of the MLM approach over the traditional OLS approach on multilevel mediation analysis were demonstrated in Krull and MacKinnon (1999, 2001). Krull and MacKinnon (2001) also defined several different multilevel mediation models depending on the level of variables in the model (i.e., independent variable, mediation variable, and dependent variable). For example, a mediation model with all three variables measured at level 1 is labeled as 1-1-1. A model is labeled as 2-1-1 when the independent variable is measured at level 2 and the mediator and dependent variable are measured at level 1.

Researchers have extended Krull and MacKinnon's (2001) work by applying MLM in different multilevel mediation models, with fixed or random coefficients (Bauer, Preacher, & Gil, 2006; Kenny, Korchmaros, & Bolger, 2003; MacKinnon, 2008; Pituch, Murphy, & Tate, 2009; Pituch & Stapleton, 2008; Pituch, Stapleton, & Kang, 2006; Pituch, Whittaker, & Stapleton, 2005). Researchers also investigated the impact of centering method on the performance of the MLM mediation test (Tofighi, 2010; Zhang et al., 2009). The findings of these studies showed that group-mean centering produces better outcomes than grand mean centering or non-centering. The MLM mediation test with group mean centering is called centering within context (CWC) by Tofighi (2010) and Zhang et al. (2009) or un-conflated multilevel modeling (UMM) by Preacher et al. (2011).

The MLM approach has been shown to have more power and flexibility over the single level mediation method in multilevel data. However, MLM still has some limitations in assessing mediation effects in multilevel data. According to Preacher et al. (2010), the MLM method is not capable of analyzing outcome variables measured at the upper level. This limitation also was mentioned by Krull and MacKinnon (2001) when they indicated that the dependent variables are only at level 1 and each effect in the causal chain involves a variable affecting another variable at the same or lower level. For example, MLM is not applicable with a multilevel mediation model that has the predictor and the mediator measured at level 1 while the outcome variable is measured at level 2. Another limitation is that MLM conflates the within and between components of effects which results in incorrect estimates of indirect effects or bias toward the within effect (Preacher et al., 2010).

The limitations of the MLM approach in multilevel mediation analysis can be overcome by the Multilevel Structural Equation Modeling (MSEM) approach (Preacher et al., 2010). The MSEM approach was first introduced by Muthén and Asparouhov (2008). As an advancement of structural equation modeling (SEM), MSEM applies the SEM technique to multilevel data. Ludtke, Marsh, Robitzsch, Trautwein, Asparouhov, & Muthén (2008) showed that the MLM approach might have substantially biased estimates of between-effects and underestimate the associated standard error while the MSEM approach can provide unbiased estimates of high level constructs under some conditions. Regarding multilevel mediation analysis, Preacher et al. (2010) showed that the MSEM method will provide better estimates of indirect effects and also allows analyzing more multilevel mediation models than can be done in the MLM approach such as mediation models with the outcome variable at level 2 (Preacher et al., 2010). His simulation study also found that MSEM performs better than MLM in terms of confidence interval (CI) coverage, efficiency and convergence (Preacher et al., 2011).

Even though the mediation models have been extended from a simple single-level model in Baron and Kenny's paper to the MLM and MSEM frameworks, these approaches all require an assumption of normality. However, non-normal data are frequently encountered in applied social science research (Micceri, 1989). Despite the extensive attention of researchers on applying MLM and SEM approaches in multilevel analyses, few studies have addressed the impacts of non-normal data on the model estimation in both MLM and MSEM approaches (Zhang, 2005). Maas and Hox (2004) have found that standard errors are affected by the violation of distribution assumption in multilevel modeling. Byrd (2008) examined the effect of non-normality on MSEM analyses and found that standard errors of parameter estimates are distorted at both level 1 and level 2. Zhang (2005) compared the performance of MLM and MSEM approaches in the presence of non-normal data. He found that both SEM and MLM approaches have the same statistical power and are quite robust to the violation of the normality assumption but the SEM approach seems to be more useful in complex path models.

Other researchers examined the impact of violation of the normality assumption on the results of a mediation analysis. Most of their research has focused on the single level mediation analysis (Biesanz, Falk, & Savalei, 2010; Finch, West, & MacKinnon, 1997; Zu, 2009). The only one that investigated the multilevel modeling context is Pituch and Stapleton's paper (2008). This study examined the performance of some MLM methods for an upper level mediation model or 2-1-1 model as labeled by Krull and MacKinnon's (2001) method. Pituch and Stapleton found that the robust standard error and bootstrapping methods perform better than the standard methods. Regarding the MSEM approach, even though its advantage was pointed out by Preacher et al. (2010), its performance has not been fully examined, especially under the condition of non-normal data.

The Purpose of this Dissertation

The purpose of this dissertation was to examine the performance of the MSEM approach on non-normal hierarchical data. This study also aimed to compare the MSEM method with the MLM method proposed by MacKinnon (2008) and Zhang et al. (2009). The study focused on the null hypothesis testing which were presented by Type I error, statistical power, and convergence rate.

This dissertation focused on the mediation model called upper level model with fixed slopes and continuous variables. The upper level mediation model named by Pituch and Stapleton (2008) has the independent variable measured at level 2, the mediator, and the dependent variable, measured at level 1. It is a common model in applied literature (Preacher et al., 2011) and can be labeled as the 2-1-1 model using the notation convention in Krull and MacKinnon (2001).

Research Questions

Based on the purposes, this study aims to answer the following research questions:

1. What are the differences between the MSEM approach and the MLM approach in terms of Type I error control, statistical power, and convergence rate with non-normal data?
2. How do the simulation design factors, including the magnitude of the population indirect effect, the population distribution shape, sample size at level 1 and level 2, and the ICC level, affect the performance of the MSEM approach and the MLM approach?

Significance of Study

The multilevel mediation analysis using Multilevel Structural Equation Modeling (MSEM) was recently proposed by Preacher et al. (2010). In spite of potential advantages of the MSEM approach over the MLM approach, little research has investigated the MSEM performance, especially when the normality assumption is violated. On the other hand, the performance of the MLM approach on multilevel mediation analysis has been examined by many researchers with normally distributed data. However, the effect of violating statistical assumptions, such as non-normal data in the MLM approach has not been well studied.

Therefore, it was informative to provide empirical evidence of the MSEM approach's performance on multilevel mediation with non-normal data. This study also presented a comparison between MSEM and MLM.

Definition of Terms

The section provides the definition of some statistical terms that are often used in this dissertation.

- **Mediator:** a variable that transfers whole or partial effect of an independent variable on a dependent variable (MacKinnon & Dwyer, 1993).
- **Mediation analysis:** a statistical analysis used to investigate how variables (mediators) mediate the effect of predictors on outcome variables (MacKinnon, 2008).
- **Multilevel data:** also called hierarchical data. It refers to data that contain some hierarchical or nested structure. Multilevel data are collected from more than one level of research units (MacKinnon, 2008).
- **Multilevel mediation model:** a mediation model applied to multilevel modeling. In this dissertation, a multilevel mediation model refers to a simple mediation model with multilevel modeling (MacKinnon, 2008).
- **Multilevel Modeling:** a statistical approach that allows analyzing multilevel data without restructuring them. This approach partitions the residual variance into residual components at different data levels. It allows the different effects of independent variable as well as intercepts across different groups (Hox, 2002).
- **MLM model with fixed effect:** A MLM model in which the effect of variables do not vary across higher level unit (Hox, 2013).

- **MLM model with random effect:** A MLM model in which the effect of variables is assumed to vary across higher level units (Hox, 2013).
- **Simple mediation model:** a path model with three variables: a predictor (X), an outcome variable (Y), and a mediator (M). In this model, *a* is the coefficient for X in a model predicting M from X, and *b* and *c'* are the coefficients in a model predicting Y from both M and X, respectively. In the language of path analysis, *c'* is the direct effect of X, whereas the product of *a* and *b* is the indirect effect of X on Y through M. If all three variables are observed, then $c = c' + ab$ (Hayes, 2009).
- **Structural Equation Modeling:** a statistical modeling approach widely used in social sciences to study the relationship among unobserved (or latent) variables and observed variables. It hypothesizes that the population covariance matrix of observed variables is a function of model parameters. The SEM procedure aims to minimize the difference between hypothesized and observed covariance matrices (Bollen, 1989).
- **Un-conflated multilevel modeling (UMM):** a multilevel mediation modeling approach with the group-mean centering procedure. It allows for the separation of between and within variances of mediators (Zhang et al., 2009).
- **Upper level mediation model:** A multilevel mediation model with the predictor measured at level 2 while the mediator and the outcome variable measured at level 1 (Pituch et al., 2005). This mediation model can be labeled as the 2-1-1 model using the notation convention in Krull and MacKinnon (2001).

CHAPTER TWO: LITERATURE REVIEW

This chapter provides information about studies related to mediation analysis. For that purpose, it has the following sections: an introduction to mediation analysis and a general overview of mediation models in single level data, a review of mediation analysis using multilevel modeling (MLM), a review of mediation analysis using Multilevel Structural Equation Modeling (MSEM), and a review of studies conducted with non-normal data in multilevel mediation analysis. These sections give an overview of research problems and go through findings of previous studies that influence the design of this study.

Mediation Modeling in Single Level Data

Mediation analysis can be used to examine how a variable affects another variable. Mediation analysis assumes a casual effect among variables where a mediation variable, or mediator, transmits a partial or whole effect of an independent variable onto a dependent variable. If we consider a model with two variables, independent X and dependent Y as shown in Figure 1, the effect of X on Y, represented by c is called X's total effect on Y. This total effect is interpreted as the expected change in Y when X is changed by one unit. However, the effect of X on Y might come directly or indirectly.

The simplest mediation model is presented in Figure 2. In this model, variable X has a direct effect on Y, denoted as c'. The variable X is also hypothesized to affect the mediator, M, which then has an effect on variable Y. The effect of variable X on variable Y through variable M is called the indirect effect or mediated effect.

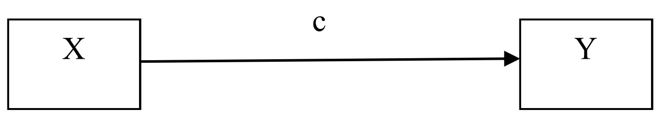

*Figure 1.*A causal relationship between variable X and Y

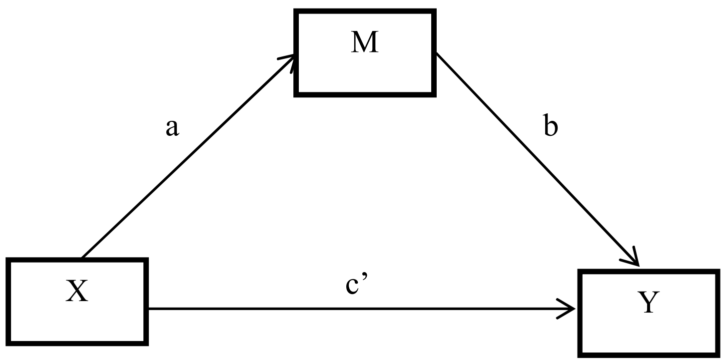

*Figure 2.*Simple mediation model

The simple mediation model can be assessed by the following three equations (Baron & Kenny, 1986; MacKinnon, 2008).

$$Y = d_1 + c\,X + \varepsilon_1 \quad (1)$$

$$M = d_2 + a\,X + \varepsilon_2 \quad (2)$$

$$Y = d_3 + c'X + b\,M + \varepsilon_3 \quad (3)$$

Coefficient c denotes the total effect of X on Y. Coefficient a represents the effect of X on M while b and c' are coefficients for the model predicting Y by both X and M. By definition, the indirect effect of X on Y is the product of a and b. The ε_1 is a residual term of regression equation (1). It is assumed to be normally distributed. The terms ε_2 and ε_3 denote the residuals for equation (2) and (3) with the assumption of a bivariate normal distribution. Terms d_1, d_2, and d_3 represent intercepts for the three equations (1), (2), (3), respectively. If all three variables in the models are observed, $c = c' + ab$ (MacKinnon, 1995; Hayes, 2009). Simple algebra shows that $ab = c\text{-}c'$. In other words, the indirect effect is the difference between the total effect and the direct effect.

The estimated indirect effect, or mediation effect, is shown by $\hat{a}\hat{b}$ where "$\frown$" sign represents the estimator of each respective coefficient. Another way to measure the indirect effect is the difference $\hat{c} - \widehat{c'}$.

Many approaches have been proposed to test the mediation effect in single level data. The most common approach is the causal steps approach developed by Judd and Kenny (1981) and Baron and Kenny (1986). This approach requires estimates of all three path coefficients of the mediation model (a, b, and c') as well as the total effect c. The mediation effect is supported if coefficients *a, b*, and *c* are statistically significant. If *c'* is statistically significant, we have a partially mediated effect. If c' is insignificant, the effect of X on Y is completely mediated by mediator M.

The second approach, called the difference of coefficients approach, uses the difference between the total effect (*c*) and the direct effect (*c'*). The test statistic is calculated by the ratio of $(\hat{c} - \widehat{c'})$ to its standard deviation. A *t*-test is conducted to test the null hypothesis $H_0: c - c' = 0$. The *t* statistic can be calculated by several methods which differ in the formula for the standard deviation of (*c-c'*) (Clogg, Petkova, & Shihadeh, 1992; Freedman & Schatzkin, 1992).

Another approach, called the product of coefficients approach, uses the estimate of ab. The z statistic as the ratio of $\hat{a}\hat{b}$ and its standard error is used to test the null hypothesis $H_0: ab = 0$. Several methods can be used to calculate the standard error of $\hat{a}\hat{b}$. The most popular one was proposed by Sobel (1982, 1986). The Sobel test is based on the assumption that the product of $\hat{a}\hat{b}$ has the normal distribution. However, the product of two random normal variables is not normally distributed (Stone &Sobel, 1990). As a result, the critical value of the z-test is incorrect. Bootstrapping and the empirical-M test are alternative methods to fix this problem.

In the empirical-M test, the distribution of *ab* is determined as the distribution of the product of two normally distributed variables. The distribution function of the product of two normal variables was presented by Meeker, Cornwell, and Aroian (1981). A table of critical values of the distribution of the product was then calculated by Mackinnon, Lockwood, and Williams (2004) using a series of simulations. The test using this table to test the indirect effect is called empirical-M test.

Bootstrapping is an empirical procedure that approximates the distribution of a statistic by repeatedly resampling an observed sample. As a result, researchers can make statistical inferences without any information about the distribution. Given that the product of *a*b is not normally distributed, bootstrapping is a useful method to estimate and test the indirect effect.

Some simulations have been conducted to compare the performance of these tests. MacKinnon, Lockwood, Hoffman, West, and Sheets (2002) compared 14 different methods of testing the mediation effect and found that the causal steps approach has the lowest power in detecting indirect effects while Type I error rates below the nominal value in all sample sizes. The power of the difference in coefficients method is higher but Type I error rate still remains low. The product of coefficients method yielded the highest statistical power. In comparison to the methods based on the normal distribution, the Bootstrapping and empirical *M*-test methods provide the most accurate confidence limits and greater statistical power (MacKinnon et al., 2004).

Mediation Models in MLM

Multilevel data are common in social sciences, especially in educational research where contextual factors often have an important role. For example, a researcher interested in students' behaviors might need to consider the effect of teachers or classrooms. The teachers and classrooms data will be collected accompanied by students' data. However, if students are not randomly selected but grouped by their teachers, this would be an example of multilevel data, or nested data. In this example, students are nested in their teachers or classrooms. Information in nested data is collected from different levels. Data at level one would include students' information which varies across students. Data at level two would be obtained from teachers' or classrooms' information which is identical for all students in the same group. A problem with multilevel data is that individual observations are in general not independent. Students who belong to the same classroom or who share the same teacher have more interactions, and share a more similar environment. As a result, their characteristics might be more similar to each other than those of students from different classes. The effect of group membership is measured by the intra-class correlation coefficient or ICC. The ICC index is calculated by the ratio of between group variance to total variance in the following equation. The total variance is the sum of between group variance and within group variance.

$$ICC = \tau_{00}/(\tau_{00} + \delta^2)$$

where τ_{00} is the between group variance and δ^2 is the within group variance. As we can see from the formula, as τ_{00} increases relative to δ^2, the ICC increases.

Standard statistical tests depend heavily on the assumption of independence. The violation of the independence assumption might result in large biases in the standard errors (Hox, Maas, & Brinkhuis, 2010). Multilevel data need to be restructured to apply single level analysis techniques. Traditionally, researchers can choose between two restructuring methods: aggregation or disaggregation. However, both of these methods have their own weaknesses. In the aggregation approach, new variables are created at the upper level by combining information in each group of low level variables. The analysis is conducted at the upper level where each group is considered as an observation. This method discards a lot of information in the lower level by analyzing data in the upper level. It also reduces significantly the power because of limiting the sample size to the number of groups. The disaggregation approach, conversely, will disaggregate variables from the upper level to the lower level. New variables at the lower level have the same value for all individuals in the same group. Individuals across groups are treated as if they are randomly selected from the same population. However, observations in the same group are likely more similar to each other than they are to the members from other groups. By considering clustered data as independent data, the disaggregation approach violates the assumption of independence. Moreover, information from the upper level is treated as independent information at a much larger sample size of lower level data. The increase of the sample size results in the deflation of standard error.

Using restructuring approaches can also lead to the conceptual fallacies when interpreting the analysis results. These fallacies refer to the incorrect assumption that the relationship between variables at the higher level is the same at the lower level. The first fallacy happens when the conclusion regarding variability across unit at higher level is based on data of units at the lower level. This fallacy is call atomistic fallacy and is the result of the aggregation approach. For example, the reading score might be found to associated with the number of books students have at home but it is not necessarily true that the number of books at the school library has an significant impact on the average reading score of students in that school. The second fallacy is called ecological fallacy which simply applies the relationship among the higher level variables to the lower level unit. An assumption that a student who comes from a class with the higher experience teacher will have a higher score is an example of the ecological fallacy.

Multilevel modeling (MLM) or hierarchical linear modeling (HLM) is an alternative that allows analysis of multilevel data without restructuring them. The MLM method partitions the residual variance into residual components at different data levels. For example, two-level data that group student outcomes within teachers would include the student and teacher levels. The MLM model of these data has two residual components, between-teacher (a variance of the teacher level residual) and within teacher (the variance of the student level residual). The teacher residuals, also called teacher effects represent the effect of teacher characteristics on student outcomes. Students from the same group (the same teacher) have the same teacher effect. Another advantage of the MLM method is to allow the intercept to vary from group to group and the effects of independent variables to be different across groups. In addition, the MLM method allows analysis of cross-classified data which is impossible by using traditional disaggregation or aggregation approaches. The cross-classified data is a special case of hierarchical data when a subject is classified by two or more clusters these clusters are not hierarchical or nested within each other. An example of cross-classified data is students from the same neighborhood go to different schools. In this case, level-one student information is nested in both variables neighborhood and school which are in the same level.

The MLM approach, as an extension of linear regression, requires some similar assumptions as the OLS approach (Maas & Hox, 2004). MLM assumes a linear relation between dependent and independent variables. It also requires homoscedasticity and independence of observations. The normality assumption in MLM is slightly different from that in OLS because of the nature of multilevel data. MLM requires level-1 residuals being normally distributed and level-2 random effects having a multivariate normal distribution.

Krull and MacKinnon (1999, 2001) suggested using MLM in mediation analysis with multilevel data. Each equation in the context of the single level mediation model is reformulated as a multilevel model. The conversion of the single equation into the multilevel model depends on the level where variables are measured. As a result, there are different multilevel mediation designs when variables are measured at different levels. A notation with a set of three numbers I-J-K is used to label a simple multilevel mediation model with three variables (independent variable X, dependent variable Y, and mediating variable M) where I is the level of X, J is the level of M and K is the level of Y (Krull & MacKinnon, 2001).

In this dissertation, I only focus on the multilevel data with an independent variable measured at level 2, a mediator and a dependent variable measured at level 1. This multilevel mediation mode is also called an upper level mediation model (Pituch et al., 2005) and labeled as the 2-1-1 model by Krull and MacKinnon (1999). If *c* denotes the total effect of X_j on Y_{ij}, *c'* denotes the direct effect of X_j on Y_{ij}, *a* denotes the effect of X_j on M_{ij}, *b* denotes the effect of M_{ij} on Y_{ij}, the 2-1-1 mediation model would be presented in figure 3. In this figure, the independent variable (X_j) is measured at level 2; and the mediator (M_{ij}) and outcome variable (Y_{ij}) are measured at level 1. The subscripts *i* and *j* denote the individual-level and cluster-level units, respectively.

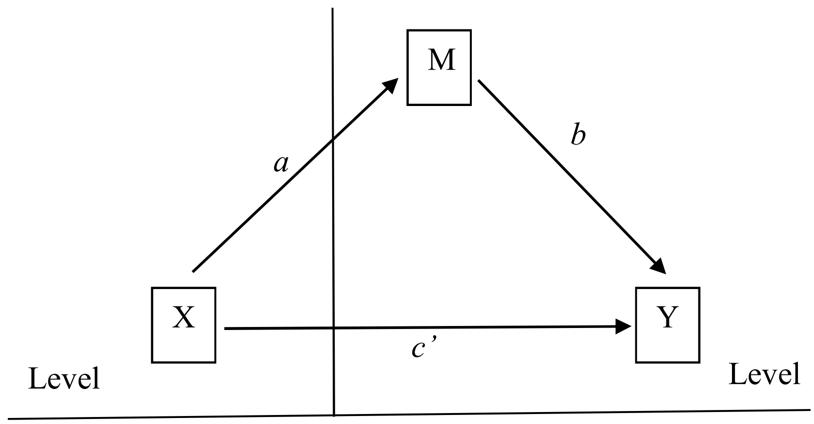

Figure 3. Multilevel mediation model 2-1-1

The effects shown in Figure 3 can be estimated by a set of three multilevel models. The first one estimates c, the effect of X_j on Y_{ij} without a mediator M_{ij}.

The level 1 equation: $Y_{ij} = \beta_{Y0j} + r_{Yij}$ (4)

The level 2 equation: $\beta_{Y0j} = \gamma_{Y00} + cX_j + u_{Y0j}$ (5).

where Y_{ij} is the outcome of observation i in group j, β_{y0j} is the group intercept, and r_{yij} is the error term associated with Y_{ij}. Independent variable X_j is included as a predictor for the group intercept β_{y0j} while u_{Y0j} is the level 2 residual.

The second multilevel model represents the relationship between the level-2 variable X_j and the individual-level mediator M_{ij}.

The level 1 equation: $M_{ij} = \beta_{M0j} + r_{Mij}$ (6)

The level 2 equation: $\beta_{M0j} = \gamma_{M00} + aX_j + u_{M0j}$ (7)

where β_{M0j} is the intercept for group j; r_{Mij} is the level-1 residual, u_{M0j} is the level-2 residual; and a is the effect of X_j on M_{ij}.

The third model shows the effect of variable X and mediator M on the outcome Y with the assumption of the fixed effect of M on Y across groups.

The level 1 equation: $Y_{ij} = \beta_{X0j} + bM_{ij} + r_{Xij}$ (8)

The level 2 equation: $\beta_{X0j} = \gamma_{X00} + c'X_j + u_{X0j}$ (9)

where β_{X0j} is the intercept for group j; r_{Xij} and u_{X0j} are the Level-1 and Level-2 residuals, respectively.

Based on this set of models, the formulas for the indirect effect are still the same as the formulas for the single level mediation model. The indirect effect by the difference of direct effect of X on M and direct effect of M on Y is $c - c'$. The indirect effect by the product of coefficient method is ab.

The algebraic equivalence of the two point estimates of the mediation effect $\hat{c} - \widehat{c'}$ and $\hat{a}\hat{b}$ in single level mediation models does not hold in multilevel mediation models. Correspondingly, estimates of direct and indirect effects will not sum to the estimate of the total effect and the estimates of $\hat{c} - \widehat{c'}$ and $\hat{a}\hat{b}$ provide different values (Krull & MacKinnon 1999, 2001). Krull and MacKinnon (1999, 2001) have also noted that the MLM method is only applicable to the mediation models where a variable affects the other variable(s) in the same level or the lower level. In other words, the MLM method cannot be applied to multilevel models in which a variable at the lower level affects a variable measured at the higher level.

Since Krull and MacKinnon (1999, 2001) first discussed, researchers have demonstrated the usefulness of the MLM approach in many study designs. The MLM models for multilevel mediation analysis have been developed to cover many common data structures with different estimation methods. The advantage of multilevel mediation models over single level mediation models on multilevel data has been well documented. The multilevel mediation models that have been discussed may have fixed slopes (e.g. Krull & MacKinnon, 1999, 2001; Pituch et al., 2005;

Pituch et al., 2006; Pituch & Stapleton, 2008) or random slopes across higher level units (Kenny, Bolger, & Korchmaros, 2003; Bauer, Preacher, & Gil, 2006; Pituch, Murphy, & Tate, 2010).

Krull and MacKinnon (1999) proposed the MLM models to test the mediation effect in 2-1-1 data structure. Simulated data were used to compare the performance of the MLM approach and the OLS approach. Simulated data included a dichotomous independent variable X, and two continuous variables, mediator M and outcome variable Y. Simulated factors included the sample size, the values of *a* and *b*, and the ICC values. Specifically, the sample size was comprised of 10, 20, 30, 50, 100 or 200 groups with a either small or moderate group size. For the conditions with a small group size, a half of the groups had 5 individuals and the other half included 10 individuals. For the moderate conditions, a half of the groups had 20 individuals and the other half of 30 individuals. The ICC values of variables in level 1 were set up to either 0.05 or 0.15. Krull and MacKinnon (2001) used similar simulation conditions to examine the performance of the MLM approach in 1-1-1, 2-2-1, and 2-1-1 models for both dichotomous and continuous variables. In these studies, the multilevel models with no centering allowed random intercepts but constrained coefficients associated with mediators or predictors to be fixed. The authors found out that the standard error of the mediated effect with multilevel data in single level models is downwardly biased by 20% or more in multiple conditions. The eta squared analysis was used to identify the relationship between group size, ICC and the bias of the estimate of the standard error of the mediated effect. The eta squared values (ranging from .06 to 0.19) showed that group size and ICC were the factors that increase the extent of single-level underestimation for all examined mediation models.

Holding the constraint of fixed effect across upper units, Pituch et al. (2005) examined 2-2-1 and 2-1-1 models and compared the performance of four common methods of testing mediation effects in single level mediation models in multilevel mediation data. The first one is the Baron and Kenny (1986) method which requires a significant *t* test of paths a, b, and c. The second is the joint significance test proposed by MacKinnon (2002) that requires statistically significant *t* tests for paths *a* and *b* only. The third one is the Sobel (1982) method. This method requires the point estimate of the indirect effect, which is the product of paths *a* and *b* and its standard error. The forth method is called the asymmetric confidence limit proposed by MacKinnon at el. (2002). In this method, a confidence interval of the indirect effect, *ab*, is calculated. The mediation effect is supported if this interval does not contain zero. The critical values using to obtain the upper and lower limits are not taken from the z distribution as in the Sobel method but from the distribution of the product of two normally distributed variables. The simulated factors included the values of path *a* and *b*, ICC values, the number of groups, and group sizes. The values of path *a* and *b* varied from 0 to 0.6 and created the effect sizes of 0, 0.03, 0.12 and 0.24. The number of groups ranged from 10, 20 to 30 with small or moderate group sizes. Conditions with the small group size had half groups of 5 participants and the other half of 10. For the moderate group size conditions, a half groups consisted of 15 while the other half groups had 20 participants per group. The ICC values were set at either 0.05 or 0.15 which is the same setting as Krull and MacKinnon (1999, 2001). Without using group-mean centering, the asymmetric confidence limit method was shown to have more accurate Type I error rates than the other methods which in general had low Type I error rates. For conditions with one path (*a* or *b*) equal to 0, at the alpha level of 0.05, the asymmetric confidence limit method had Type I error rates varying from 0.42 to 0.51 while the other methods had Type I error rates from 0.007 to 0.45. Regarding the power for the test of the indirect effect, the asymmetric confidence limit test also had the greatest power among four tests. At the alpha level of 0.05, the average power of asymmetric confidence limit methods is .72 with 52% conditions having power greater than 80%. The other three methods had average power from .61 to 0.705. Results also showed that none of these test provided adequate power when sample size is small and effect size was .12 or smaller.

Pituch et al. (2006) continued to test the performance of six estimation methods in 2-2-1 and 2-1-1 mediation models in a Monte Carlo study using similar simulated conditions in Pituch et al. (2005). Among six methods, there are three bootstrap methods (parametric percentile bootstrap, bias-corrected bootstrap, and iterated bias-corrected bootstrap) and three single sample methods (the z test of *ab* product, the empirical-M test, and the joint significance test). The test performance was compared in terms of Type I error rates, statistical power, and the accuracy of confidence interval. MLM models were estimated with no centering and the results show that the bias-corrected bootstrapping had the most accurate Type I error rates. The other methods provided fairly accurate Type I error rates except the z test method which had low Type I error rates. The bias-corrected bootstrap method also had greatest power while the z test provided the least power. Further, the confidence interval produced by the bias-corrected bootstrap method was the most accurate.

Multilevel mediation models with random effects were examined by Kenny et al. (2003). They proposed a 2-step method to estimate the indirect effect and its standard error in lower level multilevel mediation models, the 1-1-1 model. The authors showed that the total effect, c, can be decomposed to the direct effect *c'*, the indirect *ab*, and the covariance of *ab*, δ_{ab}. In the first step, the path coefficients (*a*, *b*, *c*, and *c'*) and their variances are estimated separately using multilevel models with no centering. In the second step, the covariance between *a* and *b* is estimated using estimated coefficients *a* and *b*. A simulated dataset and an actual dataset were used to examine the performance of this approach. The results showed that Kenny's method provides more accurate model parameters than using single level methods. However, the authors admitted that the 2-step method still has some limitations including the inaccurate estimation of the total effect and covariance of *ab* due to the use of maximum likelihood estimation. The other limitation is the assumption of a joint multivariate normal distribution of path coefficients since the violation of this assumption might be more serious in the case of multilevel mediation models (Kenny et al., 2003). Bauer et al. (2006) extended Kenny's work by examining the 1-1-1 model and presented equations of MLM models with no centering to test the mediation effect. This method allows calculating all random indirect and total direct effect as well as their standard errors. Pituch et al. (2010) extended the models and methods for testing indirect effect to three-level designs. The authors presented equations of MLM models using no centering to examine the 3-1-1, 3-2-1 and 3-3-1 mediation models in which the effect of higher level variables on lower level variables was random. The procedures to estimate the indirect effect in these mediation models were illustrated using an actual educational dataset.

Researchers also examined the performance of two estimates of the mediation effect. Krull and MacKinnon (1999, 2001) showed that both $\hat{c} - \widehat{c'}$ and $\hat{a}\hat{b}$ estimates in multilevel models are unbiased as the estimates in single level models (using OLS methods). The standard error approximations in multilevel estimates were considerably less biased than that of single level estimates. They also demonstrated that even though the $\hat{c} - \widehat{c'}$ and $\hat{a}\hat{b}$ estimates are not algebraically equivalent, the discrepancy between these two estimates is typically small and close to zero with large sample sizes. This result was confirmed by Zhang et al. (2009) when they found no discrepancy between $\hat{c} - \widehat{c'}$ and $\hat{a}\hat{b}$ estimates under grand-mean centering and only very small differences under group-mean centering. According to Krull and MacKinnon

(1999), $\hat{c} - \hat{c'}$ is the estimate of the total effect while $\hat{a}\hat{b}$ is the estimate of a single unique mediated effect. The estimate of $\hat{a}\hat{b}$ can be summed to provide the total mediation effect, but the estimation of $\hat{c} - \hat{c'}$ cannot be proportioned to determine the contribution of each mediator in multi-mediator models. Therefore, in general, the use of $\hat{a}\hat{b}$ estimate is recommended over the $\hat{c} - \hat{c'}$ estimate (Krull & MacKinnon, 1999).

Zhang et al. (2009) pointed out that the relationship between two variables at level 1 in a multilevel mediation model still can be partitioned into within-group and between-group components. Under grand-mean centering, the within group coefficient is held equal to the between group coefficient. This constraint may provide confounded and incorrect estimates of the mediation effect. Specifically, when the within-group effect is greater than the between-group effect, the estimate of the mediation effect is greater than the true effect. On the other hand, when the within-group effect is smaller than the between effect, the mediation effect is underestimated. Zhang et al. (2009) suggested using group-mean centering to separate within and between variances of mediators and named this method as Unconflated Multilevel Modeling (UMM). The traditional MLM method using the grand-centering procedure proposed by Krull and MacKinnon (1999, 2001) was called Conflated Multilevel Modeling (CMM) method by Zhang et al. (2009). Equations to estimate mediation effects in the UMM method are different from those in the CMM method. For the 2-1-1 multilevel mediation model with the fixed effect assumption, equations (8) and (9) are changed into the equations (10) and (11), respectively as follows:

$$Y_{ij} = \beta_{X0j} + \beta_{X1j}(M_{ij} - M_{.j}) + r_{Xij} \qquad (10)$$

$$\beta_{X0j} = \gamma_{X00} + c'X_j + bM_{.j} + u_{X0j} \qquad (11)$$

where $M_{.j}$ is the mean of group *j*

Under group mean centering, the indirect mediation effect calculated by the product of the coefficients is equal to ab. Using Monte Carlo simulation, Zhang et al. (2009) examined the extent of confounding in using the grand-mean centering MLM procedure to estimate the 2-1-1 mediation model. The simulated factors were the within-group effect, the between-group effect, and the sample size. The within-group coefficient varied from -.59 to 0.59 while the between –group value ranged from 0, .14, .39 to .59. The sample size was fixed at 600 with four combinations of the group size (n = 5, 8, 12, and 20) and the number of groups. The estimate of the mediated effect using grand-mean centering was compared with the true population mediated effect. The result showed that grand-mean centering procedure inaccurate point-estimates of the indirect effect in most conditions. When the within-group coefficient was smaller or greater than the between-group coefficient, the difference between the point-estimate and the population value of the indirect effect was negative or positive, respectively. This pattern showed the confounded mediation estimate under grand-mean centering. Therefore, group-mean centering is necessary in some multilevel settings that have two level-1 variables, for example, the 2-1-1 model or the 1-1-1 model (Zhang et al., 2009).

Among the studies in multilevel mediation model, many applied the Monte Carlo method while varied in the simulation factors and the values of each factor. Simulation factors and significant findings of studies using the simulation method are summarized and presented in Table 1.

Mediation Models in MSEM

Structural Equation Modeling (SEM) is a general statistical modeling approach that encompasses many traditional statistical techniques from simple regression, and path analysis to discriminant analysis, canonical correlation, and factor analysis (Hox, 1998). The SEM approach is widely used in social sciences to study the relationship among unobserved (or latent) variables and observed variables. It hypothesizes that the population covariance matrix of observed variables is a function of model parameters. In other words, the population covariance matrix can be reproduced exactly by a correct model and the accurate set of parameters. Emphasizing the modeling of means and covariances, the SEM procedure aims to minimize the difference between hypothesized and observed covariance matrices (Bollen, 1989). A structural equation model normally includes a measurement model and a structural model. The structural model (or the latent model) summarizes the relationship between latent variables while the measurement model focuses on how a latent variable is measured by a set of observed variables.

Table 1
Summary of simulation factors and significant findings on some MLM mediation papers

Paper	Simulation factors	Findings
Krull and MacKinnon (1999, 2001)	The number of group, the group size, the values of *a* and *b,* and the ICC values	Group size and ICC increase the under-estimation of single level model
Pituch et al. (2005)	The number of group, the group size, the values of *a* and *b,* and the ICC values	The asymmetric confidence limit had better control of Type I error rate and higher power
Pituch et al. (2006)	The number of group, the group size, the values of *a* and *b,* and the ICC values	The bias-corrected bootstrap method had greatest power while the bias-corrected bootstrapping had the most accurate Type I error rates
Bauer et al. (2006)	The values of a and b, the covariance between a and b, the number of group, and the group size	Estimates are unbiased under most conditions
Zhang et al. (2009)	The within-group effect, the between-group effect, the number of group, and the group size	Results show the necessary in some multilevel mediation models

Multilevel Structural Equation Modeling (MSEM) is an advancement of SEM in order to apply the SEM framework into multilevel data analysis. Several MSEM methods have been proposed by researchers. However, early MSEM methods have some limitations. For example, the MSEM method proposed by Muthén (1994) requires decomposing observed scores into within and between covariance matrices and fit separate within and between models to estimate the parameters. Other methods were either unable to accommodate random slopes (e.g., Bauer, 2003; Curran, 2003; Mehta & Neale, 2005) or to handle unbalanced groups (Muthén, 1989, 1994; Muthén & Satorra, 1989).

Muthén and Asparouhov (2008) proposed a MSEM method using a model called the growth mixture model. Applying the SEM framework into multilevel data, this model not only allows random slopes but also handles missing data and unbalanced group sizes. The Muthén and Asparouhov MSEM model implements the maximum likelihood estimation method with the expectation maximization (EM) algorithm or accelerated expectation maximization (AEM) algorithm (Muthén & Asparouhov, 2008). The general multilevel mixture model proposed by Muthén and Asparouhov (2008) for a two-level data structure is represented in equations 12 through 14. This model is assumed to have *m* latent endogenous variables, and *n* latent exogenous variables. It is also assumed that the two-level multilevel data structure includes *p* observations (or individual cases) in *q* groups. The measurement component of Muthén and Asparouhov's two-level model (2008) is represented in equation (12).

$$Y_{ij} = v_j + \Lambda_j \eta_{ij} + K_j X_{ij} + \varepsilon_{ij} \quad (12)$$

where subscripts *i* and *j* refer to level-1 units and level-2 units (e.g., groups), respectively while Y_{ij} is a vector containing all dependent measured variables; X_{ij} is a q-dimensional vector of exogenous variables while K_j is a $p \times q$ matrix of slopes for the q exogenous variables; Λ_j is loading matrix while η_{ij} is an $m \times 1$ vector of random effects; finally, v_j is a p-dimensional vector of variable intercepts and ε_{ij} is a p-dimensional vector of error terms which is assumed to be multivariate normally distributed with a covariance matrix of Θ.

The structural component of Muthén and Asparouhov's two-level model (2008) is expressed in equation (13).

$$\eta_{ij} = \alpha_j + B_j\eta_{ij} + \Gamma_j X_{ij} + \zeta_{ij} \quad (13)$$

where α_j is an m-dimension vector of intercept terms, B_j is an *m x m* matrix of structural regression parameters, Γ_j is an *m x q* matrix of slope parameters for exogenous covariates, and ζ_{ij} is an m-dimensional vector of latent variable regression residuals which is assumed to be multivariate normally distributed with a covariance matrix of Ψ.

The multilevel part of Muthén and Asparouhov's two-level model (2008) is shown in the level-2 structural model in equation (14).

$$\eta_j = \mu + \beta\eta_j + \gamma X_j + \zeta_j \quad (14)$$

where η_j is a vector of random effects while vectors *μ, β, γ* contain estimated fixed effects. The residual term ζ_j is assumed to be multivariate normally distributed and independent across groups (Preacher et al., 2010).

Preacher et al. (2010) suggested adopting the Muthén and Asparouhov MSEM model into multilevel mediation analysis. The authors also showed that the MSEM approach in general has some advantages over the MLM approach in analyzing multilevel data. First, the MSEM by including latent variables can take into account measurement errors whereas MLM relies only on observed variables with an assumption of no measurement errors. Second, the between and within effect of level 1 variables can be separated in MSEM allowing the estimation of direct and indirect effects in each level. The MLM approach, on the other hand, conflates between and within effects of variables at level 1 (Preacher et al., 2010). Lüdtke et al. (2008) demonstrated that MSEM reduces the bias in estimates of contextual effects relative to a group mean-centered MLM approach. An additional advantage of MSEM is to provide an analysis of more multilevel mediation models that cannot be analyzed using the MLM approach such as mediation models with outcome variable in level 2.

Because this dissertation focuses on the 2-1-1 mediation model, the MSEM model of the 2-1-1 mediation model is described in Figure 4. In this figure, M_{ij} and Y_{ij} are observed scores of the mediator and the outcome variable for an individual *i* in group *j*, respectively; X_j is the observed score of predictor for group *j*; M_j and Y_j are means of mediator scores and outcome score for group j, respectively; c'_B, a_B, and b_B are the between-group effect of the predictor on the outcome, the between group effect of the predictor on the mediator, and the effect of the mediator on the outcome, respectively; b_W is the within-group effect of the mediator on the outcome.

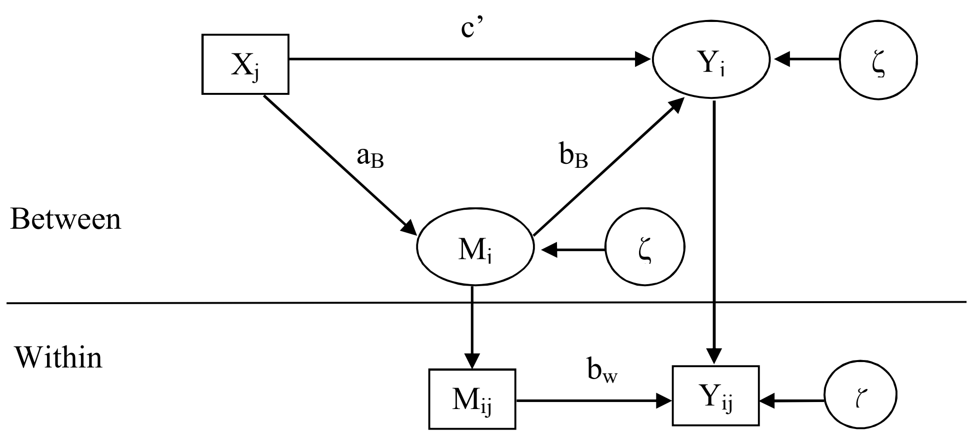

Figure 4. Multilevel structural equation model for 2-1-1 data

The MSEM model as applied to a mediation model is a special case of Muthén and Asparouhov's general model that does not have latent variables. A 2-1-1 MSEM mediation model with fixed slopes would apply following constrains: $v_j = 0$, $K_j = 0$, $\Gamma_j = 0$. Three equations (12), (13), (14) would be simplified to the equations (15), (16), (17), respectively.

Level 1 measurement model: $Y_{ij} = \Lambda_j \eta_{ij}$

$$= \begin{bmatrix} X_{ij} \\ M_{ij} \\ Y_{ij} \end{bmatrix} = \left[\left(\begin{array}{cc|ccc} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{array}\right)\right] \begin{bmatrix} \eta_{Mij} \\ \eta_{Yij} \\ \hline \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} \quad (15)$$

Level 1 structural model: $\eta_{ij} = \alpha_j + B_j \eta_{ij} + \zeta_{ij}$

$$= \begin{bmatrix} \eta_{Mij} \\ \eta_{Yij} \\ \hline \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \hline \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ b_w & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_{Mij} \\ \eta_{Yij} \\ \hline \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{Mij} \\ \zeta_{Yij} \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \hline \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} 0 \\ b_W * \eta_{Mij} \\ \hline 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \zeta_{Mij} \\ \zeta_{Yij} \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (16)$$

Level 2 structural model: $\eta_j = \mu + \beta \eta_j + \zeta_j$

$$= \begin{bmatrix} \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} = \begin{bmatrix} \mu_{\alpha\eta Xj} \\ \mu_{\alpha\eta Mj} \\ \mu_{\alpha\eta Yj} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ a_B & 0 & 0 \\ c'_B & b_B & 0 \end{bmatrix} \begin{bmatrix} \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{bwj} \\ \zeta_{\alpha\eta Xj} \\ \zeta_{\alpha\eta Mj} \end{bmatrix} \quad (17)$$

The fixed within-slope of Y_{ij} regressed on M_{ij} (b_W) is contained in the matrix of structural regression parameters, B. There is no within indirect effect because the model has only two variables (M_{ij} and Y_{ij}) with within variation. The between-indirect effect is calculated by multiplying the between-effect of X_j on M_j (a_B) by the between-effect of M_j on Y_j (b_B).

The suggestion of applying MSEM into multilevel mediation analysis has been extended by a few studies. Several mediation models for both 2-level and 3-level structural data using the MSEM approach have been proposed (Preacher, 2011; Preacher et al., 2011; Preacher et al., 2013). The performance of the MSEM method was compared with that of two MLM methods in the 2-1-1 mediation model with fixed slopes (Preacher, Zhang, & Zyphur, 2011). Two MLM methods are Conflated Multilevel Modeling proposed by Krull and MacKinnon (1999, 2001) and Unconflated Multilevel Modeling proposed by Zhang et al. (2009). The simulated conditions were defined by the values of ICC (from .05 to .40), the number of groups (from 20 to 1000) and group sizes (5, 20, or 50). Results of the study showed that the MSEM method significantly reduces bias and provides better coverage in the between indirect effect. On the other hand, the MSEM method has less efficiency in some small sample size and low ICC conditions.

Mediation analysis with non-normal data

Inferential statistics are based on several statistical assumptions (e.g. independence assumption, normality assumption, or equal variance assumption). Among them, the assumption of normality is common in many statistical techniques with continuous data. Statistical techniques used in mediation analysis also require the assumption of normality. The OLS technique used in single-level mediation analysis requires the residual to be normally distributed. In multilevel mediation analysis, the multivariate normality assumption is required for either the MLM method or the MSEM method.

The concept of normality can refer to either univariate normality or multivariate normality. Univariate normality is measured by the skewness and kurtosis of a variable's distribution. In terms of univariate normality, a variable is normally distributed if the absolute values of its skewness and kurtosis are both equal to zero. However, in practice, a variable with both the absolute values of kurtosis and skewness smaller than 1 is considered to satisfy the normality assumption because the effect of non-normality on the estimation result is insignificant (Muthén & Kaplan, 1985). Multivariate normality, on the other hand, is measured by multivariate skewness and kurtosis. There are various definitions of multivariate skewness and kurtorsis but the first and most common one was proposed by Mardia (1970). Let assume that x_1, x_2, …, x_N are the sample vectors of size N from a p-variate population with mean vector of μ and covariance matrix of Σ. The sample mean $\bar{x}$ and the sample covariance matrix S are calculated as follows:

$$\bar{x} = \frac{1}{N}\sum_{j=1}^{N} x_j \quad (18)$$

$$S = \frac{1}{N}\sum_{j=1}^{N}(x_j - \bar{x})(x_j - \bar{x})' \quad (19)$$

The sample measures of multivariate skewness and kurtosis defined by Mardia (1970) are:

$$b_1 = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\{(x_i - \bar{x})'S^{-1}(x_j - \bar{x})\}^3 \quad (20)$$

$$b_2 = \frac{1}{N}\sum_{i=1}^{N}\{(x_i - \bar{x})'S^{-1}(x_j - \bar{x})\}^2 \quad (21)$$

where b_1 and b_2 are the sample measures of multivariate skewness and kurtosis, respectively.

Both MLM and SEM approaches require the assumption of normality. However, non-normal data are frequently encountered in applied social science research (Micceri, 1989). Therefore, a great deal of researchers has conducted studies about the impacts of non-normal data on parameter estimates and standard errors. Maas and Hox (2004) examined the consequences of the violation of the normality assumption at the second level of the 2-level MLM model using simulated data. Four simulated factors were varied in this study including the number of groups, the group size, the ICC value, and the level-2 residual distribution. The number of groups varied from 30 to 100 while the group sizes were 5, 30 or 50. The ICC values were chosen among 0.1, 0.2 and 0.3. Three non-normal distributions of level-2 residuals were examined including a chi-square distribution with one degree of freedom, a uniform distribution, and a Laplace distribution. The authors found that the violation of the normality assumption on the second level has little or no effect on the fixed effects estimates but do have an effect on the random effect estimates. The estimates of random effects were unbiased but standard errors for the random effects at the second level are highly inaccurate if the distributional assumption was violated. For each parameter, the 95% confidence interval was established using the asymptotic standard normal distribution. The coverage probability of the 95% confidence interval was then used to assess the accuracy of standard errors of the each parameter. The coverage of 95% confidence interval of level-2 variance estimates was as low as .64 for the chi-square distribution and 0.85 for the Laplace distribution.

Byrd (2008) used Monte Carlo methods to evaluate the performance of the multilevel SEM method under conditions of non-normal data and different estimators (maximum likelihood, weighted least squares, and generalized least squares). Three two-level models were examined including (1) a model with one level-one predictor and one level-two predictor, (2) a model with two level-one predictors and one level-two predictor, and (3) a model with two level-one predictors and two level-two predictors. All three models have the dependent variable measured in level one. The simulated factors included degrees of normality, the number of groups and the group size while the normality conditions were explored for the level-one independent variables. The investigated group sizes included 10, 30, and 50 while the numbers of groups were 30, 50 and 100. Nine levels of non-normality were examined with skewness varying from -0.002 to 1.96 and kurtosis varying from -0.001 to 6.561. The results revealed that the coefficient estimates have little or no bias among the investigated conditions. However, the standard errors of parameter estimates at both level 1 and level 2 were affected by the number of groups. When the sample size was small (the number of groups was 30 with the group size of 10), the level-2 standard errors were biased downward by more than 25%. In addition, the standard error bias increased when the correlation among variables and kurtosis increased. The bias associated with the level-2 standard errors became near zero with the large sample size (the number of groups was 100 and group sizes were 30 or 50).

Zhang (2005) compared the performance of MLM and multilevel SEM approaches in the presence of non-normal incomplete data. The author used simulated two-level data with one level-one predictor, one level-two predictor, and one level-1 outcome variable. The cross-level interaction between two independent variables was also used to predict the dependent variable. Data normality was explored for the first level predictor pooled across all the observations and groups. Three configurations of data distribution were examined including: normality, moderate non-normality (skewness of 0.5 and kurtosis of 1.0), and severe non-normality (skewness of 1 and kurtosis of 3.75).The sample size and the percentage of incomplete data were other simulated factors. Finding of the study revealed that neither the MLM approach nor multilevel SEM approach was sensitive to the violation of normality assumption. Both MLM and multilevel SEM methods had the same power in detecting the main effect while MLM appeared to have better power for cross-level interaction. A possible explanation of results was that the study inflated the statistical power in non-normal conditions without control for Type I error rates. Zhang (2005) reported the result from a separate study showing that the non-normality might nearly double the Type I error risk when Type I error rates were found to be around .10 instead of the nominal level of .05 under severely non-normal data.

Problems caused by the violation of normality assumption in a mediation analysis (the indirect effect) were also explored in several studies. However, most of studies have investigated the problems in the single level mediation model. The only study that directly addressed the multilevel modeling context is Pituch and Stapleton's work (2008). Regarding the MSEM approach, even though its advantage was pointed out by Preacher et al. (2010, 2011), its performance has not been fully examined. Especially, no study has investigated the performance of MSEM on mediation analysis under the condition of non-normal data. Studies in the single level mediation models revealed that the violation of the normality assumption affects the estimate of standard error of indirect effect when using the maximum likelihood estimator (Finch et al., 1997; Zu, 2009). Among methods to test the indirect effect in single level mediation models under non-normal conditions, bootstrapping and the hierarchical Bayesian MCMC methods have the best performance (Biesanz et al., 2010). Bootstrapping is also the best method to test the mediated effect in multilevel mediation models with non-normal data (Pituch & Stapleton, 2008).

Finch et al. (1997) used Monte Carlo methods to investigate the effect of sample size and non-normal data on the estimation of mediated effects in latent variable models. The simulated factors included (1) the sample size, (2) the non-normality level of data, (3) the population values of the model parameters, and (4) the types of estimators. Three types of estimators were examined including Maximum Likelihood (ML), Asymptotically Distribution Free (ADF), and Maximum Likelihood robust (ML-robust). Three distributional conditions were examined including: normal distribution (skewness and kurtosis equal to 0), moderate non-normal distribution (skewness = 2 and kurtosis = 7), and extreme non-normal distribution (skewness = 3 and kurtosis = 21). The results showed that maximum likelihood estimates of the standard error of the mediated effects were biased under severely non-normal conditions. However, the ADF and ML-robust estimates of the standard errors of both the mediated effect and the structural parameters were not affected by the non-normality.

Zu (2009) argued that classical methods using the Maximum Likelihood (ML) estimator are affected by non-normality. The author proposed three robust procedures for mediation analysis under the conditions of the violation of the normality assumption. These procedures use robust M-estimators instead of ML estimators. Simulation studies were then conducted to compare the performance of the robust M-estimator with that of the ML estimator in terms of estimating the mediated effect, the standard error of the mediated effect and the confidence interval of the mediated effect. The manipulated factors were sample sizes, data distributions, and effect sizes. Five distributions were considered in this study: normal distribution, T_5 distribution (skewness of 0 and kurtosis of 6), contaminated normal distribution (skewness of 0 and kurtosis of 5.49), skewed distribution (skewness of X, M, and Y varied from 2.975 to 3.262 while kurtosis values varied from 20.479 to 23.540), and heteroscedastic distribution (the variance of the outcome variable given all the predictors is not a constant). Findings of the study showed that mediation analysis using the ML estimator is sensitive to the violation of the normality assumption and the proposed robust procedures of mediation analysis have better results. Specifically, the estimate of the mediated effect had a little bias across estimation methods under non-normal conditions. However, the standard error was substantially underestimated for most non-normal conditions when the ML estimator was used. In the small sample size conditions (n=50), the relative bias ranged between -0.596 and -0.299. The lager sample size conditions failed to reduce the relative bias of the estimate of the standard error.

Biesanz et al. (2010) examined different methods to assess indirect effects with normal and non-normal data as well as with complete and incomplete data using simulated data. Non-normal data were generated by transforming the dependent variable to have skewness of 2 and kurtosis of 7. Methods to assess the indirect effect examined in this study included the classic standard error method, bootstrapping methods, the distribution of the product methods, the hierarchical Bayesian MCMC method, the causal steps method, and the partial posterior method.

The performance of these methods were compared in terms of Type I error rates, power, and coverage of confidence intervals. Results showed that the bootstrapped percentile confidence interval and the hierarchical Bayesian MCMC methods provide the best performance overall by holding Type I error rates, presenting reasonable power, and having stable and accurate coverage rates.

Pituch and Stapleton (2008) evaluated the performance of standard and robust methods to test indirect effects for upper level mediation models with the fixed direct effect under the presence of non-normal data. The standard methods assume that data are normally distributed while the robust methods do not need the assumption of normality. The distribution of simulated data were systematically varied from a normal condition (skew = 0, kurtosis = 0) to three non-normal conditions which are moderately non-normal condition (skew = 1.63, kurtosis = 4), substantially non-normal condition (skew = 2.28, kurtosis = 12), and a "worst-case" condition (skew = -2.28, kurtosis = 12). The non-normality was simulated at both level-1 and level-2 data. Four standard methods were investigated including z test of the product of ab, empirical-M test, and two parametric bootstrap methods: the parametric percentile residual bootstrap, and the bias-corrected parametric percentile residual bootstrap. Six robust methods being examined included two robust versions of the z test and empirical-M test, and four nonparametric bootstrap methods which were the modification of the parametric bootstrap methods. The robust versions of z and empirical-M tests involved using robust standard errors for the indirect effect. The robust standard errors of the indirect effect were calculated by the robust estimates of standard errors of path coefficients *a* and *b*. The nonparametric bootstrap methods used the estimated residuals instead of sampling from a population distribution of residuals as in parametric bootstrap methods. The results showed Type I error rates were elevated when departures from normality increased. All methods had Type I error rates of .088 or higher under worst-case non-normality conditions. While the stratified bias-corrected nonparametric bootstrap and standard bias-corrected nonparametric bootstrap were found to have the greatest power, the average statistical power of these methods ranged from .395 to .526 under worst case non-normality conditions. In general, the empirical-M test and bias-corrected parametric bootstrap methods perform better than the other methods. Regarding the confidence interval estimation, robust methods have low coverage rates and sometimes the coverage was too small even under the large sample size.

Summary

Chapter II provides a review of the literature of mediation analysis with multilevel data. As evidenced in chapter II, many methods to test the indirect effect have been proposed under the MLM framework and the MSEM framework. Researchers considered the sample size, the ICC value, and the effect size as the factors having the impact on the test of the indirect effect. The violation of the normality assumption is another issue that has a potential impact on the test of the indirect effect. The population distribution was found to have a little or no effect on the coefficient estimates but have significant effects on standard errors of coefficient estimates under either MLM or MSEM. The standard error of mediated effect was also reported to be biased under non-normal situation with the single mediation model. Regarding multilevel mediation analysis, to date, only Pituch and Stapleton (2008) investigated the performance of the MLM tests under non-normal data. Their findings were that the Type I error rates were elevated when data are not normally distributed. Multilevel mediation tests under MSEM framework were recently proposed by Preacher et al. (2010) with many advantages over MLM tests. However, it is not clear how the MSEM tests perform with non-normal data. Given that non-normal data are common in most educational settings, it is important to understand how the MSEM mediation test performs under such conditions.

CHAPTER THREE: METHOD

This study was designed to examine the statistical performance of multilevel mediation analysis methods under various degrees of non-normal data using Monte Carlo method. Multilevel data of 2-1-1 was simulated and analyzed by MLM and MSEM methods. The performance of MLM and MSEM methods was examined in terms of convergence rate, Type I error control, and statistical power. The impact of simulated factors and possibility of interaction among factors on performance of these analysis methods was explored. The study provided some guidelines for researchers in conducting data analysis in similar conditions.

Design

This simulation study involved five independent factors that combined to create 4x3x5x3x3 = 540 simulated conditions. The simulated factors include: population between-indirect effect, the degree of non-normality, sample size that includes number of groups and group size, and the population intra-class correlation (ICC) of M and Y.

Population between-indirect effect

The values of the population between-indirect effect was set up to 4 conditions including one null condition and three non-null conditions. The null conditions, in which ab = 0, was obtained by setting a = 0, b= 0. This condition was used to calculate the Type I error rates of the tests. The three non-null conditions were obtained with true *ab* effect equal to: 0.03 (a=0.3, b=0.1), 0.12 (a=0.3, b=0.4), and 0.24 (a=0.6, b=0.4). This set of effect values represents the small, moderate, large and no effect sizes as used in previous studies (e.g. Krull & MacKinnon, 2001; Pituch & Stapleton, 2008).

Degree of non-normality

In multilevel models, the violation of the normality assumption can happen either in the level one, or level two, or even in both levels. Zhang (2005) and Byrd (2008) investigated the effect of non-normality at level one on the direct effect in both MLM and MSEM models. The impact of non-normality at the level two on the direct effect in MLM models was explored by Maas and Hox (2004). Regarding the mediation effect, only the effect of non-normality on the level two was investigated in MLM models by Pituch and Stapleton (2008). This study tried to fill a portion of this gap in literature by investigating the impact of the level one non-normality on mediation effect. The degree of non-normality was varied systematically for variables in level 1: the mediator and the outcome variable. Three data distributions were configured: normal data (skewness = 0 and kurtosis = 0), moderately non-normal (skewness = 1.63, kurtosis = 4), and severely non-normal (skewness = 2.28, kurtosis = 12). The combinations of skewness and kurtosis in moderately non-normal and severely non-normal distributions were the same as the setting in a previous simulation study of multilevel mediation with non-normality (Pituch & Stapleton, 2008). For the normal conditions, univariate normal sample data were generated for each variable. The Cholesky decomposition approach was used to create a multivariate normal distribution with a given correlation matrix. For the non-normal conditions, the Fleishman's transformation method (Fleishman, 1978) was used to obtain a given non-normality degree of a variable. However, the processes of generating multivariate normal data and transforming each variable to a specified distributional shape interact with each other and cause the sample data to deviate from the desired correlation matrix (Vale & Maurelli, 1983). Vale and Maurelli (1983) proposed a procedure of calculating an intermediate correlation matrix to counteract the impact of nonnormality transformation on the correlation matrix. The procedure of generating correlated non-normal data, therefore, includes three steps:

(1) Calculate the intermediate correlation matrix.

(2) Generate multivariate normal data with the calculated intermediate correlation matrix.

(3) Transform each variable to a specific population distribution.

Sample size

Sample size conditions were defined by the combination of both number of groups and the group size. The number of groups selected in this study was similar to the values used in the Preacher et al. (2011) study. The number of groups, or the number of level-2 units was specified as J = 20, 50, 100, 300, or 1000. The sizes of group at level 1 were unequal but average group sizes were similar to the group sizes used by Preacher et al. (2011). The average group sizes were set up with three values: small size of 5, medium size of 20, and large size of 50. For the small group size conditions, half of the groups had 3 units and the other half of 7 units. For the medium group size conditions, half of the groups were composed of 15 units, the other half of 25 units. For the large group size, half of the groups had 39 units and the other half of 61 units. Combining values of the group size and the number of groups creates 3 x 5 = 30 sample size conditions.

The ICCs of M and Y

The effect of ICC values on the result of multilevel mediation analysis was not clear. Krull and McKinnon (2001) identified ICC as a factor that extends the underestimation of the standard error of the mediated effect in single-level models with cluster data. The simulated factor of ICC values was also found having an impact on the bias of the estimate of the indirect effect as well as the power of the test of the indirect effect in MSEM mediation analysis (Preacher et al., 2011). However, Maas and Hox (2004) found that the estimation of fixed effects and their standard errors are not different across ICC level on the robust and standard multilevel analysis methods. Hox et al. (2010) also noted that the difference in ICC has no effect on the parameter estimates and their standard errors on the MSEM models. Pituch et al. (2005) did not find any effect of ICC values on the MLM mediation analysis. The ICC value was then considered as a simulated factor in this study. The population ICC was identical for both mediation variable and dependent variable and set to a value of 0.05, 0.20, or 0.40. These ICC values were considered among the common ICC values in several previous studies (Hox, 2002; Muthén, 1994; and Preacher et al., 2011).

Constant values

In addition to the parameters varied as simulated factors, the values of other parameters in the multilevel models will be held constant throughout the simulation. The direct effect of the independent variable X on the dependent variable Y (the c' on the MLM model or c'_B in the MSEM model) will be set to 0.1. The within direct effect of M on Y, b_W, will be set to 0.2. These constant values are similar to the setup in the Preacher et al. (2011) study and the Pituch et al. (2008) study. The population variance of the independent variable X as well as population variances of group means of the mediator M and outcome variable Y all will be specified as 1 while their population means all will be set to zero.

Data

The data matrix of the three variables in the mediation model and one variable identifying group membership were generated using SAS/IML. All three variables were generated as continuous variables. For each simulation condition, 1000 samples were generated and analyzed.

The procedure of generating sample data includes two steps: level-2 data were generated in the first step, while level-1 data were created in the step 2 using information from level-2 data.

The level-2 data matrix includes the independent variable X and group means of the mediator M and the outcome Y. The population correlation matrix of these three variables was defined using between-group path coefficients, a_B, b_B, and c'_B. The Cholesky decomposition approach was used to generate a multivariate normal data matrix of three variables with means of zero and variances of 1.

The level-1 data matrix includes the level-1 values of mediator and outcome variable in each group. The sample data were generated separately for each group. The correlation matrix of the mediator and the dependent variable was defined based on the within direct effect of the mediator on the outcome variable. Because the within direct effect is constant, the population level-1 correlation matrix was fixed across groups. The population variances of level-1 variables were calculated based on the ICC value and the variances of group means (which are equal 1) in the level-2 data. The population means of level-1 variables in each group were derived from the generated level-2 sample data. Multivariate non-normal sample data of two variables were generated for each group by the procedure described in the degree of non-normality section above.

Procedure

This study used the MLM method proposed by Zhang et al. (2009) called UMM method which uses group-mean centering to partition the between-group and within-group effects. The MSEM method used here is the one proposed by Preacher et al. (2010). The nature of MSEM allows separating the between-group and within-group variation with no centering. Each simulated sample was fit with the MLM and MSEM mediation models. The MSEM mediation model was estimated by using Mplus while the MLM model was estimated by using the Statistical Analysis System software (SAS). The Restricted Maximum Likelihood (REML) estimation method was considered since it might provide better estimates of variances and covariances of random effects than Maximum Likelihood (ML) estimation method (Preacher et al., 2011). However, the REML is included in SAS but is not available in Mplus. Hence, both estimation procedures were conducted with the Maximum Likelihood (ML) estimation method.

The estimate of the indirect effect can be calculated by either the product of coefficients method or the difference of total effect and direct effect method. Both methods yield the same result in single level mediation models. They also have very similar results in MLM mediation models. However, the product of coefficients method was still recommended for multilevel data since the estimates of this method can provide greater amounts of information in models with more than one mediator (Krull & MacKinnon, 1999). The two coefficients here are the direct effect of the independent variable X on the mediation variable M (denoted by *a*) and the direct effect of the mediation variable M on the dependent variable Y (denoted by *b*).

The estimate of the between-indirect effect in this study was calculated using product of coefficients method. It is a_B*b_B for the MSEM model or $a*b$ for the MLM model. Various testing methods to test the indirect effect in multilevel mediation analysis have been proposed. Pituch and Stapleton (2008) listed and compared the performance of ten different testing methods. Among these tests, bootstrapping and the empirical-M test have been shown having the best performance, especially, in the presence of non-normal data (Pituch & Stapleton, 2008). In this study, the bootstrapping method was used to test the significance of the indirect effect.

Bootstrapping is a resampling procedure that allows building a sampling distribution of a statistics from observed data without knowing about the distribution of this statistics.

Bootstrapping procedure when applying to the multilevel models needs to be adjusted to take into account the hierarchical structure of data. In general, there are three main bootstrap methods for multilevel models: the case bootstrap, parametric bootstrap and non-parametric bootstrap. The parametric bootstrap and non-parametric bootstrap methods require both model and residual distributions to be specified correctly while the case bootstrap method does not require any assumption except that the hierarchical structure of data is correctly specified. Therefore, the case bootstrap method is more robust to model misspecification but inefficient otherwise (Carpenter, Goldstein, & Rasbash, 2003). However, residual bootstrap methods, either parametric or nonparametric are preferred over case bootstrap for several reasons. First, the values of explanatory variables would be the same if only the residuals are resampled. Second, the resampling of residuals only also maintains the assumption of the independence of sampled responses. Further, resampling cases at the highest level might change the correlation structure of the initial data set if the sample size of upper level is not large enough (Pituch and Stapleton, 2008). Between parametric residual bootstrap and non-parametric residual bootstrap, Carpenter et al. (2003) showed that the latter takes into account non-normality in the distribution of residuals and hence gives more accurate inferences.

In this study, the non-parametric residual bootstrap method was used to test the statistical significance of an indirect effect. The number of resamples in the bootstrap procedure was 1000 which is similar to the setting used by Pituch and Stapleton (2008). The bootstrap procedure approximates the distribution of the indirect effect for each simulated sample. The confidence interval derived from the bootstrap distribution was used to test the indirect effect. The indirect effect is insignificant for a replication if the confidence interval of sample data contains the value of zero.

The non-parametric residual bootstrap procedure was proposed by Carpenter et al. (2003) and was described in detail to conduct in SAS by Wang, Carpenter, and Kepler (2006). Applying the procedure into the 2-1-1 mediation model with fixed slope and random intercept, the non-parametric residual bootstrap was conducted by the following steps:

Step 1: The multilevel model was run with the original simulated sample. The random and fixed effects of intercepts and coefficients were saved. The level-1 and level-2 residuals were obtained for the next steps.

Step 2: Level-1 and level-2 residuals were rescaled by centering to avoid bias estimates of the intercept. The rescaled residuals then were transformed to have the covariance matrix equal the model estimated residual covariance matrix.

Step 3: Random samples were drawn with replacement from transformed residual data and saved as bootstrap residual data for both level 1 and level 2. The bootstrap residuals were then merged with original level 1 and level 2 data to calculate bootstrap data.

Step 4: The multilevel model was run with the bootstrap resample to obtain the estimated fixed effects.

Step 5: Steps 3 and 4 were repeated 1000 times to get 1000 sets of the estimated fixed effect. The confidence interval of the fixed effects of the original simulated sample was approximated based on these bootstrap estimates.

The procedure of transforming residuals in step 2 requires the estimated variances of intercepts and coefficients to differ from zero. However, the default setting of the SAS software program that is used to estimate the MLM model prevents the variances from having negative or zero values. Estimating the MLM model using default setting in SAS resulted in variances of zero in many samples, especially in small size samples. Pituch and Stapleton (2008) suggested allowing software to estimate negative variances to produce the consistent bootstrap estimates. Therefore, the variances were allowed to be negative when estimating the MLM model in this study. The number of simulated samples with variances equal to zero, however, was still recorded for each simulation condition.

The performance of the MLM and MSEM methods was examined based on the following statistical indices in each simulated condition: statistical power, Type I error rate, and convergence rate. The indices were calculated separately for each method.

Type I error rate (at alpha level of .05) and statistical power were calculated based on the test result associated with the indirect effect for each replication. The Type I error rate was calculated using null conditions. For each null condition, the Type I error rate is the proportion of replications in which the statistical test shows the mediation effect is significant. The statistical power, on the other hand, is the proportion of replications in each non-null condition in which the null hypothesis of no mediation effect is rejected.

Convergence rates are important because of their implications for the practical application of each method. Convergence rates were calculated for each condition as the proportion of replications in which the statistical model converged. Results for Type I error and power were based on the converged replications only.

Boxplots describing the distribution of each outcome variables (i.e. Type I error rate, power rate, and convergence rate) were first investigated for a general view and comparison of two methods. In addition, the results of simulation were evaluated by using analysis of variance to calculate the eta-squared associated with each simulated factors as well as their first-order interactions. The eta-squared analyses were conducted for all performance indices. The eta-squared values exceeding Cohen's (1988) medium effect size criteria of .0588 were considered significant and were further explored.

The simulation procedure was conducted in three phases. Simulated data were created in the first phase using SAS/IML. In the second phase, the MLM model was fitted into the simulated data using Base SAS MIXED procedure. The SAS software was also used to generate the bootstrapping data and fit the MLM model into the bootstrapping data in this phase. In the third phase, the MSEM model was fitted into the same generated data using Mplus. The SAS software again was used to generate the bootstrapping data and the MSEM model was fitted into each bootstrapping data. The SAS code and Mplus code for whole simulation procedure can be found in the Appendix C.

CHAPTER FOUR: RESULTS

This study is intended to compare the MLM and MSEM methods in detecting mediation effect under non-normal data across different sample sizes and values of ICC. Two types of analyses were used to evaluate the performance of these methods. First, descriptive statistics and boxplots were used to provide the overview of the performance of two methods. Eta-squared analyses were then applied to detect simulated factors that have a significant effect on the performance of each method in various statistical criteria: Type I error and statistical power, and convergence rate. To simplify the interpretation and presentation of results among five simulated factors, only eta squared (η^2) associated with each simulation design factor and their first order interactions were calculated and report in this study. The eta squared values exceeding Cohen's (1988) medium effect size criteria of .0588 were considered significant and were further explored. Descriptive statistics tables and boxplots were then employed to compare these criteria of the two methods across different levels of each significant simulated factor.

Convergence Rate of MSEM and MLM Model

Convergence rates were considered in this study because of their important implications for the practical application of each method. Convergence rates were calculated for each condition as the proportion of replications in which the statistical model converged. Convergence rate was examined on both null and non-null conditions with the total of 540 conditions for both the MSEM and MLM models. The MSEM model had perfect convergence rate for all 540 examined conditions.

Regarding the MLM model, when the default constraint of non-negative variances was applied; estimated variances of intercepts and coefficients were equal to zero for some simulated data, especially with the small sample size. The zero variances made the bootstrap procedure inapplicable in these cases. In order to solve this problem, Hox (2002) and Pituch & Stapleton (2008) suggested allowing negative variances in the computation to produce consistent bootstrap estimates. This study followed the Pituch and Stapleton (2008)'s suggestion by allowing SAS to compute negative variances. Convergence rates of the MLM model without constraint from negative variances were also excellent across simulated conditions.

The convergence rate of MLM model estimated with a constraint in variances was calculated and reported in Table 2. The mean estimated convergence rate was grouped by ICC, group number and cell size.

The convergence rate was low with the small sample size and low ICC condition. The convergence rate increased as ICC increased. Sample size, either group number or cell size, also showed positive relationship with the convergence rate. Perfect convergence rates were observed with the high level of ICC and the large sample size.

Type I error control of MSEM and MLM mediation models.

The Type I error control of two mediation analysis methods was examined using 135 null conditions. In these conditions, the data were generated with a mediation effect of zero which was obtained by both coefficients *a* and *b* that are equal to zero. For each null condition, the Type I error rate is the proportion of replications in which the statistical test shows that the mediation effect is statistically significant.

Table 2
Convergence Rates of MLM Model with Variance Constrain by Group Number, Cell Size, and ICC Values

		ICC		
Cell size	Group Number	.05	.20	0.40
5	20	0.33183	0.86100	0.98967
	50	0.59883	0.99650	1
	100	0.80883	0.99967	1
	300	0.98033	1	1
	1000	0.99950	1	1
20	20	0.82683	0.99967	1
	50	0.99250	1	1
	100	1	1	1
	300	1	1	1
	1000	1	1	1
50	20	0.98733	1	1
	50	1	1	1
	100	1	1	1
	300	1	1	1
	1000	1	1	1

The Type I error rates of the MLM and MSEM models are summarized in Table 3. The average of rejection rates of null conditions for MSEM method were found to be close to zero across group numbers and cell sizes. Using the nominal alpha level of 0.05, the performance of the MSEM model was very conservative. Due to this result, the eta-squared analysis was not conducted for the Type I error rates of the MSEM model.

Table 3
The Average of Type I Error Rates for MLM an MSEM Models by Group Number and Cell Size

K	MLM Model			MSEM Model		
	$n_j = 10$	$n_j = 20$	$n_j = 50$	$n_j = 10$	$n_j = 20$	$n_j = 50$
20	0.00911	0.01044	0.01000	0.00133	0.00222	0.00267
50	0.02311	0.01689	0.01467	0.00222	0.00089	0.00222
100	0.03556	0.03222	0.02400	0.00133	0.00133	0.00178
300	0.05178	0.05022	0.05178	0.00178	0.00133	0.00044
1000	0.05044	0.05244	0.05022	0.00089	0.00133	0.00089

Note. K = number of groups; n_j = within group sample size; MLM = multilevel modeling; MSEM = multilevel structural equation modeling.

Eta-squared analyses were performed with the rejection rates of null conditions for the MLM method. The result showed that the number of groups (with eta squared of .77) was the only simulated factor that has the significant impact on the variation of Type I error rates. The different of Type I error rates across the number of groups can be observed in Table 3. This table shows that the MLM mediation model using bootstrapping method was under control for Type I error rates with the small number of groups. The Type I error rates increased when the number of group increased and it reached to the expected alpha level of .05 at the group number of 300.

Statistical Power of MLM and MSEM Mediation Models

The statistical power of the MLM and MSEM mediation models were examined using 405 non-null conditions which vary systematically in the mediation effect size, sample size, ICC value, and the degree of non-normality. The power is the proportion of replications in each non-null condition in which the null hypothesis of no mediation effect was rejected. Mean and standard deviation of the estimated power of these 405 non-null conditions were calculated for the MLM and MSEM models. The MLM model was found to have higher power (with mean of 0.677 and standard deviation of 0.354) than the MSEM model (with mean of 0.411 and standard deviation of 0.386). The eta-squared (η^2) analyses were conducted for the power rates of the MSEM and MLM models separately. The outcomes of both eta-squared analyses are summarized in Table 4. The number of groups ($\eta^2 = 0.459$ and 0.755 for the MSEM and MLM models, respectively), mediation effect size ($\eta^2 = 0.243$ and 0.104 for the MSEM and MLM models, respectively) emerged as the major factors related to power for both models. The interaction between the group number and effect size ($\eta^2 = 0.069$ and 0.066 for the MSEM and MLM models, respectively) was also associated with the variability in statistical power of both models. The value of ICC ($\eta^2 = 0.065$) and the cell size ($\eta^2 = 0.061$) were also found to have the significant effect on the power of the MSEM model but not the MLM model.

Table 4
Eta-squared Analysis of Statistical Power for MLM and MSEM Models

Factors	MSEM	MLM
Size2	**0.45981**	**0.75537**
ME	**0.24311**	**0.10357**
ME*Size2	**0.06891**	**0.06633**
ICC	**0.06548**	0.01808
Size1	**0.06071**	0.01815
Size2* ICC	0.01112	0.00461
Size2* Size1	0.01015	0.00514
Size1* ICC	0.01005	0.00871
ME* ICC	0.00953	0.00112
ME* Size1	0.00791	0.00086
Size2* shape	0.00017	0.00001
Size1* shape	0.00003	0.00001
shape	0.00003	0
ICC* shape	0.00002	0.00001
ME* shape	0.00001	0

Note. The effect sizes with significant main effect and interaction in power appear in bold; Size2 = number of group; Size1 = group size, ME = effect size; ICC = ICC value; shape = population shape.

This study aimed to examine the performance of these testing approaches under the non-normal distribution conditions. Data at two levels of non-normality were generated accompanied with normal data to make the comparison. However, the factor of population shape was found to have no significant effect on power for both the MLM and MSEM models. Figure 5 provides an evidence for the similarity in term of power between normal conditions and non-normal conditions for both models. It again shows that the MLM model had higher power than the MSEM model in either non-normal conditions or normal conditions. Within each testing method, the power of both normal and non-normal groups had the similar mean and variability.

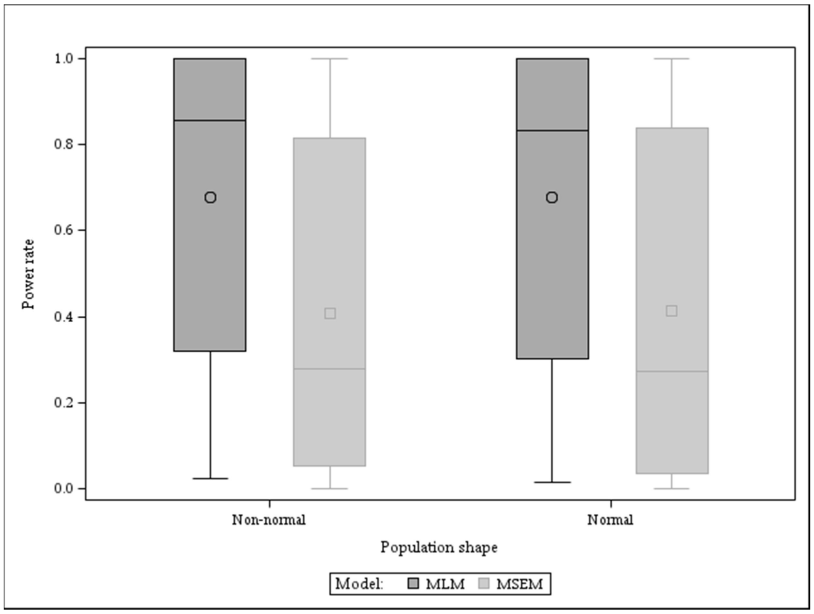

Figure 5. Statistical Power of MLM and MSEM Models by Population Shape

Eta squared analyses showed that the group number had significant effects on power of both the MLM and MSEM models (η^2 of 0.755 and 0.460, respectively). The effect of the group number on power for both models is presented in Figure 6. The MSEM model had the smallest power (power mean of 0.055) when the group number was 20. Power increased as the number of groups increased with the average power of 0.2 and 0.402 for the group number of 50 and 100, respectively. The MSEM model had many conditions with the power ranged from .8 to 1, especially the conditions with group number greater than 300. However, the average powers were 0.644 and 0.753 for conditions with group numbers of 300 and 1000, respectively. A further investigation revealed that the conditions having reasonable power (power of .8 or greater) had either medium or large effect sizes. Most of these conditions had a large sample size with the group numbers of 300 or 1000. It was noticed that the ICC value also had an effect on the power since there were some conditions with the group number of 100 and the cell size of 50 with reasonable power. These conditions were all had large ICC value of 0.4.

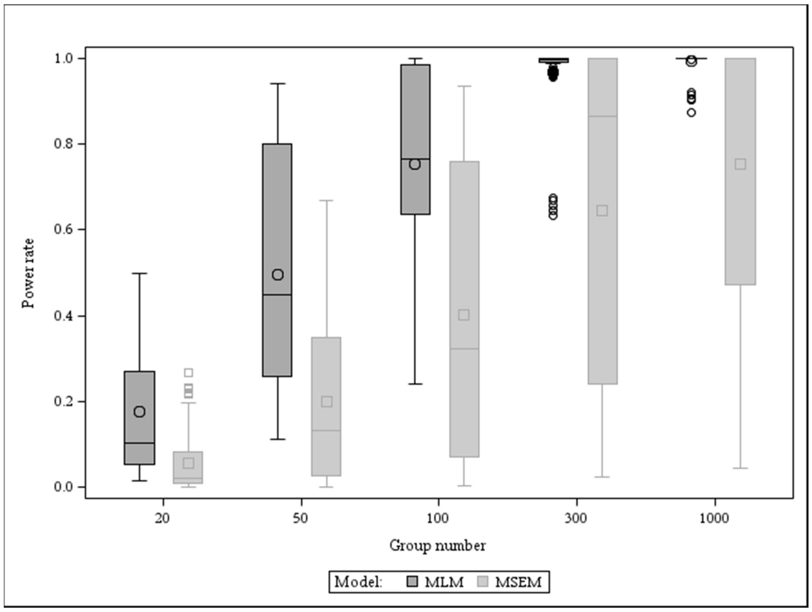

Figure 6. Statistical Power of MLM and MSEM Models by the Number of Groups

Regarding the MLM model, power was also smallest (power mean of 0.178) with the group number of 20 and increased as the group number increased. It is noted that the MLM model had a higher power than the MSEM model in each level of group numbers. The MLM model had average power of 0.495 and 0.753 for the group number of 50 and 100, respectively. The power of conditions with the group number of 300 and 1000 reached to 1.0 for most conditions with average power of 0.968 and 0.993, respectively. The variance of MLM model' power is larger than that of MSEM model's power with the small group numbers (group numbers of 20 and 50) but much smaller in cases of large group numbers (the group numbers of 100, 300 and 1000).

The impacts of effect size on the power are presented in Figure 7. Power of the MSEM model was smallest as the mediation effect of 0.03 (power mean of 0.055 and standard deviation of 0.07). The power increased as the effect size went up with power mean of 0.199 and 0.401 for effect size of 0.12 and 0.24. Power of the MLM model also got bigger when effect size increased. The MLM model has higher power than MSEM model in each effect size level with power mean of 0.559, 0.642, and 0.831 for the effect size of 0.03, 0.12, and 0.24, respectively.

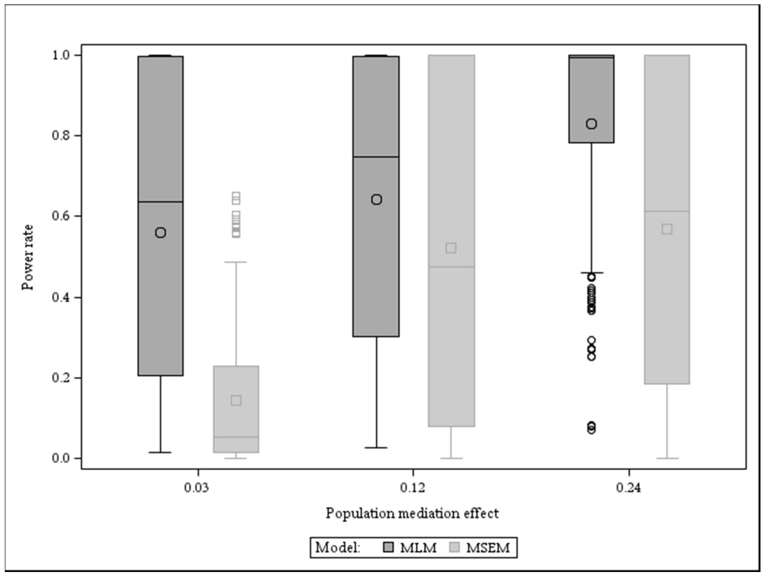

Figure 7. Statistical Power of MLM and MSEM Models by Mediation Effect Sizes

The interaction of effect size and the number of groups showed a substantial relationship with the power for both MLM and MSEM models ($\eta^2 = 0.066$ and 0.069, respectively). Figure 8 shows the effect of the interaction between group number and effect size on power for the MLM model. Power was observed to go up as the number of groups increase for all three effect size levels. The effect size of .24 has the highest power while the effect size of .03 has the smallest power. The difference in power among effect size levels was largest at the group number of 50 and decreased when the number of groups increased. Power of all effect sizes got closed to 1 at the group number of 1000. Figure 9 shows the effect of the interaction between group number and effect size on power for the MSEM model. The difference in power among effect sizes is smallest at the group number of 20. Similar to the MLM model, power increased as the number of groups went up for all three levels of effect sizes. However, the increase was smaller for effect size of .03. It was also noted that the effect sizes of .12 and .24 have similar power across the group number of 100, 300 and 1000.

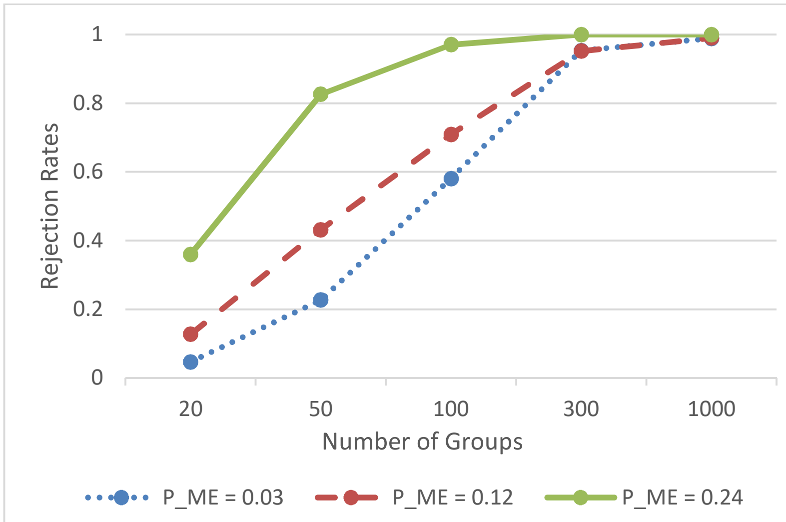

Figure 8. The Interaction between Group Number and Effect Size on Power for the MLM Model

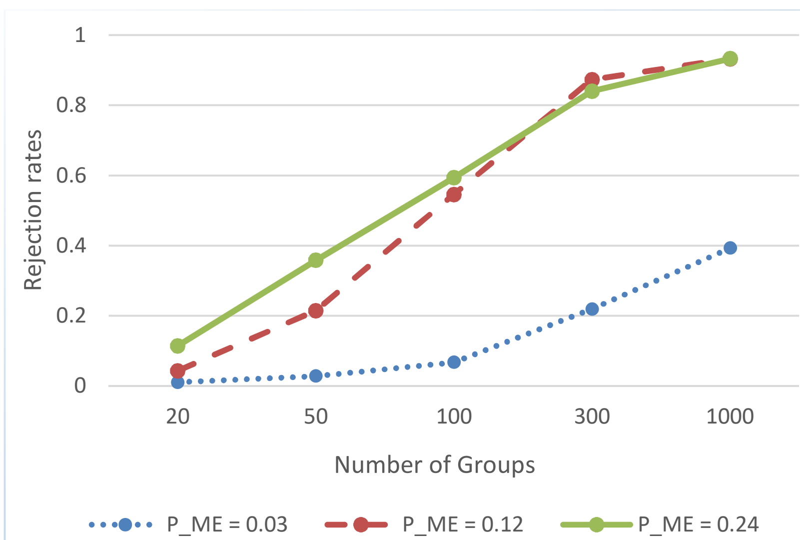

Figure 9. The Interaction between Group Number and Effect Size on Power for the MSEM Model

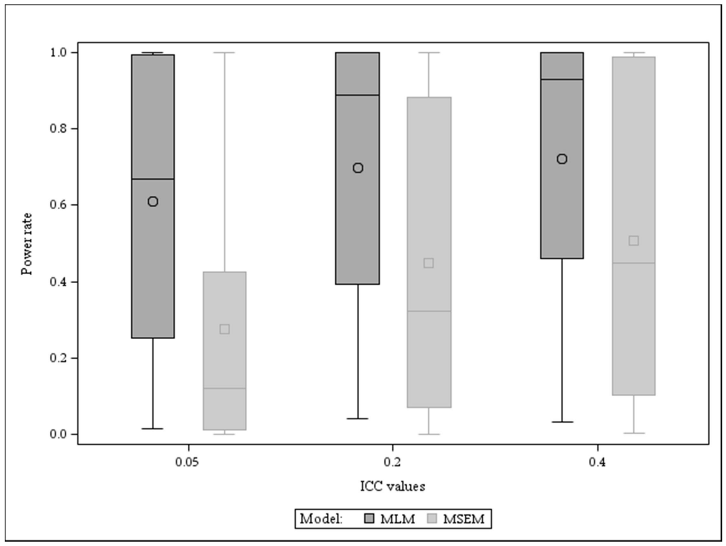

Figure 10. Statistical Power of MLM and MSEM models by ICC values

The magnitude of ICC was found to significantly associate with power of the MSEM model (η2 of 0.065). Figure 10 presents the distributions of mean estimated power across ICC levels for the MSEM model. The distributions of mean estimated power for MLM model were also included in this figure for the comparison purpose. The MSEM model had the smallest power at the ICC level of 0.05 (mean of 0.275 and standard deviation of 0.349). Power went up as the ICC values increased with power mean of 0.448 and 0.508 for ICC level of .2 and .4, respectively. The variances of estimated power were also higher at the higher ICC levels as 0.387 and 0.386 for ICC of 0.2 and 0.4, respectively. It was seen that two methods had the opposite patterns in term of variability. The explanation for this observation was that having some high power conditions mixing with low power conditions made the MLM model having high variability in term of power at the low ICC level. When the ICC level increased, the power increased and the variability of power decreased. Regarding the MSEM model, the variability of power at the low ICC level was small since most of the conditions had low power. The increasing in the ICC level made the power of some conditions larger but some other ones still remained low hence the variability of power for the MSEM model got bigger.

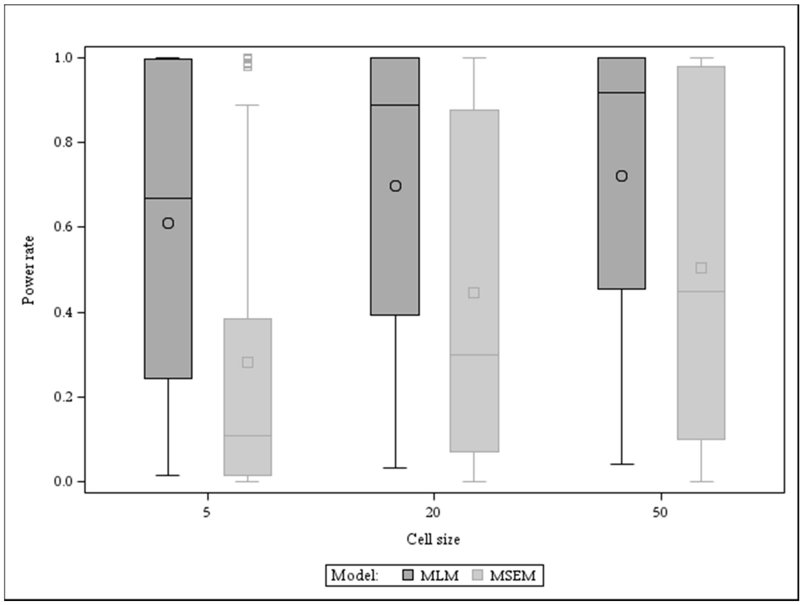

Figure 11. Statistical Power of MLM and MSEM Models by Cell Size

Cell size is another factor that has a significant impact on statistical power of MSEM model (η^2 = 0.061). As expected, power was highest with the biggest cell size (n = 50). It was also noted that power decreased as cell size decreased. However, the standard deviation of power decreased when the cell size decreased. Figure 11 shows the overall distributions of estimated power by cell size for the MSEM model. The distribution of power by cell size for the MLM model is also included in Figure 11.

Table 5 summarizes the variation of statistical power of the MLM and MSEM models across four simulation factors: effect size, group number, cell size, and ICC value. This table shows that the MLM model had the reasonable power event with the small effect size and small ICC value but with the group number of 300 or higher. With the higher effect size or the higher ICC value, the MLM model can have high power (80% or higher) with some conditions in which the group number was 100 or 50. Regarding the MSEM model, the power was very low for most small effect size conditions. There were only a few conditions with power of around 60% with small effect size where the ICC level was high and the sample size was very large. The power was also very low for all small sample size conditions event thought with the high level of effect size and ICC value. On the other hand, the MSEM model was found to have high power with medium and large effect size conditions when the group number of 300 or higher. With the medium effect size (effect size of 0.12), the MSEM model had power closed to 1 for most conditions with the group number of 300 and 1000, except ones with small ICC value (of 0.05). The similar pattern was observed with the high effect size conditions. Additionally, for the high effect size conditions, the MSEM model also had very high power (of 90%) in some conditions with the group number of 100 and high ICC value of 0.4.

Chapter Summary

This chapter provided a description of the results of the study. The MSEM method had the perfect convergence rates for all examined conditions. Regarding the MLM model, the convergence rate was perfect when the model was estimated without the constraint in the variances. With the constraint of variances being greater than zero, the MLM model had high convergence rate with most conditions but some with small sample size and low ICC. The MLM model had a good control of Type I error rates with large sample sizes while the MSEM model was very conservative in controlling for Type I error rates. The MLM model was found to have better power than the MSEM model in most conditions, especially the small sample size conditions. Regarding the relationship between the simulated factors and the performance of the MLM and MSEM models, the sample size, the effect size, and ICC values emerged as the most significant factors.

Table 5

Variation of power across group number, cell size, effect size, and ICC value for MLM and MSEM models

			MLM model			MSEM model		
Effect size	Group number	Cell size	ICC = 0.05	ICC = 0.20	ICC = 0.40	ICC = 0.05	ICC = 0.20	ICC = 0.40
	20	5	0.02467	0.05133	0.05667	0	0.00467	0.01200
	20	20	0.03533	0.05733	0.04200	0.00333	0.01200	0.00533
	20	50	0.04933	0.05733	0.04533	0.00800	0.02133	0.02000
	50	5	0.11467	0.20867	0.24000	0.00133	0.01467	0.02800
	50	20	0.19000	0.23800	0.26467	0.01333	0.02800	0.04933
	50	50	0.25533	0.27733	0.25267	0.02400	0.04000	0.04533
	100	5	0.26133	0.51467	0.61067	0.00533	0.03867	0.07200
0.03	100	20	0.52533	0.66200	0.64400	0.03067	0.07733	0.09333
	100	50	0.63667	0.67600	0.69533	0.04800	0.10533	0.12667
	300	5	0.66267	0.96333	0.99667	0.02933	0.12933	0.20000
	300	20	0.96800	0.99733	0.99867	0.12267	0.26933	0.30400
	300	50	0.99600	0.99667	0.99867	0.20933	0.34000	0.36133
	1000	5	0.89933	1	1	0.05333	0.24000	0.36000
	1000	20	1	1	1	0.21733	0.47867	0.58133
	1000	50	1	1	1	0.40133	0.57333	0.62800
	20	5	0.03200	0.08467	0.12600	0	0.00933	0.03467
	20	20	0.08867	0.14667	0.19267	0.00533	0.06533	0.07267
	20	50	0.10067	0.17467	0.19800	0.02400	0.07467	0.09267
	50	5	0.13667	0.31467	0.45000	0.00133	0.06667	0.15000
	50	20	0.33267	0.50800	0.55933	0.05067	0.29600	0.37867
	50	50	0.45333	0.55400	0.57400	0.17933	0.35733	0.44667
	100	5	0.29533	0.63000	0.75000	0.00867	0.26533	0.59333
0.12	100	20	0.60667	0.79667	0.85733	0.31200	0.71733	0.81333
	100	50	0.75600	0.83667	0.85133	0.56800	0.81333	0.81600
	300	5	0.64533	0.96867	0.99667	0.12667	0.87867	1
	300	20	0.96667	0.99733	0.99933	0.86933	0.99733	1
	300	50	0.99267	1	0.99867	0.98933	0.99867	0.99867
	1000	5	0.90867	0.99933	1	0.39333	0.99867	1
	1000	20	1	1	1	1	1	1
	1000	50	1	1	1	1	1	1
	20	5	0.07733	0.26467	0.39533	0	0.02800	0.09600
	20	20	0.27267	0.43200	0.48067	0.01200	0.13600	0.21200
	20	50	0.37933	0.46067	0.47467	0.08400	0.22267	0.23067
	50	5	0.39867	0.78133	0.88333	0.00400	0.14533	0.36133
	50	20	0.80867	0.89067	0.92800	0.15467	0.43467	0.59200
	50	50	0.88400	0.92000	0.94133	0.30533	0.57467	0.64933
	100	5	0.78733	0.98333	0.99467	0.02933	0.32800	0.64800
0.24	100	20	0.98600	0.99667	0.99667	0.27867	0.76933	0.90133
	100	50	0.99133	0.99800	0.99933	0.58800	0.88667	0.90933
	300	5	0.99733	1	1	0.14267	0.74400	0.98400
	300	20	1	1	1	0.72267	0.99867	1
	300	50	1	1	1	0.97200	1	1
	1000	5	1	1	1	0.39867	0.99867	1
	1000	20	1	1	1	1	1	1
	1000	50	1	1	1	1	1	1

CHAPTER FIVE: DISCUSSION

This chapter outlines the summary of study and presents the answers for research questions. The discussion and conclusions were drawn based on results. Limitations of this study were also discussed.

Summary of the Study

Purpose

The purpose of this study was to examine the performance of the MSEM approach in the multilevel mediation analysis with non-normal data. This study also compared the performance of MSEM method with that of MLM methods. The performance criteria considered were Type I error rates, statistical power, and convergence rate.

This study focused on the multilevel mediation model named the upper level mediation model by Pituch and Stapleton (2008) that had the independent variable measured at level 2, the mediator and the dependent variable measured at level 1. This mediation model was also labeled as the 2-1-1 model by the notation convention of Krull and McKinnon (2001). In this study, the upper level mediation model was constrained with fixed effect and continuous variables only.

Research questions

1. What are the differences between the MSEM approach and the MLM approach in terms of Type I error control, statistical power, and convergence rate with non-normal data?

2. How do the simulation design factors, including the magnitude of the population indirect effect, the population distribution shape, sample size at level 1 and level 2, and the ICC level, affect the performance of the MSEM approach and the MLM approach?

Methods

The study used Monte Carlo simulation method to examine the performance of the MLM and MSEM approaches. The simulated data factors included: (1) number of groups (k = 20, 50, 100, 300, and 1000); (2) cell size (average n_j = 5, 20, and 50); (3) ICC of level-1 variables (small = 0.03, medium = .2, and large = 0.4); (4) between indirect effect (null condition with effect = 0, non-null conditions with following effects: small = 0.03, medium = 0.12, and large = 0.24); (5) degree of non-normality for level-1 variables (normal data where skewness = 0 and kurtosis = 0, moderately non-normal where skewness = 1.63, kurtosis = 4, and severely non-normal where skewness = 2.28, kurtosis = 12). The combination of these factors created 540 different simulation conditions.

The data in this study were generated by two steps using PROC IML in SAS (version 9.4). The level-2 variable and group means of level-1 variables were generated in step 1. Group means of level-1 variables in step 1 were then used to generate level-1 variables for each group in step 2. 1000 samples were generated for each simulation condition. For each sample, the MLM and MSEM mediation models were estimated which yielded performance criteria of rejection status and convergence status. The performance the MLM and MSEM models on each condition was summarized from estimates of generated samples.

The simulation results were analyzed using PROC GLM in SAS 9.4 where dependent variables were Type I error rate, statistical power, and convergence rate while the independent variables were five simulation factors.

Research question 1: What are the differences between the MSEM approach and the MLM approach in terms of Type I error control, statistical power, and convergence rate with non-normal data?

This study aimed to compare the performance of the MSEM and MLM models across various outcome variables. The first performance variable was the convergence rates of both models. The convergence rates were excellent for both MSEM and MLM models. The MSEM model was estimated using Mplus 7.1 and the result showed 100% of convergence rates for all conditions. The MLM model was estimated using SAS 9.4. Following the suggestion of Pituch & Stapleton (2008) to remove the constraint of non-negative variances, the results also showed 100% of convergence rate for all conditions. When the constraint was not removed, the convergence rates were significantly lower for the small sample size conditions.

The second outcome variable was the Type I error rate of both models. The MSEM model showed a very conservative performance in term of controlling for the Type I error with the rejection rates of null conditions were zero or closed to zero across all conditions. The Type I error rates of the MLM model were found to be lower than the expected alpha level at the smallest sample size conditions but the rejection rate increased as the sample size increased in either the group number or the cell size. The Type I error rates of MLM model reached to the expected alpha level as the group numbers were 300 or higher.

Regarding the statistical power, the MLM model had the higher means power across all simulated conditions than the MSEM model. The MLM model also had higher power than the MSEM model for each level of the number of groups. While the MLM model had an acceptable power at the small group number (group number of 50), the MSEM model only had enough power at the large number of group (group number of 300).

Research question 2: How do the simulation design factors affect the performance of the MSEM approach and the MLM approach?

The performance of the MSEM and MLM models was evaluated based on several criteria including Type I error rate control, power, and convergence rate. The associations among the simulated factors as well as their first-order interactions with the performance criteria were examined to answer this research question.

The number of groups was found to have a significant effect to many performance criteria in both the MSEM and MLM models including Type I error rates of the MLM model, and statistical power of both the MLM and MSEM models. In general, the higher number of groups resulted in the better performance of the associated criteria. Type I error rates of the MLM model were closed to zero when the group number equals to 20 but got closed to the expected alpha level of .05 with the greater group numbers. Similarly, the conditions with the small group number of 20 had very low power for both MSEM and MLM models. Adding more groups raised power for both models. However, power of the MLM model increased much faster than that of the MSEM model. While the MLM model had a reasonable power with the group number of 300, the MSEM model still had conditions with low power even with group number of 1000, specifically the conditions with the small effect size and the low ICC value.

Mediation effect sizes were associated with power in the MLM model as well as power in the MSEM model. When the effect size was small, power was very low for most conditions in the MSEM model, and then the variation of power among these conditions was small. When the effect size increased, the average power increased. However, the variation of power among conditions also went up. The smallest average power in the MLM model was found with the mediation effect of 0.03. Average power increased as the effect size increased. On the other hand, the variation of power for the MLM model decreased at the higher levels of effect size since many conditions had a perfect rejection rate, especially at the highest effect size level of 0.24.

The interaction between the group number and the effect size affected the estimate of power for both the MSEM and MLM models. Regarding the MSEM model, the power was very low at the small group number for all three effect size levels. As the group number went up the power increased for all effect size levels. However, the change was much smaller for the small effect size group while the medium and large effect size groups had similar power across the group numbers, especially the large group numbers. Regarding the MLM model, power was higher as the effect size was larger. It was evidenced that power was significantly different across the effect size levels as the group number was small. The power became similar across effect size levels as the number of group reached to 300 and 1000.

ICC also had an impact on power of the MSEM model. When the ICC value was small, the model tended to have less power in detecting the mediation effect. As the ICC increased, the estimated power increased and also had a smaller variation.

Analysis results showed a significant relationship between cell size and statistical power of the MLM model. Smaller average power estimates were seen with the small level-1 sample size. Increasing the cell size generated a larger power for the MSEM model. There was also an evidence of decrease in the variation of power estimates as the cell size increased.

Discussion and Conclusion

Preacher et al. (2010) proposed using the MSEM model to test the mediation effect in multilevel data as an alternative of the MLM model. It was shown that the MSEM model was more flexible in some testing situation where the MLM model cannot be employed. This study extended the Preacher et al. (2011)'s work by exploring the performance of both methods under various conditions of non-normality. The bootstrapping method was applied in order to overcome the non-normal distribution of the mediation effect.

There were two main goals for this study. The first was to examine the performance of the MSEM model under the conditions of non-normal data. The effect of non-normality on the estimates of the MLM model was also considered. The results of this study showed no significant effect of the degree of non-normality on any performance criteria of either MSEM or MLM models. This result was similar to the finding of Maas and Hox (2004) as they found the violation of non-normality has little or no effect on the fixed effect estimates of MLM models. Byrd (2008) and Zhang (2005) also did not report the association between non-normal data and the estimates of MLM and MSEM model. As pointed out in chapter two, there were several studies examined the association between the non-normality and the performance of either the MLM model or the MSEM model. Among those studies, Pituch and Stapleton (2008)'s paper was the one that focused on multilevel mediation analysis but it examined the MLM model only. Pituch and Stapleton (2008) reported that Type I error rates elevated as departures from normality increased. However, examining more detail from their results show that the inflated Type I error rate only happened with null conditions that had either parameter *a* or *b* being different from zero. The null conditions with both *a* and *b* coefficients equal to zero were conservative in controlling for Type I error rates. This outcome was similar to the result in this study since Pituch and Stapleton (2008)'s conclusion was based on the group number of 20 and 40, which were only considered the small group number in this study.

The second main goal of this study was to examine the performance of the MSEM method and the MLM method, specifically the MLM method proposed by MacKinnon (2008) and Zhang et al. (2009), under various conditions. It was surprising that the MSEM model was conservative in controlling the type I error rate even with the largest sample. In contrast, the MLM model showed a good control in Type I error rates when the sample size was large enough (group number of 300) while the Type I error rates were lower than the expected alpha level for smaller sample size conditions.

It was evident that the MSEM model was outperformed by the MLM model in term of power for most simulated conditions. Simulation results showed that the MLM model required a smaller sample size to reach a reasonable power level compared to the MSEM model. This result was consistent with the finding of the previous study comparing the MSEM and the MLM models in the multilevel mediation analysis (Preacher et al., 2011). Preacher at el. (2011) however, only compared the MLM model and MSEM model on normal data and found that the MLM model had adequate power at the reasonably small sample sizes while the MSEM model had small power at the small sample sizes.

This study also confirmed the effect of sample size on power for both MLM and MSEM models which was reported by Pituch and Stapleton (2008) and Preacher et al. (2011). The association between ICC and power for the MSEM model reported by Preacher et al. (2011) was also observed here in the current study.

Among the simulation factors examined in this dissertation, the mediation effect size emerged as the most important one. Effect size was seen to be highly associated with each of the considered performance criteria. This study also support the finding of previous studies (Preacher et al., 2011; Zhang, 2005) about the relationship between sample size, especially the number of group, and the performance of either the MLM or MSEM models.

For researchers conducting mediation analysis with multilevel data, results of this study provided some implications. First, this study focused on the performance of mediation tests under non-normal data conditions. It provided the evidence that the bootstrapping method minimizes the effect of non-normality on the performance of the testing methods; even though the study only considered the non-normality in the level 1 of multilevel data. Researchers conducting a multilevel mediation analysis hence might consider applying the bootstrapping method in order to reduce the impact of the violation of normality assumption. Second, this study aimed to compare the performance of the MLM and MSEM approach in testing the mediation effect. Emphasizing on the null hypothesis test, results of this study showed that the MLM approach might be the better option than the MSEM approach since the MLM approach provided better control of Type I error rates and also had higher statistical power in most examined conditions.

Limitations of Study

Given the research design, there were some limitations in this study. The study investigated only the 2-1-1 mediation model with fixed effects among many multilevel mediation models. In addition, the variables investigated here were all continuous variables while binary variables were also common in educational and social sciences. Another limitation of this study was that the MSEM model and the MLM model were estimated in different statistical software. Even though the same estimation method was used for both models it still might have some unintended differences due to the software programs but not the methods. The non-normality conditions were examined in this study only happened at the level-1 data while it is possible to have the data that is non-normal in either level-2 or both levels. Therefore, the impact of non-normality on model estimates was not fully explored in this study. The null conditions considered here in this study was only the special case with both coefficients *a* and *b* of a mediation effect equal to zero. There are possible null conditions with either *a* or *b* equal to zero that were not examined in this study. Therefore, the conclusion in this study related to the type I control of both MLM and MSEM models was limited. Finally, the results of this study is limited to the indirect effect which was calculated by the product of coefficients method but not the difference of coefficients method and tested by bootstrapping method but not the common Sobel's test.

Comparing the MLM model and the MSEM model but only focusing only on the null hypothesis test is another limitation of the present study. It would be more informative if the performance of the MSEM model and the MLM model in term of accuracy and precision was fully explored. The accuracy and precision normally are examined by calculating statistical bias and standard error for the point estimate or the confidence interval width and confidence interval coverage for confidence interval. However, only relative bias and confidence interval width were examined in this study. The performance of the MLM and MSEM model in terms of relative bias and confidence interval width were presented in Appendix A and Appendix B.

REFERENCES

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology,* 51, 1173-1182.

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics,* 28, 135-167.

Bauer, D. J., Preacher J. K., & Gil, M.K. (2006). Conceptualizing and Testing Random Indirect Effects and Moderated Mediation in Multilevel Models: New Procedures and Recommendations. *Psychological Methods,* Vol. 11, No. 2, 142–163, DOI: 10.1037/1082-989X.11.2.142

Biesanz, J. C., Falk, C. F., & Savalei, V. (2010). Assessing mediational models: Testing and interval estimation for indirect effects. *Multivariate Behavioral Research,* 45:4, 661-701, DOI: 10.1080/00273171.2010.498292.

Bollen, K. A. (1989), *Structural Equations with Latent Variables*, New York: John Wiley & Sons.

Byrd, J. K. (2008). *A Monte Carlo investigation of three different estimation methods in multilevel structural equation modeling under conditions of data nonnormality and varied sample sizes* (Doctoral Dissertation). Retrieved from PsycINFO (Accession No. 2009-99150-535).

Carpenter, J. M., Goldstein, H., & Rasbash, J. (2003). A novel bootstrap procedure for assessing the relationship between class size and achievement. *Applied Statistics,* 52, 431–443.

Clogg, C. C., Petkova, E., & Shihadeh, E.S (1992). Statistical methods for analyzing collapsibility in regression models. *Journal of Educational Statistics*, 17(1), 51-74.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New York: Lawrence Erlbaum Associates.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research,* 38, 529-569.

Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling: A Multidisciplinary Journal,* 4:2, 87-107, DOI: 10.1080/10705519709540063

Fleishman, I. A. (1978). A method for simulating non-normal distributions. *Psychometrika,* Vol. 43, No 4, 521-532.

Freedman, L. S., & Schatzkin, A. (1992). Sample size for studying intermediate endpoints within intervention trials of observational studies. *American Journal of Epidemiology,* 136, 1148-1159.

Hayes, F.A. (2009). Beyond Baron and Kenny: statistical mediation analysis in the new millennium. *Communication Monographs,* 76:4, 408-420.

Hox, J. J., & Bechger, T. M. (1998). An introduction to structural equation modeling. *Family Science Review,* 11, 354-373.

Hox, J. J. (2002). *Multilevel Analysis: Techniques and Applications.* Mahwah, NJ: Erlbaum.

Hox, J.J., Maas, M. C. Jr., & Brinkhuis, S. M. Jr. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica,* Vol. 64, nr. 2, 157–170, DOI:10.1111/j.1467-9574.2009.00445.x.

Hox, J. J. (2013). Multilevel regression and multilevel structural equation modeling. Pp 281-294 in T. D. Little (Ed). *The Oxford Handbook of Quantitative Methods*. *Vol. 2*. Oxford: Oxford University Press.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review,* 5:602–619.

Kenny, A. D., Korchmaros D. J., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods,* Vol. 8, No. 2, 115–12, DOI: 10.1037/1082-989X.8.2.115.

Krull, J. L. & MacKinnon, D. P. (1999*).* Multilevel mediation modeling in group-based intervention studies. *Evaluation review,* Vol. 23 No. 4, August 1999, 418-444.

Krull, J. L. & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, 36 (2), 249-277.

Loukas, A., Suzuki, R., & Horton, K. D. (2006). Examining school connectedness as a mediator of school climate effects. *Journal of Research on Adolescence*, 16, 491-502.

Ludtke, O., Marsh H.W., Robitzsch, R., 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,* Vol. 13, No. 3, 203–229.

MacKinnon, P. D. (2008). *Introduction to statistical mediation analysis.* New York: Lawrence Erlbaum Associates.

MacKinnon, P. D. & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, 1993 17:144. DOI: 10.1177/0193841X9301700202

MacKinnon, P. D., Warsi, G., & Dwyer, H. J. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research,* 30(1), 41-62.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods,* 7, 83-104.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research,* 39, 99-128.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika,* 57, 519–530.

Maas, C. & Hox, J. J. (2004). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics and Data Analysis*, 46, 427-440.

Matsumura, L. C., Garnier, H. E., & Spybrook, J. (2013). Literacy coaching, reading comprehension instruction and student achievement: A multi-level mediation model. *Learning and Instruction*, 25, 35-48.

Meeker, W.Q., Jr; Cornwell, L.W., Aroian, L.A. (1981). *Selected tables in mathematical statistics, volume VII: The product of two normally distributed random variables.* Providence, RI: American Mathematical Society.

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

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin,* Vol. 105. No.1, 156-166.

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

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research,* 22, 376-398. (Eds.). *Longitudinal data analysis* (pp. 143-165). Boca Raton, FL: Chapman & Hall/CRC.

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology,* 38, 171-189.

Muthén, B., & Satorra, A. (1989). Multilevel aspects of varying parameters in structural models. In R. D. Bock (Ed.). *Multilevel analysis of educational data* (pp. 87-99). San Diego: Academic.

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology,* 1997, 25, 267-316.

Muthén, B. O., & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs

Pituch K.A., Whittaker T.A. & Stapleton LL.M. (2005). A comparison of methods to test for mediation in multisite experiments. *Multivariate Behavioral Research,* 40:1, 1-23, DOI: 10.1207/s15327906mbr4001_1.

Pituch K.A., Stapleton L. M. & Kang J. W. (2006). A comparison of single sample and bootstrap methods to assess mediation in cluster randomized trials. *Multivariate Behavioral Research,* 41:3, 367-400, DOI: 10.1207/ s15327906mbr4103_5.

Pituch, K.A., & Stapleton, L. M. (2008). The performance of methods to test upper-level mediation in the presence of nonnormal data. *Multivariate Behavioral Research,* 43:2, 237-267, DOI: 10.1080/00273170802034844.

Pituch, K. A., Murphy, D. L. & Tate, R. L. (2009). Three-level models for indirect effects in school- and class-randomized experiments in education. *The Journal of Experimental Education,* 78:1, 60-95.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010): A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods,* Vol. 15, No. 3, 209–233

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: the advantages of multilevel sem. *Structural Equation Modeling: A Multidisciplinary Journal,* 18:2, 161-182

Quiroga, C. V., Janosz, M., Bisset, S., & Morin, A. J. S. (2013). Early adolescent depression symptoms and school dropout: Mediating processes involving self-reported academic competence and achievement. *Journal of Educational Psychology*, Vol. 105, No. 2, 552–560

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods, Second Edition.* Newbury Park, CA: Sage.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In: Leinhardt, S., editor. *Sociological methodology*. American Sociological Association; Washington, DC (p290-312).

Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. *Sociological methodology*, *16*, 159-186.

Stone, C.A., & Sobel, M. E. (1990). The robustness of estimates of total indirect effects in covariance structure models estimated by maximum likelihood. *Psychometrika,* 55, 337-352.

Tofighi, D. (2010). *Multilevel Mediation Analysis: Statistical Assumptions and Centering* (Doctoral Dissertation). Retrieved from PsycINFO (Accession No. 2011-99160-418).

Vale, D. C. & Maurelli, V.A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika,* Vol. 48, No. 3.

Wang, J., Carpenter, J. R., & Kepler, M. A. (2006). Using SAS to conduct nonparametric residual bootstrap multilevel modeling with a small number of groups. *Computer methods and programs in biomedicine,* Vol. 82, 130–143.

Zhang, D. (2005). *A Monte Carlo investigation of robustness to nonnormal incomplete data of multilevel modeling* (Doctoral Dissertation). Retrieved from PsycINFO (Accession No. 2007-99003-005).

Zhang, Z., Zyphur, M. J., & Preacher K. J., 2009. Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods.* Vol. 12, No. 4, 695-719.

Zu, J. (2009). *Robust procedures for mediation analysis* (Doctoral Dissertation). Retrieved from PsycINFO (Accession No. 2010-99220-471).

APPENDIX A: PERFORMANCE OF THE MLM MODEL AND THE MSEM MODEL IN TERM OF RELATIVE BIAS

Bias is the difference between the average obtained values and the true population values. It can be positive or negative. In this simulation study, the relative percentage bias was use to assess the performance of bias. The relative percentage bias of the mediation effect was calculated by the difference between estimated mediation effect and the true population value divided by the true population value. The relative bias was calculated for each simulation condition and equal to the average relative bias of all replications on that condition. The formula of relative percentage bias is presented in the equation (22).

$$Relative\ Bias = 100\left[\left(\hat{a}_B\hat{b}_B - a_B b_B\right)/a_B b_B\right]\% \quad (22)$$

where $\hat{a}_B$ and a_B are the estimated and population coefficients of the between effect of X on M, respectively while $\hat{b}_B$ and b_B are the estimated and population coefficients of the between effect of M on Y in a 2-1-1 mediation model.

Simulation results

Statistical relative bias of the MLM and MSEM mediation models was examined using 405 non-null conditions which vary systematically in the mediation effect size, sample size, ICC value, and the degree of non-normality. This study used the relative percentage bias to assess the performance of models on bias. The relative bias was calculated as the difference between the average obtained value and the true population values divided by the population value of indirect effect.

The distributions of relative bias by each mediation model were illustrated in Figure 12. Estimated relative bias was much close to zero for many simulated conditions in the MSEM model than in the MLM model. The MLM model had an overall mean of relative bias of -0.551% (SD = 29.203). The range of bias for the MLM model is from -44.66% to 128.12%. The MSEM model had an overall mean of relative bias of 0.310% (SD = 20.915) with range from -103.86% to 145.97%.

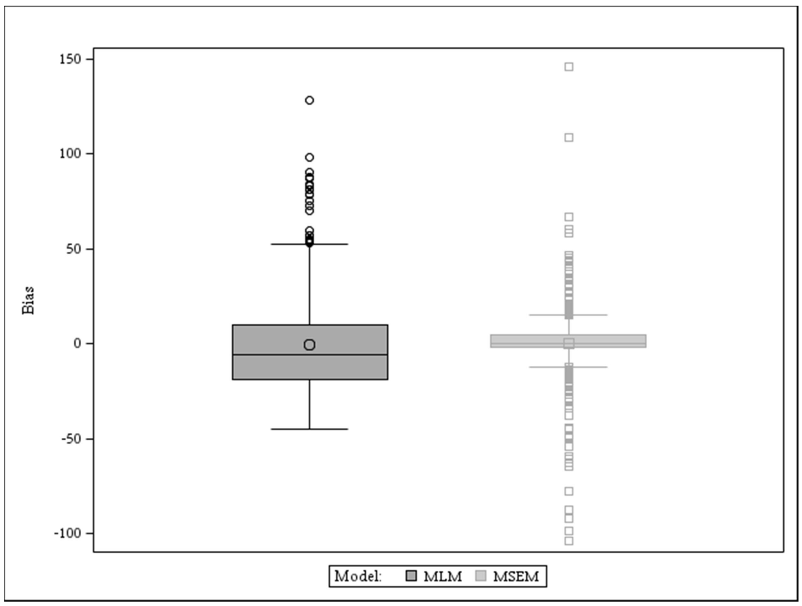

Figure 12. Distribution of Relative Bias by Models

The eta-squared analyses were conducted for all simulation factors (number of groups, group size, effect size, degrees of non-normality, and ICC values) and possible two-way interactions. The eta-squared values were examined and those values exceeding the Cohen's (1988) medium effect size criteria of 0.0588 were further explored. Results of eta-squared analyses were summarized in Table 6. No main factor or interaction was found to have significant effect on the bias for the MSEM model. However, effect size (η^2=0.583) emerged as a factor highly related to estimated bias for the MLM model. The interaction between effect size and ICC (η^2=0.194) and the interaction between effect size and cell size (η^2=0.166) were also associated with the variability of bias in the MLM model.

Table 6
Eta-squared Analyses of Relative Bias for MLM and MSEM Models

Factors	MLM Model	MSEM Model
ME	**0.58269**	0.01435
ME* ICC	**0.19436**	0.00617
ME* size1	**0.16571**	0.01181
size2* size1	0.00259	0.01398
ME* size2	0.00110	0.02517
size1* shape	0.00073	0.00593
size1	0.00058	0.00073
size1* ICC	0.00055	0.00345
ICC* shape	0.00045	0.01194
size2* shape	0.00042	0.02478
size2* ICC	0.00032	0.01189
shape	0.00019	0.00023
ME* shape	0.00010	0.00455
size2	0.00008	0.00318
ICC	0.00002	0.00512

Note. The effect sizes with significant main effect and interaction in relative bias appear in bold. size2 = number of group; size1 = group size, ME = effect size; ICC = ICC value; shape = population shape.

The distributions of bias across the indirect effect levels for both the MLM and MSEM models were shown in Figure 13. In comparison between the two models, the MSEM model had smaller bias than the MLM model in each mediation effect level. However the variance of the distribution of relative bias for the MSEM model was slightly higher than that of the MLM model. The variances of the relative bias of the MSEM model were 29.7, 17.7, and 10.0 while that of the MLM model were 27.4, 12.3, and 12.9 for the effect size level of 0.03, 0.12, and 0.24, respectively. Relative bias was observed to be largest when the indirect effect equals to 0.03 and became smaller when the indirect effect got larger (equals to 0.12 and 0.24) for both models. The variance of bias distribution, on the other hand, was inversely related to the effect size.

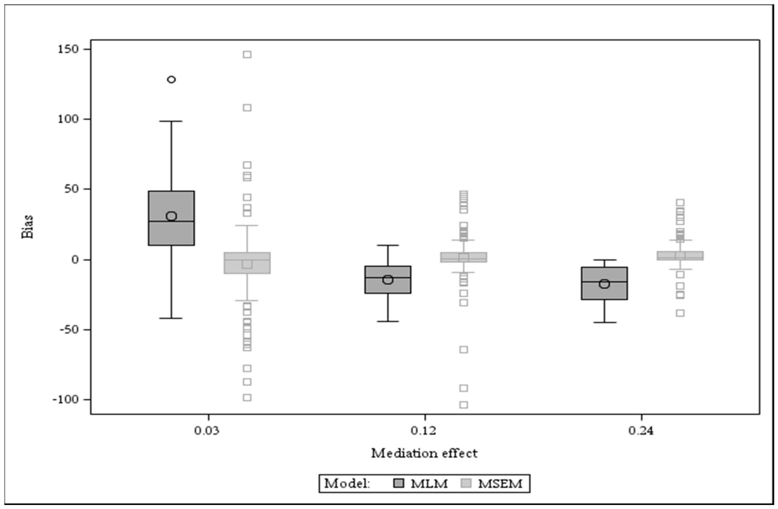

Figure 13. Distribution of Relative Bias by Indirect Effect for MLM and MSEM Models

Results of the eta-squared analyses also present the significant effect of the interaction between effect size and ICC value on the bias for the MLM model (η^2=0.194). Figure 14 displays the average of relative bias for the MLM model under different combinations of the indirect effect and ICC values. The difference in relative bias among different ICC levels was large when the effect size was small. But the difference became smaller when the effect was medium or large. Figure 14 also shows the larger ICC values having the smaller relative bias at every level of effect size.

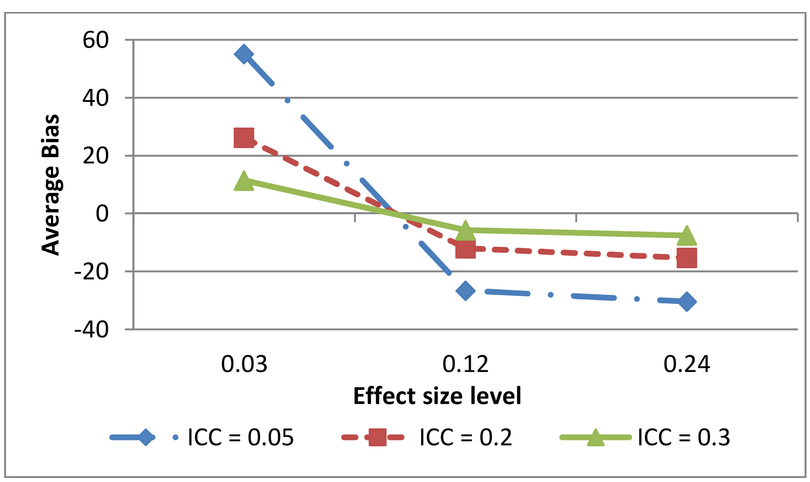

Figure 14. Interaction between Effect Size and ICC on Relative Bias for MLM Model

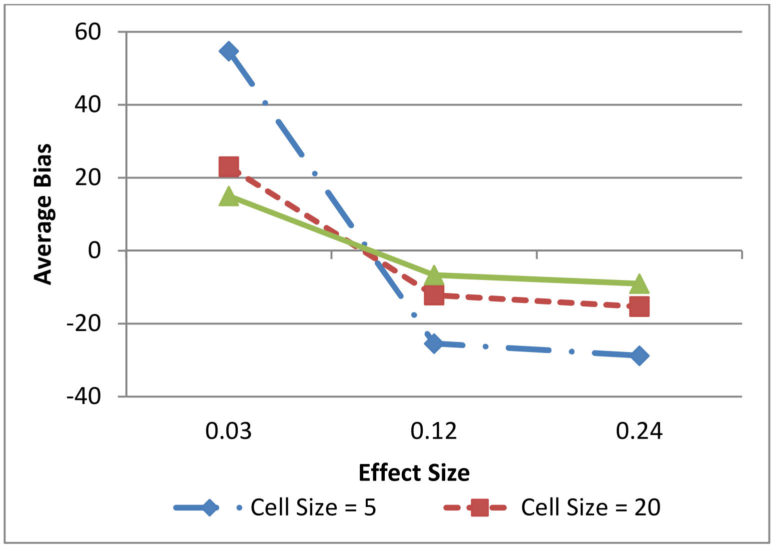

Figure 15. Interaction between Effect Size and Cell Size on Relative Bias for MLM Model

The interaction between cell size and effect size was also found to have a significant effect on bias for the MLM model ($\eta 2$=0.166). The relationship between effect size, cell size and the estimated bias of the MLM model was presented in Figure 15. This figure shows negative relationship between cell size and bias when the larger cell size had the smaller bias in each level of effect size. The difference in bias among cell size levels was large with the small effect size while became smaller with medium and large effect sizes.

Table 7
Variation of relative bias across effect sizes, cell sizes, and ICC values

Effect size	Cell Size	ICC = 0.05	ICC = 0.2	ICC = 0.4
	5	74.5293	49.7758	29.8070
0.03	20	47.9077	19.9692	1.1419
	50	32.8062	8.7983	3.2690
	5	-39.3481	-23.4245	-13.6519
0.12	20	-25.3573	-8.8033	-2.4247
	50	-15.2480	-3.6729	-1.1719
	5	-42.1429	-28.5483	-15.7854
0.24	20	-29.7282	-11.8024	-4.5555
	50	-19.3848	-5.5701	-2.2571

The variation of relative bias across cell size, effect size, and the ICC value was summarized in Table 7. The relative bias was high with all small ICC value conditions. The small cell size conditions were also found to have large relative bias also. On the other hand, the relative bias was small for the conditions that have a medium or large cell size and large ICC value.

Summary

As described in the previous sections, the MSEM model outperformed the MLM model in term of relative bias. Examining across the mediation effect levels showed that relative bias of the MSEM model were small and centered on the zero point across all effect size levels. The MLM model had positive relative bias for the small effect size conditions and negative relative bias for medium and large effect size conditions. Even thought, relative bias decreased for the MLM model when effect size increased, the average relative bias of the MLM model still significantly higher than that of the MSEM model. The MSEM model had the similar bias across all simulated conditions and its bias was significantly smaller than the estimated bias of the MLM model, especially in small sample-size conditions. This result was consistent with previous studies in calculating bias for the MSEM model. Byrd (2008) found that the coefficient estimated in the MSEM model had little or no bias when examining the performance of the MSEM model under various non-normal data conditions. Preacher et al. (2011) also showed that the MSEM model had much smaller bias than the MLM model for all examined conditions. In addition, the average relative bias across sample sizes (both the group number and cell size) was found to be similar to the estimated relative bias in the Preacher et al. (2011)'s study.

Among the simulated factors, the interaction between effect size and ICC emerged as a factor associated with relative bias in the MLM model. In general, conditions with the larger ICC had the smaller bias. The difference in bias among ICC levels however depends on the level of effect size. The bias difference was the greatest at the smallest effect size level and became smaller at the medium and large effect sizes. The interaction between effect size and cell size was also related to bias of the MLM model. The small cell size conditions had the biggest bias while the difference in bias between two groups of medium and large cell size was small. As effect size increased, bias of the small cell size group got closed to the bias of medium and large cell size groups.

APPENDIX B: PERFORMANCE OF MLM MODEL AND MSEM MODEL IN TERM OF CONFIDENCE INTERVAL WIDTH

Confidence interval (CI) width is the difference between the upper limit and lower limit of the confidence interval. The bootstrapping procedure draws 1000 bootstrap samples from a simulated data. The indirect effect was calculated for each bootstrap sample using either MLM or MSEM methods. The estimated indirect effects from these bootstrap samples were used to approximate a distribution of indirect effect for each generated sample. The upper limit of 95% confidence interval was defined as the value at 97.5th percentile while the lower limit was defined as the value at 2.5th percentile. A confidence interval width was then calculated as the difference between upper limit and lower limit. It was calculated for each replication and hence, the CI width for each simulation condition was the average CI width of its replications.

Simulation results

Confidence interval (CI) width is used to measure the precision of a testing method. A test is more precise if CI width is smaller. In this study, CI width was calculated by the difference between upper limit and lower limit of 95% confidence interval. The CI width was calculated for each simulated sample. Hence, the CI width for a condition is the average of CI width from all of its replications.

CI widths of both the MSEM and MLM models were examined using 405 non-null conditions. The distributions of CI width for each model are illustrated in Figure 16. The MLM model had an average of CI width of 0.352 (SD= 0.27) and a range of CI width from 0.069 to

1.373 while the MSEM model had a larger CI width with mean of 0.689 (SD= 0.828) and a range of CI width from 0.054 to 4.807.

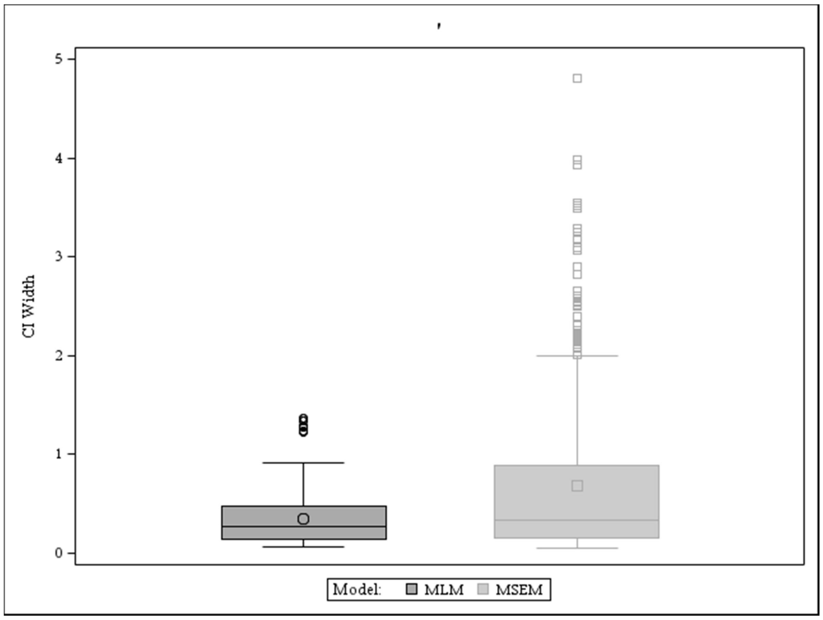

Figure 16. Confidence Interval Width by Models

The variation of confidence interval width was further examined using eta-squared analyses with all simulated factors and first order interactions. Analyses were conducted separately between the MLM and MSEM models. Results from eta-squared analyses were summarized in Table 8. The number of group was found to have significant impact on CI width for the MLM model with $\eta^2 = .835$. Regarding the MSEM model, several factors impacted the confidence interval width: the number of group ($\eta^2 = .431$), the ICC values ($\eta^2 = .145$), the cell size ($\eta^2 = .132$), and the interaction between ICC and the group number ($\eta^2 = .0.59$).

The distributions of CI width across the number of groups for both models were illustrated in Figure 17. It was observed that when the number of groups went up, CI width as well as its variance went down for the MSEM model. The same pattern happened to the MLM model. The MSEM model had a large CI width in small group number conditions with means CI width of 1.680 and 0.815 for the group number of 20 and 50, respectively. CI width of the MSEM model significantly reduced when the number of groups increased with the mean CI width of 0.507, 0.279, and 0.165, for the group number of 100, 300, and 1000, respectively. The MLM model had much smaller CI width than the MSEM model in small group numbers which were 0.794 and 0.421 for group numbers of 20 and 50, respectively. For the large group number conditions, the MLM model still had smaller CI width but the differences were not significant (0.283, 0.158, and 0.104 for group number of 100, 300, and 1000, respectively).

Table 8
Eta-squared Analysis of CI Width for MLM and MSEM Models

Factors	MLM Model	MSEM Model
Size2	**0.83454**	**0.43050**
ICC	0.02291	**0.14523**
Size1	0.02288	**0.13229**
Size2*ICC	0.01705	**0.05919**
ME	0.03165	0.05646
Size2*Size1	0.02010	0.05549
Size1*ICC	0.02026	0.03749
ME*Size2	0.00989	0.01898
ME*Size1	0.00139	0.00860
ME*ICC	0.00134	0.00834
ICC*shape	0.00001	0.00058
Size1*shape	0.00008	0.00019
Size2*shape	0.00009	0.00015
shape	0.00004	0.00010
ME*shape	0	0.00009

Note. The effect sizes with significant main effect and interaction in CI width appear in bold. Size2 = number of group; Size1 = group size, ME = effect size; ICC = ICC value; shape = population shape.

ICC value was showed to have a substantial relationship with CI width ($\eta 2 = .145$) in the MSEM model. The overall distributions of estimated CI width by ICC levels for the MSEM model are presented in Figure 18. For the comparison purpose, Figure 18 also demonstrates the distributions of CI width by ICC values for the MLM model. While there was no significant difference among ICC levels for the MLM model, CI width obviously decreased as ICC values increased for the MSEM model. Large CI width (with average of 1.123) was observed at the ICC level of 0.05. The MSEM model also showed some conditions with large CI width at the ICC level of 0.4 but on overall, it had the same CI width as the MLM model with the CI width mean of 0.384.

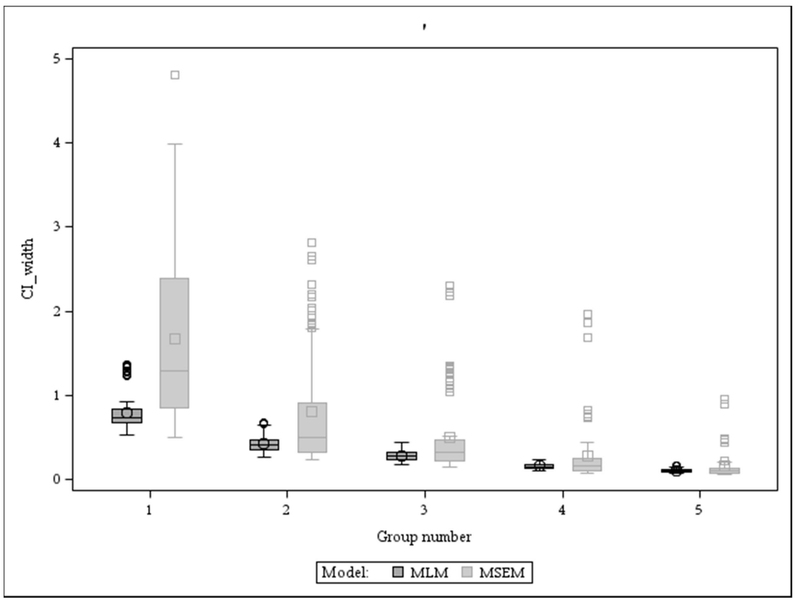

Figure 17. Confidence Interval Width by Group Number for MLM and MSEM Models

The overall distributions of estimated CI width by cell size for MSEM and MLM models were presented in Figure 19. While CI width was the same across cell size levels for the MLM model, it was found to decrease as the cell size increased with the MSEM model. Large mean CI width (1.102) was observed at the smallest cell size level (cell size = 5) for the MSEM model. At the cell size levels of 20 and 50, CI width of the MSEM model had the same average as that of the MLM model even though there were still some conditions with a large CI width.

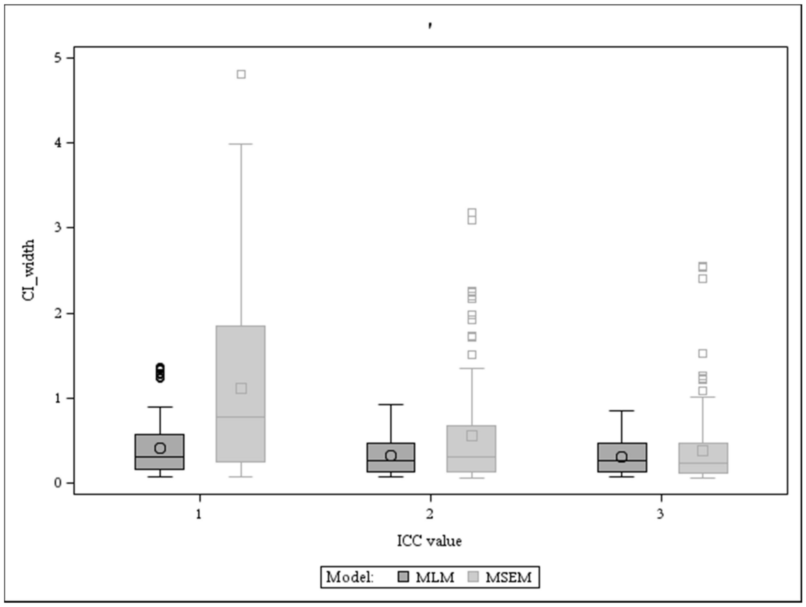

Figure 18. Confidence interval width by ICC level for MLM and MSEM model

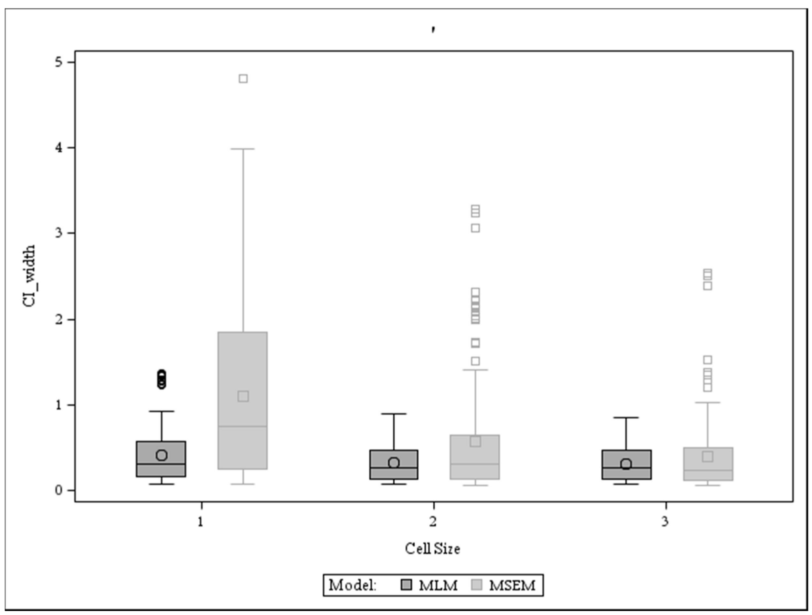

Figure 19. Confidence interval width by cell size level for the MLM and MSEM models

In analysis of variability for the MSEM model, the interaction between the group number and ICC showed the significant effect (η2 = .0.59). The effect of this interaction on the estimated CI width was shown in Figure 20. The impact of ICC on CI width was more noticeable for the group number of 20, with ICC of 0.05 having the largest mean CI width (2.591), followed by the ICC of 0.2 (mean CI width = 1.435), and the ICC of .4 (mean CI width = 1.013). As observed in Figure 20, the mean CI width decreased as the number of group increased and all ICC levels showed similar mean estimated CI width when the number of group equals to 1000.

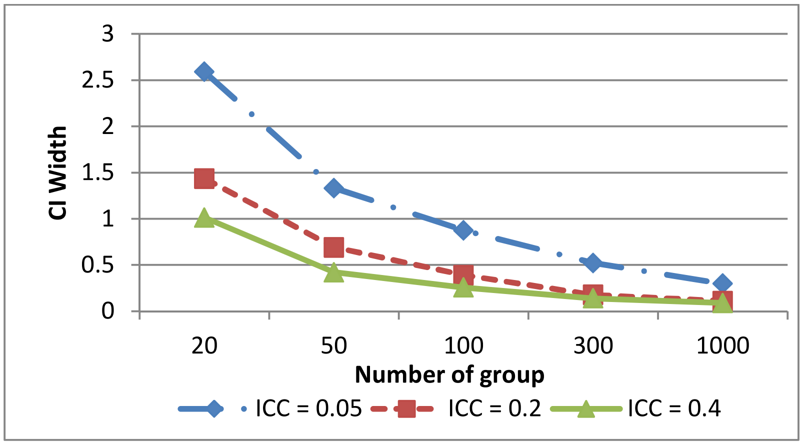

Figure 20. Confidence Interval Width by Group Number and ICC Level

Summary

In general, the MLM model had better CI width than the MSEM model. Specifically, the CI width of the MLM model is significantly smaller than that of the MSEM model when the number of group was small (group number of 20 and 50). The difference in the CI width between two models became trivial when the group numbers were 300 and 1000. ICC was found to associate with confidence interval width of the MSEM model. In general, larger ICC values outperformed the smaller ones in term of the CI width. The estimated CI width also became less disperse as the ICC values increased. Cell size also affected the CI width of the MSEM model. When the cell size was small, the MSEM model was less precise in estimating the mediation effect. As cell size increased, the confidence interval width as well as the variation of CI width decreased. It was evident that the interaction between group number and ICC had a relationship with confidence interval width in the MSEM model. The MSEM model had the largest CI in the small ICC and small number of groups. Increasing in either ICC values or the number of groups was followed by decreasing in the estimate of CI width.

APPENDIX C: SAS CODE AND MPLUS CODE FOR THE SIMULATION PROCEDURE

SAS code to generate data, fit the MLM model, and to generate bootstrapping data

```
filename junk dummy;

proc printto log = junk print = junk;

%global reps bootstrap pme Pshape size1 size2 picc;

%let reps = 20;
%let bootstrap = 100;
%let pme = 1; * from 1-4;
%let size2 = 1; * from 1-5;
%let size1 = 1; * from 1-3;
%let picc = 1; * from 1-3;
%let Pshape = 1; * from 1-3;

*-------------------------------------------------------------------------------
The macro Gendata will generate a 2-1-1 data
-------------------------------------------------------------------------------
Inputs:
Outputs:
        SIMdata - 4 variables: groupID, X, M, Y where X is in level 2, M and Y in level 1
-------------------------------------------------------------------------------;
%Macro GenData;
        Proc IML;
                Start gendata_lvl2(nn,seed1,variance,bb,cc,dd,mu,r_matrix,rawdata);
                        L=eigval(r_matrix);        neg_eigval=0;

                        do r=1 to nrow(L);
                                if L[r,1] < 0 then neg_eigval=1;
                        end;

                        if neg_eigval=0 then do;
                                cols=ncol(r_matrix);
                                g = root(r_matrix);
                                rawdata=rannor(repeat(seed1,nn,cols));
                                rawdata=rawdata*g;
                                do r=1 to nn;
                                do c=1 to cols;
                                        rawdata[r,c]=(-1*cc) + (bb*rawdata[r,c]) + (cc*rawdata[r,c]##2) +
(dd*rawdata[r,c]##3);
                                        rawdata[r,c]=(rawdata[r,c] * sqrt(variance[1,c])) + mu[1,c];
                                end;
                                end;
```

```
			end;

			if neg_eigval=1 then do;
				cols=ncol(r_matrix);
				v=eigvec(r_matrix);
				do i=1 to nrow(L);
				do j=1 to ncol(v);
						if L[i,1] > 0 then v[j,i] = v[j,i] # sqrt(L[i,1]);
						if L[i,1] <= 0 then v[j,i] = v[j,i] # sqrt(.000000001);
					end;
				end;
				rawdata=rannor(repeat(seed1,nn,cols));
				rawdata=V*rawdata`;
				rawdata=rawdata`;
				do r=1 to nn;
				do c=1 to cols;
						rawdata[r,c]=(-1*cc) + (bb*rawdata[r,c]) +
(cc*rawdata[r,c]##2) + (dd*rawdata[r,c]##3);
					rawdata[r,c]=(rawdata[r,c] * sqrt(variance[1,c])) + mu[1,c];
				end;
				end;
			end;
		finish;

		Start gendata_lvl1(nn1,nn2,seed1,variance1,data2,bb,cc,dd,r1_matrix,rawdata);
		  n_group = nrow(data2);
		  haft_n_group = n_group/2;
		  mean_tmp1 = data2[k,2:3];
		  mean_tmp2 = data2[k+haft_n_group,2:3];
		  x_k1 = data2[k,1];
		  x_k2 = data2[k+haft_n_group,1];
		  run gendata_lvl2(nn1,seed1,variance1,bb,cc,dd,mean_tmp1,r1_matrix,rawdata_k1);
		  group_index1 = j(nn1,2,k);
		  group_index1[,2] = x_k1;
		  run gendata_lvl2(nn2,seed1,variance1,bb,cc,dd,mean_tmp2,r1_matrix,rawdata_k2);
		  group_index2 = j(nn2,2,k+haft_n_group);
		  group_index2[,2] = x_k2;
		  rawdata_k1 = group_index1||rawdata_k1;
		  rawdata_k2 = group_index2||rawdata_k2;
		  rawdata_k = rawdata_k1//rawdata_k2;
		  if k =  1 then rawdata = rawdata_k;
			else rawdata = rawdata//rawdata_k;
		  end;
		finish;

		Start genR_matrix(a,b,c,r_matrix);
			r_matrix = J(3,3,1);
			r_matrix[1,2] = a;
			r_matrix[2,1] = a;
			r_matrix[1,3] = b*a+c;
			r_matrix[3,1] = b*a+c;
			r_matrix[2,3] = c*a + b;
			r_matrix[3,2] = c*a + b;
		finish;

		start intermediate_r(b1,c1,d1,target,icor);
```

```
				r = .5;     b2 = b1;c2 = c1;d2 = d1;
				maxiter = 5;
				converge = 0.000001;
				ratio = 1;
				do iter = 1 to maxiter while(abs(ratio) > converge);
					deriv =
(3*r**2*6*d1*d2+2*r*2*c1*c2+(b1*b2+3*b1*d2+3*d1*b2+9*d1*d2));
					function =
(r**3*6*d1*d2+r**2*2*c1*c2+r*(b1*b2+3*b1*d2+3*d1*b2+9*d1*d2)-target);
					ratio = function/deriv;
					r = r - ratio;
				end;
				icor = r;
			finish;

			pop_var2 = J(1,3,1);
			pop_mean2 = J(1,3,0);
			p_r = j(3,3,1);
			seed1=round(1000000*ranuni(0));
			sn = &reps;
			btn = &bootstrap;
			pmei = &pme;
			Pshapei = &Pshape;
			size2i = &size2;
			size1i = &size1;
			picci = &picc;

			if pmei = 1 then; do; run genR_matrix(0, 0, 0.1, p_r2); *print p_r2; end;      * a = 0, b = 0, c' = 0.1;
			if pmei = 2 then; do; run genR_matrix(0.3, 0.1, 0.1, p_r2); *print p_r2; end; * a = 0.3, b = 0.1 c' =
0.1;
			if pmei = 3 then; do; run genR_matrix(0.3, 0.4, 0.1, p_r2); *print p_r2; end; * a = 0.3, b = 0.4 c' =
0.1;
			if pmei = 4 then; do; run genR_matrix(0.6, 0.4, 0.1, p_r2); *print p_r2; end; * a = 0.6, b = 0.4 c' =
0.1;

			if Pshapei = 1 then; do; bb = 1; cc = 0; dd = 0; end;
			if Pshapei = 2 then; do; bb = 0.8798319; cc = 0.2581376; dd = 0.016749; end;
			if Pshapei = 3 then; do; bb =  0.5934025; cc = 0.2262259; dd = 0.1059367; end;

			if size2i = 1 then group_k = 20;
			if size2i = 2 then group_k = 50;
			if size2i = 3 then group_k = 100;
			if size2i = 4 then group_k = 300;
			if size2i = 5 then group_k = 600;
			if size1i = 1 then do; n1_i = 3; n2_i = 7; end;
			if size1i = 2 then do; n1_i = 17; n2_i = 23; end;
			if size1i = 3 then do; n1_i = 39; n2_i = 61; end;

			if picci = 1 then icc = 0.05;
			if picci = 2 then icc = 0.20;
			if picci = 3 then icc = 0.40;

			variance1 = (1/icc)-1;
			covariance1 = variance1*0.2;

			Cov_matrix1 = J(2,2,0);
```

```
			Cov_matrix1[1,1] = variance1;
			Cov_matrix1[2,2] = variance1;
			Cov_matrix1[2,1] = covariance1;
			Cov_matrix1[1,2] = covariance1;

			var_matrix1 = J(1,2,0);
			var_matrix1[1] = variance1;
			var_matrix1[2] = variance1;

			SD1 = sqrt(diag(Cov_matrix1));
			SD1Inv = inv(SD1);
			r_matrix1 = SD1Inv * Cov_matrix1 * SD1inv;

			target1 = r_matrix1[1,2];

			run intermediate_r(bb,cc,dd,target1,icor1);

			r_matrix1[1,2] = icor1;
			r_matrix1[2,1] = icor1;

			run gendata_lvl2(group_k,seed1,pop_var2,1,0,0,pop_mean2,p_r2,rawdata2);
			run gendata_lvl1(n1_i,n2_i,seed1,var_matrix1,rawdata2,bb,cc,dd,r_matrix1,rawdata);
			call sort(rawdata, 1); /** sort matrix c_data by the 1st col **/
			create SIMdata from rawdata [colname={group X M Y}];
			append from rawdata;
		Quit;
%MEnd;

*--------------------------------------------------------------------------------
This macro estimates the MLM model
Model A: regression of M on X and model B: regression of Y on X and M
For meodel A: the coefficien a is the effect of X on M
For Model B: the coefficien b is the effect of MM on Y while MM is group mean of M
				the coefficien c is the effect of MC on Y while MC is centering M
--------------------------------------------------------------------------------
Inputs:
   inputdata - a SAS multilevel 2-1-1 dataset
Outputs:
		ModelEstimates - a SAS dataset with 9 columns of 1 observation: The fixed effect, random effect and
convergence status of models
		resid1 - a SAS dataset of level-1 residuals of model A and B which are e1 and e2, respectively
		resid2 - a SAS dataset of level-2 residuals of model A and B which are u1 and u2, respectively
--------------------------------------------------------------------------------;
%macro MLM (inputdata, ModelEstimates, resid1, resid2);

		* Estiamte model A: regression of M on X;
		proc mixed data = &inputdata noclprint method = ML NOBOUND;
		  class group;
		  model m = x / solution outp = resA1;
		  random intercept/ s subject = group;
		ods output SolutionF = FixedeffectA Solutionr = resA2 CovParms = VarA ConvergenceStatus =
ConvergenceA;

		* transpose table of fixed effect of model A;
		proc transpose data=FixedeffectA out=fixedA prefix = F suffix = A;
			id effect;var estimate;
```

```
* transpose table of random effect of model A;
proc transpose data=varA out=randomA prefix = R suffix = A;
        id CovParm;     var estimate;

data convergenceA;      set convergenceA;
        statusA = status;  keep statusA;

* Estimate model B: regression of Y on M and X;
*Calculate group mean and centering M;

Proc SQL;
        Create table Centering_simdata as select *, mean(M) as MM, M - mean(M) as MC from
&inputdata group by group;
Quit; run;

proc mixed data = Centering_simdata noclprint method = ML NOBOUND;
  class group;
  model Y = MC MM X / solution outp = resB1;
  random intercept/ s subject = group;
ods output SolutionF = FixedeffectB Solutionr = resB2 CovParms = VarB ConvergenceStatus =
ConvergenceB;

* transpose table of fixed effect of model B;
proc transpose data=FixedeffectB out=fixedB prefix = F suffix = B;
        id effect; var estimate;

* transpose table of random effect of model B;
proc transpose data=varB out=randomB prefix = R suffix = B;
        id CovParm; var estimate;

data convergenceB;      set convergenceB;
        statusB = status;  keep statusB;

* level-1 residual of model A in e1;
proc sql;
 create table reside1 as select resid as e1 from resA1
quit;
* level-1 residual of model B in e2;
proc sql;
 create table reside2 as select resid as e2 from resB1
quit;

* level-2 residual of model A in u1;
proc sql;
 create table residu1 as select estimate as u1 from resA2
quit;
* level-2 residual of model B in u2;
proc sql;
 create table residu2 as select estimate as u2 from resB2
quit;

data &resid1; merge reside1 reside2;
data &resid2; merge residu1 residu2;

* Create table ModelEstimates;
```

```
	data &ModelEstimates;
		merge fixedA randomA fixedB randomB convergenceA convergenceB;
		MLMstatus = 0;
		MLM_IDE = FXA*FMMB;

	*MLMstatus variable presents the convergence status of MLM mediation model;

	data &ModelEstimates;
		set &ModelEstimates;
		if statusA ne 0 or statusB ne 0 then MLMstatus = 1;
		else do;
			if RinterceptA <= 0 and RinterceptB > 0 then MLMstatus = 2;
			if RinterceptB > 0 and RinterceptB <= 0 then MLMstatus = 3;
			if RinterceptB <= 0 and RinterceptB <= 0 then MLMstatus = 4;
		end;
	run;

%mend;

*--------------------------------------------------------------------------------
The macro Rescale calculate the centering residuals of MLM models
Model A: regression of M on X and model B: regression of Y on X and M
--------------------------------------------------------------------------------
Inputs:
	resid1 - a SAS dataset of level-1 residuals of model A and B which are e1 and e2, respectively
	resid2 - a SAS dataset of level-2 residuals of model A and B which are u1 and u2, respectively

Outputs:
	newresid1 - a SAS dataset of grand mean centering level-1 residuals of estimated MLM models
	newresid2 - a SAS dataset of grand mean centering level-2 residuals of estimated MLM models
--------------------------------------------------------------------------------;

%Macro Rescale(resid1, resid2, newresid1, newresid2);
	* Rescale level-1 residuals ;
	Proc SQL;
		Create table &newresid1 as select e1-mean(e1) as c_e1, e2-mean(e2) as c_e2 from &resid1;
	Quit;
	* Rescale level-2 residuals ;
	Proc SQL;
		Create table &newresid2 as select u1-mean(u1) as c_u1, u2-mean(u2) as c_u2 from &resid2;
	Quit;
%MEnd;

*--------------------------------------------------------------------------------
The macro Transform will transform residuals  the centering residuals of MLM models
Model A: regression of M on X and model B: regression of Y on X and M
--------------------------------------------------------------------------------
Inputs:
	newresid1 - a SAS dataset of level-1 residuals after centering
	newresid2 - a SAS dataset of level-2 residuals after centering
	ModelEstimates - a SAS data set of  estimates of random and fixed effect from MLM models with
simulated data

Outputs:
	New_resA1 - a SAS dataset of transformed level-1 residuals of estimated MLM model 1
	New_resB1 - a SAS dataset of transformed level-1 residuals of estimated MLM model 2
```

```
        New_resA2 - a SAS dataset of transformed level-2 residuals of estimated MLM model 1
        New_resB2 - a SAS dataset of transformed level-2 residuals of estimated MLM model 2
-------------------------------------------------------------------------------;

%Macro Transform(newresid1, newresid2, ModelEstimates, New_resA1, New_resA2, New_resB1, New_resB2);
        Proc IML;

                Use &newresid1; * read level-1 residuals;
                Read all ;
                Use &newresid2; * read level-2 residuals;
                Read all ;

                total_G = nrow(c_u1); * group size;
                total_N = nrow(c_e1); * sample size;

                Use &ModelEstimates; * estiamtes of random and fixed effects of model A and B;
                read var {FInterceptA FXA RInterceptA RResidualA FInterceptB FMCB FMMB FXB
RInterceptB RResidualB MLMstatus};

                if MLMstatus[1] = 0 then do; * check for MLM model status,
                                                        transform only when model is
converged
                                                        and the variance of intercept is not
equal to 0;
                        * Start transform level-1 residuals;
                        * For Model A;
                        Ls1 = sqrt((c_e1)`*(c_e1)/total_N);
                        Lsigma1 = sqrt(RResidualA[1]);
                        A1 = Lsigma1/Ls1;
                        New_e1 = c_e1 * A1`;
                        *S1 = (New_e1` * New_e1)/total_N;
                        Create &New_resA1 from New_e1 [Colname = 'New_e1'];
                        Append from New_e1;
                        * For Model B;
                        Ls1 = sqrt((c_e2)`*(c_e2)/total_N);
                        Lsigma1 = sqrt(RResidualB[1]);
                        A1 = Lsigma1/Ls1;
                        New_e2 = c_e2 * A1`;
                        *S1 = (New_e1` * New_e1)/total_N;
                        Col = {"New_e2"};
                        Create &New_resB1 from New_e2 [Colname = Col];
                        Append from New_e2;
                        * For level-2 residuals;
                        * For Model A;
                        Ls1 = sqrt((c_u1)`*(c_u1)/total_G);
                        Lsigma1 = sqrt(RInterceptA[1]);
                        A1 = Lsigma1/Ls1;
                        New_u1 = c_u1 * A1`;
                        *S1 = (New_u1` * New_u1)/total_G;
                        Col = {"New_u1"};
                        Create &New_resA2 from New_u1 [Colname = Col];
                        Append from New_u1;
                        * For Model B;
                        Ls1 = sqrt((c_u2)`*(c_u2)/total_G);
                        Lsigma1 = sqrt(RInterceptB[1]);
                        A1 = Lsigma1/Ls1;
```

```
                New_u2 = c_u2 * A1`;
                *S1 = (New_u1` * New_u1)/total_N;
                Col = {"New_u2"};
                Create &New_resB2 from New_u2 [Colname = Col];
                Append from New_u2;
        end; ** of if statement;

        if MLMstatus[1] = 1 then do;
                Create &New_resA1 from c_e1 [Colname = 'New_e1'];
                Append from c_e1;
                Col = {"New_e2"};
                Create &New_resB1 from c_e2 [Colname = Col];
                Append from c_e2;
                Col = {"New_u1"};
                Create &New_resA2 from c_u1 [Colname = Col];
                Append from c_u1;
                Col = {"New_u2"};
                Create &New_resB2 from c_u2 [Colname = Col];
                Append from c_u2;
        end;

        if MLMstatus[1] = 2 then do; * check for MLM model status,
                                                        transform only when model is
converged
                                                        and the variance of intercept is not
equal to 0;
                * Start transform level-1 residuals;
                * For Model A;
                Ls1 = sqrt((c_e1)`*(c_e1)/total_N);
                Lsigma1 = sqrt(RResidualA[1]);
                A1 = Lsigma1/Ls1;
                New_e1 = c_e1 * A1`;
                *S1 = (New_e1` * New_e1)/total_N;
                Create &New_resA1 from New_e1 [Colname = 'New_e1'];
                Append from New_e1;
                * For Model B;
                Ls1 = sqrt((c_e2)`*(c_e2)/total_N);
                Lsigma1 = sqrt(RResidualB[1]);
                A1 = Lsigma1/Ls1;
                New_e2 = c_e2 * A1`;
                *S1 = (New_e1` * New_e1)/total_N;
                Col = {"New_e2"};
                Create &New_resB1 from New_e2 [Colname = Col];
                Append from New_e2;
                * For level-2 residuals;
                * For Model A;
                Col = {"New_u1"};
                Create &New_resA2 from c_u1 [Colname = Col];
                Append from c_u1;
                * For Model B;
                Ls1 = sqrt((c_u2)`*(c_u2)/total_G);
                Lsigma1 = sqrt(RInterceptB[1]);
                A1 = Lsigma1/Ls1;
                New_u2 = c_u2 * A1`;
                *S1 = (New_u1` * New_u1)/total_N;
                Col = {"New_u2"};
```

```
            Create &New_resB2 from New_u2 [Colname = Col];
            Append from New_u2;
        end; ** of if statement;

        if MLMstatus[1] = 3 then do; * check for MLM model status,
                                          transform only when model is
converged
                                          and the variance of intercept is not
equal to 0;
            * Start transform level-1 residuals;
            * For Model A;
            Ls1 = sqrt((c_e1)`*(c_e1)/total_N);
            Lsigma1 = sqrt(RResidualA[1]);
            A1 = Lsigma1/Ls1;
            New_e1 = c_e1 * A1`;
            *S1 = (New_e1` * New_e1)/total_N;
            Create &New_resA1 from New_e1 [Colname = 'New_e1'];
            Append from New_e1;
            * For Model B;
            Ls1 = sqrt((c_e2)`*(c_e2)/total_N);
            Lsigma1 = sqrt(RResidualB[1]);
            A1 = Lsigma1/Ls1;
            New_e2 = c_e2 * A1`;
            *S1 = (New_e1` * New_e1)/total_N;
            Col = {"New_e2"};
            Create &New_resB1 from New_e2 [Colname = Col];
            Append from New_e2;
            * For level-2 residuals;
            * For Model A;
            Ls1 = sqrt((c_u1)`*(c_u1)/total_G);
            Lsigma1 = sqrt(RInterceptA[1]);
            A1 = Lsigma1/Ls1;
            New_u1 = c_u1 * A1`;
            *S1 = (New_u1` * New_u1)/total_G;
            Col = {"New_u1"};
            Create &New_resA2 from New_u1 [Colname = Col];
            Append from New_u1;
            * For Model B;
            Col = {"New_u2"};
            Create &New_resB2 from c_u2 [Colname = Col];
            Append from c_u2;
        end; ** of if statement;

        if MLMstatus[1] = 4 then do; * check for MLM model status,
                                          transform only when model is
converged
                                          and the variance of intercept is not
equal to 0;
            * Start transform level-1 residuals;
            * For Model A;
            Ls1 = sqrt((c_e1)`*(c_e1)/total_N);
            Lsigma1 = sqrt(RResidualA[1]);
            A1 = Lsigma1/Ls1;
            New_e1 = c_e1 * A1`;
            *S1 = (New_e1` * New_e1)/total_N;
            Create &New_resA1 from New_e1 [Colname = 'New_e1'];
```

```
                    Append from New_e1;
                    * For Model B;
                    Ls1 = sqrt((c_e2)`*(c_e2)/total_N);
                    Lsigma1 = sqrt(RResidualB[1]);
                    A1 = Lsigma1/Ls1;
                    New_e2 = c_e2 * A1`;
                    *S1 = (New_e1` * New_e1)/total_N;
                    Col = {"New_e2"};
                    Create &New_resB1 from New_e2 [Colname = Col];
                    Append from New_e2;
                    * For level-2 residuals;
                    * For Model A;
                    Col = {"New_u1"};
                    Create &New_resA2 from c_u1 [Colname = Col];
                    Append from c_u1;
                    * For Model B;
                    Col = {"New_u2"};
                    Create &New_resB2 from c_u2 [Colname = Col];
                    Append from c_u2;
            end; ** of if statement;
        Quit;
%MEnd;

*--------------------------------------------------------------------------------
The macro Bootstrap will Calculate bias, CI of model estimate using bootstrap method
Model A: regression of M on X and model B: regression of Y on X and M
--------------------------------------------------------------------------------
Inputs:
        Trans_resA1 - a SAS dataset of level-1 residuals of model A
        Trans_resB1 - a SAS dataset of level-1 residuals of model B
        Trans_resA2 - a SAS dataset of level-2 residuals of model A
        Trans_resB2 - a SAS dataset of level-2 residuals of model B
        ModelEstimates - a SAS data set of  estimates of random and fixed effect from MLM models with
simulated data

Outputs:
        Bootstrap_estimate - a SAS dataset of estimated outcome using bootstrap
--------------------------------------------------------------------------------;
%Macro Bootstrap(Trans_resA1, Trans_resA2, Trans_resB1, Trans_resB2, ModelEstimates, Bootstrap_estimate);
        * get the model status;
        proc sql noprint;
            select MLMstatus into :Status
            from &ModelEstimates;
        %if &status ne 1 %then %do;
        * get the total sample size;
            proc sql noprint;
                select count(*) into : samplesize from &Trans_resA1;
            quit;
            * get the number of group;
            proc sql noprint;
                select count(*) into : groups from &Trans_resB2;
            quit;
            * get the fixed effect of Model A and B into macro variable;
            proc sql noprint;
                select FinterceptA into :FixedA1
                from &ModelEstimates;
```

```
		select FXA into :FixedA2
		from &ModelEstimates;
		select FinterceptB into :FixedB1
		from &ModelEstimates;
		select FMCB into :FixedB2
		from &ModelEstimates;
		select FMMB into :FixedB3
		from &ModelEstimates;
		select FXB into :FixedB4
		from &ModelEstimates;
	quit;
	*get level-2 data from simulated data;
	Proc sql;
		Create table Sim_lvl2 as select group, mean(X) as MX from Simdata group by group;
	quit;
	run;
	* Bootstrap procedure;
	%do bti = 1 %to &bootstrap; * Bootstrap loop;
		* Draw a random sample with replacement from transformed level-1 residuals of model
A;
		Proc surveyselect data = &Trans_resA1 out = dataA1 method = URS n=&samplesize
outhits noprint;
		* Draw a random sample with replacement from transformed level-1 residuals of model
B;
		Proc surveyselect data = &Trans_resB1 out = dataB1 method = URS n=&samplesize
outhits noprint;
		* Sort the bootstrap data by a random order;
		Data dataA1_new;
			Set dataA1;
			sorterA = RANUNI(-3);
		run;
		Proc Sort Data = dataA1_new;
			by sorterA;	run;
		Data dataB1_new;
			Set dataB1;
			sorterB = RANUNI(-3);
		run;
		Proc Sort Data = dataB1_new;
			by sorterB;	run;
		Data newdata1;
			merge Simdata dataA1_new dataB1_new;
			drop sorterB sorterA;
		run;
		* Draw a random sample with replacement from transformed level-2 residuals of model
A;
		Proc surveyselect out = dataA2 data=&Trans_resA2 method = URS n=&groups outhits
noprint;
		* Draw a random sample with replacement from transformed level-2 residuals of model
B;
		Proc surveyselect out = dataB2 data=&Trans_resB2 method = URS n=&groups outhits
noprint;
		* Sort the bootstrap data by a random order;
		Data dataA2_new;
			Set dataA2;
			sorterA = RANUNI(-3);
		run;
```

```
Proc Sort Data = dataA2_new;
          by sorterA;          run;
Data dataB2_new;
          Set dataB2;
          sorterB = RANUNI(-3);
run;
Proc Sort Data = dataB2_new;
          by sorterB;          run;
Data newdata2;
          merge Sim_lvl2 dataA2_new dataB2_new;
          drop sorterB sorterA;
proc sort data = Newdata1;
          by group; run;
* Generate Bootstrap resample;
Data Bootstrap_data ;
          Merge newdata1 newdata2;
          by group;
          new_M = &FixedA1 + &FixedA2*X + New_e1 + New_u1;
Proc SQL;
          Create table Bootstrap_data_MM as select *, mean(new_M) as new_MM from
Bootstrap_data group by group;
Quit;
Data Bootstrap_data_MM;
          set Bootstrap_data_MM;
          new_Y = &FixedB1 + &FixedB4*X + &FixedB3*new_MM +
&FixedB2*New_M + New_e2 + New_u2;
* Centering Bootstrap_ data;
Proc SQL;
          Create table Bootstrap_data_c as select *, New_M - mean(new_m) as C_new_M
from Bootstrap_data_MM group by group;
Quit;
* Run the Multilevel Model using bootstrap resample;
proc mixed data = Bootstrap_data_c noclprint method = ML NOBOUND;
  class group;
  model New_m = x / solution;
  random intercept/ s subject = group;
ods output SolutionF = FixedeffectAB;
proc mixed data = Bootstrap_data_c noclprint method = ML NOBOUND;
  class group;
  model New_y = x C_new_m new_MM/ solution;
  random intercept/ s subject = group;
ods output SolutionF = FixedeffectBB;
proc transpose data=FixedeffectAB out=fixedAB prefix = F suffix = A;
          id effect;
          var estimate;
proc transpose data=FixedeffectBB out=fixedBB prefix = F suffix = B;
          id effect;
          var estimate;
Data effectBootstrap;
          merge fixedAB fixedBB;
          Mediation_effect = FXA * Fnew_MMB;
%if &bti = 1 %then %do;
          Data allfixedB;
                    set effectBootstrap;
%end;
%else %do;
```

```
				Data allfixedB;
					set allfixedB effectBootstrap;
			%end;
		%end; * End of Bootstrap loop;
		proc sql noprint;
			select count(*) into: n_obs
				from allfixedB;
		quit;
		%let _HL = round(0.975*&n_obs);
		%let _LL = round(0.025*&n_obs);
		Proc sort data = allfixedB;
			by Mediation_effect;
		* Get the estimate of a coefficient;
		data Low_limit;
			set allfixedB;
			if _N_ = &_LL then MLM_LL = Mediation_effect;
			if MLM_LL  = . then delete;
			keep MLM_LL;
		data High_limit;
			set allfixedB;
			if _N_ = &_HL then MLM_HL = Mediation_effect;
			if MLM_HL = . then delete;
			keep MLM_HL;
		data &Bootstrap_estimate;
			merge &ModelEstimates Low_limit High_limit;
		run;
	%end; * End of if checking for model status;
	%if &status = 1 %then %do;
		data &Bootstrap_estimate;
			set &ModelEstimates;
			MLM_LL = .;
			MLM_HL = .;
	%end;
%MEnd;

%Macro SimAll;
	%do i = 1 %to &reps;
		%GenData;
		* For MLM model;
		%MLM(SIMdata, ModelEstimates, resid1, resid2);
		%Rescale(resid1, resid2, newresid1, newresid2);
		%Transform(newresid1, newresid2, ModelEstimates, Trans_resA1, Trans_resA2, Trans_resB1,
Trans_resB2);
		%Bootstrap(Trans_resA1,Trans_resA2, Trans_resB1,  Trans_resB2, ModelEstimates,
Bootstrap_estimate);
		run;
		data MLMRep_result;
			set Bootstrap_estimate;
			keep MLM_IDE MLMStatus MLM_LL MLM_HL;
		data Simdata;
			set simdata;
			rep_i = &i;
		%if &i = 1 %then %do;
			Data AllReps_MLM;
				set MLMRep_result;
			Data AllSimdata;
```

```
                    set Simdata;
            %end;
            %else %do;
                Data AllReps_MLM;
                    set AllReps_MLM MLMRep_result;
                data AllSimdata;
                    set AllSimdata Simdata;
            %end;
            run;
        %end;
%MEnd;

%SimAll;

Data AllReps_MLM1;
    set AllReps_MLM;
            Population_ME = &pme;
            if Population_ME = 1 then P_IDE = 0;
            if Population_ME = 2 then P_IDE = 0.03;
            if Population_ME = 3 then P_IDE = 0.12;
            if Population_ME = 4 then P_IDE = 0.24;
            MLM_bias = MLM_IDE - P_IDE;
            MLM_CI_width = MLM_HL - MLM_LL;
            MLM_reject = 0;
            if (MLM_LL < 0) and (MLM_HL < 0) then MLM_reject = 1;
            if (MLM_LL > 0) and (MLM_HL > 0) then MLM_reject = 1;
            MLM_coverage = 0; * CI coverage status: 0 = non-cover, 1 = cover;
            if (MLM_LL < P_IDE) and (MLM_HL > P_IDE) then MLM_coverage = 1;
            MLM_RMSE = (MLM_IDE-P_IDE)*(MLM_IDE-P_IDE); * sqaured difference between est. IDE
and population IDE to calculate RMSE later;
            if MLMstatus = 1 then delete;     * Delete the non convergence replication;

* Sumarize the outcome from replications;
* nObs - the convergence rate: = nObs/N_rep;
* reject - the rejection rate of null hypothesis, which is no mediation effect - For Type I error and Power analysis;
* bias - the avarage bias of replications in this condition;
* CI_width - the avarage width of Confidence Interval;
* ESD - estimated standard deviation;
* SE - standard error of the mean of estimated indirect effect;
* RMSE - Root mean squared error;

Proc SQL;
            Create table MLM_result as select count(*) as MLM_nObs,

    sum(MLMStatus=0) as Convergence,

    mean(MLM_IDE) as MLM_IDE,

    mean(MLM_bias) as MLM_bias,

    mean(MLM_CI_width) as MLM_CI_width,

    mean(MLM_reject) as MLM_reject,
                                                                                mean(P_IDE) as
P_IDE,
```

```
                                                    std(MLM_IDE)
as MLM_ESD,

            STDERR(MLM_IDE) as MLM_SE,

            mean(MLM_coverage) as MLM_coverage,

            sqrt(mean(MLM_RMSE)) as MLM_RMSE
                                                    from
AllReps_MLM1;

            Quit;

data Final;
            set MLM_result;
            N_Rep = 500;
            N_Bootstrap = 1000;
            Population_ME = &pme;
            Population_shape = &Pshape;
            Population_grp = &size2;
            Population_cell = &size1;
            Population_ICC = &Picc;
proc printto print = print;
Proc print data = final HEADING = H;
proc print data = AllSimdata;run;
```

SAS code to generate bootstrapping data for MSEM model

```
options PS = 500 LS = 150;
options noxwait xsync;
options nosource nonotes;

%global reps bootstrap pme Pshape size1 size2 picc;

%let reps = 500;
%let bootstrap = 1000;
%let pme = 4; * from 1-4;
%let size2 = 4; * from 1-5;
%let size1 = 1; * from 1-3;
%let picc = 1; * from 1-3;
%let Pshape = 1; * from 1-3;

data one;
 infile 'U:\Proposal\SimOutcome\MLMSim44111.lst' missover pad;
 input test_var $ 1-260 @;
 rec_num = _n_;
 if INDEX(test_var,'Obs    GROUP') ^= 0 then delete;
 if INDEX(test_var,'The SAS System') ^= 0 then delete;
 if INDEX(test_var,'Obs nObs') ^= 0 then delete;
 if INDEX(test_var,'MLM_') ^= 0 then delete;
 if INDEX(test_var,'500     1000') ^= 0 then delete;
 if INDEX(test_var,'500        1000') ^= 0 then delete;
 if INDEX(test_var,'500      1000') ^= 0 then delete;
 input @1 Obs GROUP X M Y rep_i;
run;
```

```
data two;
	set one;
	if X = . then delete;
	if group = &pme and X = &Pshape and M = &size2 and Y = &size1 and rep_i = &picc then delete;
	keep GROUP X M Y rep_i;
run;

*--------------------------------------------------------------------------------
This macro estimates the MSEM model
--------------------------------------------------------------------------------
Inputs:
   inputdata - a SAS multilevel 2-1-1 dataset

Outputs:
	ModelEstimates - a SAS dataset of 1 observation: The between effect, within effect,
						model estimates of variances of residuals and
convergence status of models
	resid1 - a SAS dataset of level-1 residuals of MSEM model, e1 is M residual while e2 is Y residual
	resid2 - a SAS dataset of level-2 residuals of MSEM model, e1 is M residual while e2 is Y residual
Note:
	For variable M: e1 is the within component of M
					The model estimated variance of e1 is variance of M under within level
section
	For Variable Y: e2 is calculated by equations provided before the code.
					The model estimated varaince of e2 is provided by Mplus under within
level section
--------------------------------------------------------------------------------;
%macro MSEM (inputdata, ModelEstimates, resid1, resid2);
	* export simdata to text file as a input file for Mplus;
	data simtmp;
		set &inputdata;
		obs_id  = _N_;
	proc sort data = simtmp;
		by obs_id;
	* export inputdata to a text file for Mplus;
	data _null_ ;
		set simtmp;
   FILE  'C:\Users\tvpham2\Desktop\Simdata.txt' ;
   PUT Group X M Y;
	run;
	* Call Mplus to run MSEM model;
	X call "C:\Program Files (x86)\Mplus\Mplus.exe" "U:\Proposal\MediationMSEM.inp"
"C:\Users\tvpham2\Desktop\MediationMSEM.out";
	run;
	* Import the Fscore table for caculation of residuals;
	data MSEM_Fscore;
		infile 'C:\Users\tvpham2\Desktop\FS.txt';
		input X M Y B_M B_M_Se B_Y B_Y_Se group;
		obs_id = _N_;
		keep B_M B_Y obs_id;
	proc sort data = MSEM_Fscore;
		by obs_id;
	* Import model estimate;
	data MSEM_coeff;
	   infile 'C:\Users\tvpham2\Desktop\MediationMSEM.out' truncover scanover flowover;
```

```
		input @ 'MODEL RESULTS' //////// dummy1 $ b_within /// dummy10 $ Variance_e1 /// dummy6
$ Variance_e2;
		input @ 'Between Level' /// dummy5 $ a_between ///  dummy2 $ b_between / dummy3 $
c_between;
		input @ 'Intercepts' / dummy8 $ intercept_M / dummy4 $ intercept_Y ////// dummy9 $
Variance_u1 / dummy7 $ Variance_u2;
		input @ 'INDB' MSEM_IDE;
		drop dummy1 dummy2 dummy3 dummy4 dummy5 dummy6 dummy7 dummy8 dummy9
dummy10;
	* calculate the level 1 residual;
	* -----------------------------------------------------------------------;
	* For variable Y: e2;
	* eta_Yij (within component) = Y_ij (observed score) - eta_Yj (between component);
	* eta_Yij = beta_within_YM * eta_Mij + epsilon_Yij;
	* eta_Mij (within component) = M_ij (observed score) - eta_Mj (between component);
	* hence epsilon_Yij = Y_ij - eta_Yj - beta_within_YM * (M_ij - eta_Mj);
	* the between components of M and Y can be obtained from Fscore table of Mplus output (FS.txt);
	* -----------------------------------------------------------------------;
	* For variable M: e1;
	* is the within component of M which is the difference between observed M and the between component
B_M;
	* -----------------------------------------------------------------------;
	data comb1;
		merge simtmp MSEM_Fscore;
		by obs_id;
	data comb2;
		set comb1;
		if _n_ eq 1 then do;
			set MSEM_coeff;
		end;
	data &resid1;
		set comb2;
		e2 = y - B_Y - b_within*(M-B_M);
		e1 = M-B_M;
	run;
	* calculate the level 2 residual;
	* -----------------------------------------------------------------------;
	* For variable Y: u2;
	* For the between part;
	* B_Y = intercept + beta_between_MY * B_M + beta_between_XY * X + u2;
	* -----------------------------------------------------------------------;
	* For variable M: u1;
	* B_M = intercept + beta_between_XM*X + u1;
	* -----------------------------------------------------------------------;
	proc sort data = comb2 out = &resid2 NODUPKEY;
		by group ;
	data &resid2;
		set &resid2;
		u2 = B_Y - intercept_Y - b_between * B_M - c_between * X;
		u1 = B_M - intercept_M - a_between * X;
		drop M Y obs_id;
	* get the estimates of coefficients of MSEM model;
	data &ModelEstimates;
		set MSEM_coeff;
		MSEMstatus = 0;
	* Check for model status;
```

```
	data &ModelEstimates;
		set &ModelEstimates;
		if a_between = . then MSEMstatus = 1;
	run;
%Mend;

*-------------------------------------------------------------------------------
The macro MSEMRescale calculate the grand mean centering residuals of MSEM model
--------------------------------------------------------------------------------
Inputs:
	resid1 - a SAS dataset of level-1 residuals, e1 and e2
	resid2 - a SAS dataset of level-2 residuals, u1 and u2
Outputs:
	C_resid1 - a SAS dataset of grand mean centering level-1 residuals, c_e1 and c_e2
	C_resid2 - a SAS dataset of grand mean centering level-2 residuals, c_u1 and c_u2
-------------------------------------------------------------------------------;

%Macro MSEMRescale(resid1, resid2, C_resid1, C_resid2);
	* Rescale level-1 residuals ;
	Proc SQL;
		Create table &C_resid1 as select e1-mean(e1) as c_e1, e2-mean(e2) as c_e2 from &resid1;
	Quit;
	* Rescale level-2 residuals ;
	Proc SQL;
		Create table &C_resid2 as select u1-mean(u1) as c_u1, u2-mean(u2) as c_u2 from &resid2;
	Quit;
%MEnd;
*-------------------------------------------------------------------------------
The macro MSEMTrans will conduct the grand mean centering and then transform residuals of MSEM model
--------------------------------------------------------------------------------
Inputs:
	c_resid1 - a SAS dataset of level-1 residuals
	c_resid2 - a SAS dataset of level-2 residuals
	ModelEstimates - a SAS data set of  estimates of coefficient of MSEM model with simulated data
Outputs:
	New_resMe - a SAS dataset of transformed level-1 residuals of variable M
	New_resMu - a SAS dataset of transformed level-2 residuals of variable M
	New_resYe - a SAS dataset of transformed level-1 residuals of variable Y
	New_resYu - a SAS dataset of transformed level-2 residuals of variable Y
-------------------------------------------------------------------------------;
%Macro MSEMTrans(c_resid1, c_resid2, ModelEstimates, New_resMe, New_resMu, New_resYe, New_resYu);
	Proc IML;
		Use &C_resid1; * read level-1 residuals;
		Read all;
		Use &c_resid2; * read level-2 residuals;
		Read all;
		total_G = nrow(c_u1); * group size;
		total_N = nrow(c_e1); * sample size;
		Use &ModelEstimates; * estiamtes of random and fixed effects of model A and B;
		read var {b_within a_between b_between c_between intercept_Y intercept_M variance_e1
variance_e2 variance_u1 variance_u2 MSEMstatus};
		if MSEMstatus[1] = 0 then do; * check for MSEM model status,
								transform only when model is
converged
								and the variance of intercept is not
equal to 0;
```

```
                        * Start trasform level-1 residuals;
                        * For M variable;
                        Ls1 = sqrt((c_e1)`*(c_e1)/total_N);
                        Lsigma1 = sqrt(variance_e1[1]);
                        A1 = Lsigma1/Ls1;
                        T_e1 = c_e1 * A1`;
                        Create &New_resMe from T_e1 [Colname = 'T_e1'];
                        Append from T_e1;
                        * For Y variable;
                        Ls1 = sqrt((c_e2)`*(c_e2)/total_N);
                        Lsigma1 = sqrt(variance_e2[1]);
                        A1 = Lsigma1/Ls1;
                        T_e2 = c_e2 * A1`;
                        Create &New_resYe from T_e2 [Colname = 'T_e2'];
                        Append from T_e2;
                        * For level-2 residuals;
                        * For M variable;
                        Ls1 = sqrt((c_u1)`*(c_u1)/total_G);
                        Lsigma1 = sqrt(variance_u1[1]);
                        A1 = Lsigma1/Ls1;
                        T_u1 = c_u1 * A1`;
                        Col = {"T_u1"};
                        Create &New_resMu from T_u1 [Colname = Col];
                        Append from T_u1;
                        * For Y variable;
                        Ls1 = sqrt((c_u2)`*(c_u2)/total_G);
                        Lsigma1 = sqrt(variance_u2[1]);
                        A1 = Lsigma1/Ls1;
                        T_u2 = c_u2 * A1`;
                        Col = {"T_u2"};
                        Create &New_resYu from T_u2 [Colname = Col];
                        Append from T_u2;
                end; ** of if statement for checking status;
                if MSEMstatus[1] <> 0 then do;
                        Create &New_resMe from c_e1 [Colname = 'T_e1'];
                        Append from c_e1;
                        Create &New_resYe from c_e2 [Colname = 'T_e2'];
                        Append from c_e2;
                        Col = {"T_u1"};
                        Create &New_resMu from c_u1 [Colname = Col];
                        Append from c_u1;
                        Col = {"T_u2"};
                        Create &New_resYu from c_u2 [Colname = Col];
                        Append from c_u2;
                end;
        Quit;
%MEnd;

*--------------------------------------------------------------------------------
The macro MSEMBootstrap will Calculate bias, CI of model estimate using bootstrap method
---------------------------------------------------------------------------------
Inputs:
        New_resMe - a SAS dataset of level-1 residuals of M
        New_resMu - a SAS dataset of level-2 residuals of M
        New_resYe - a SAS dataset of level-2 residuals of Y
        New_resYu - a SAS dataset of level-2 residuals of Y
```

```
        ModelEstimates - a SAS dataset of 1 observation: The between effect, within effect,
                                                              model estimates of variances of residuals and
convergence status of models

Outputs:
        Bootstrap_estimate - a SAS dataset of estimated outcome using bootstrap
--------------------------------------------------------------------------------;

%Macro MSEMBootstrap(New_resMe, New_resMu, New_resYe, New_resYu, ModelEstimates,
Bootstrap_estimate);
        * Check for the model status;
        proc sql noprint;
                select MSEMstatus into :Status
                from &ModelEstimates;
        %if &status = 0 %then %do;
                * get the total sample size;
                proc sql noprint;
                        select count(*) into : samplesize from &New_resMe;
                quit;
                * get the number of group;
                proc sql noprint;
                        select count(*) into : groups from &New_resMu;
                quit;
                * get the coefficient estimate of MSEM model into macro variables;
                proc sql noprint;
                        select b_within into :b_w
                        from &ModelEstimates;
                        select a_between into :a_b
                        from &ModelEstimates;
                        select b_between into :b_b
                        from &ModelEstimates;
                        select c_between into :c_b
                        from &ModelEstimates;
                        select intercept_M into :int_M
                        from &ModelEstimates;
                        select intercept_Y into :int_Y
                        from &ModelEstimates;
                quit;
                *get level-2 data from simulated data;
                Proc sql;
                        Create table Sim_lvl2 as select group, mean(X) as MX from Simdata group by group;
                run;
                * Bootstrap procedure;
                %do bti = 1 %to &bootstrap; * Bootstrap loop;
                        * Draw a random sample with replacement from transformed level-1 residuals of M;
                        Proc surveyselect data = &New_resMe out = dataMe method = URS n=&samplesize
outhits noprint;
                        * Draw a random sample with replacement from transformed level-1 residuals of Y;
                        Proc surveyselect data = &New_resYe out = dataYe method = URS n=&samplesize
outhits noprint;
                        * Sort the bootstrap data by a random order;
                        Data dataMe_new;
                                Set dataMe;
                                sorterM = RANUNI(-3);
                        run;
                        Proc Sort Data = dataMe_new;
```

```
		by sorterM;
	run;
	Data dataYe_new;
		Set dataYe;
		sorterY = RANUNI(-3);
	run;
	Proc Sort Data = dataYe_new;
		by sorterY;
	run;
	Data newdata1;
		merge Simdata dataMe_New dataYe_New;
		drop sorterM sorterY;
	run;
	* Draw a random sample with replacement from transformed level-2 residuals of M;
	Proc surveyselect out = dataMu data=&New_resMu method = URS n=&groups outhits
noprint;
	* Draw a random sample with replacement from transformed level-2 residuals of M;
	Proc surveyselect out = dataYu data=&New_resYu method = URS n=&groups outhits
noprint;
	* Sort the bootstrap data by a random order;
	Data dataMu_new;
		Set dataMu;
		sorterM = RANUNI(-3);
	run;
	Proc Sort Data = dataMu_new;
		by sorterM;	run;
	Data dataYu_new;
		Set dataYu;
		sorterY = RANUNI(-3);
	run;
	Proc Sort Data = dataYu_new;
		by sorterY;	run;
	Data newdata2;
		merge Sim_lvl2 dataMu_New dataYu_New;
		drop sorterM sorterY;
	run;
	proc sort data = Newdata1;
		by group;
	proc sort data = Newdata2;
		by group;	run;
	* Generate Bootstrap resample;
	* New_M = B_M + W_M where W_M is T_e1 while B_M = a * X + level-2 residual of
M (e1);
	* New_Y = B_Y + W_Y where W_Y equals b_w * W_M + level-1 residual of Y (e2)
					while B_Y is intercept_Y + c*X + b*B_M +
level-2 residual of Y (u2);
	* New B_M is the new_M after subtracting the new_W_M (T_e1);
	Data Bootstrap_data ;
		Merge newdata1 newdata2;
		by group;
		new_M = &int_M + &a_b * X + T_e1 + T_u1;
		new_Y = &b_w * T_e1 + T_e2 + &int_Y + &c_b * X + &b_b * (&int_M +
&a_b * X + T_u1) + T_u2;
		if new_M = . then delete;
	* Export bootstrap resample to a text file for Mplus;
	data _null_ ;
```

```
                        set Bootstrap_data;
                FILE  'C:\Users\tvpham2\Desktop\Boostrapdata.txt' ;
                PUT Group X new_M new_Y;
                    If new_M = . then delete;
                    run;
                    * Run the Run MSEM model using bootstrap resample;
                    X call "C:\Program Files (x86)\Mplus\Mplus.exe"
"U:\Proposal\MediationMSEMboostrap.inp" "C:\Users\tvpham2\Desktop\MediationMSEMboostrap.out";
                    * Import indirect effect estimate of bootstrap sample;
                    data MSEM_bootstrap;
                infile 'C:\Users\tvpham2\Desktop\MediationMSEMboostrap.out' truncover scanover flowover;
                    input @ 'INDB' ind_effect;
                    %if &bti = 1 %then %do;
                        Data All_MSEMbootstrap;
                            set MSEM_bootstrap;
                    %end;
                    %else %do;
                        Data All_MSEMbootstrap;
                            set All_MSEMbootstrap MSEM_bootstrap;
                    %end;
                %end; * End of Bootstrap loop;
                data All_MSEMbootstrap;
                    set All_MSEMbootstrap;
                    if ind_effect = . then delete;
                proc sql noprint;
                    select count(*) into: n_obs
                        from All_MSEMbootstrap;
                quit;
                %let _HL = round(0.975*&n_obs);
                %let _LL = round(0.025*&n_obs);
                Proc sort data = All_MSEMbootstrap;
                    by ind_effect;
                * Get the estimate of a coefficient;
                data Low_limit;
                    set All_MSEMbootstrap;
                    if _N_ = &_LL then MSEM_LL = ind_effect;
                    if MSEM_LL  = . then delete;
                    keep MSEM_LL;
                data High_limit;
                    set All_MSEMbootstrap;
                    if _N_ = &_HL then MSEM_HL = ind_effect;
                    if MSEM_HL = . then delete;
                    keep MSEM_HL;
                data &Bootstrap_estimate;
                    merge &ModelEstimates Low_limit High_limit;
                run;
                * Get the estimate of a coefficient;
                data &Bootstrap_estimate;
                    merge &ModelEstimates Low_limit High_limit;
                run;
            %end; * End of if checking for model status;
            %if &status = 1 %then %do;
                data &Bootstrap_estimate;
                    set &Bootstrap_estimate;
                    MSEM_LL = .;
                    MSEM_HL = .;
```

```
		%end;
	%MEnd;

	%Macro SimAll;
		%do i = 1 %to &reps;
			Data Simdata;
				set two;
				if rep_i = &i;
			* For MSEM model;
			%MSEM(SIMdata, MSEMestimates, MSEMresid_e, MSEMresid_u);
			%MSEMRescale(MSEMresid_e, MSEMresid_u, MSEMresid_ce, MSEMresid_cu);
			%MSEMTrans(MSEMresid_ce, MSEMresid_cu, MSEMestimates, Trans_resMe, Trans_resMu,
Trans_resYe, Trans_resYu);
			%MSEMBootstrap(Trans_resMe, Trans_resMu, Trans_resYe, Trans_resYu, MSEMestimates,
MSEMBootstrap_estimate);
			data MSEMRep_result;
				set MSEMBootstrap_estimate;
				keep MSEM_IDE MSEMStatus MSEM_LL MSEM_HL;
			%if &i = 1 %then %do;
				Data AllReps_MSEM;
					set MSEMRep_result;
			%end;
			%else %do;
				Data AllReps_MSEM;
					set AllReps_MSEM MSEMRep_result;
			%end;
			run;
		%end;
	%MEnd;

	filename junk dummy;
	proc printto log = junk print = junk;run;

	%SimAll;run;

	proc printto print = 'U:\Proposal\MSEMOutcome\MSEMSim44111ttt.txt';run;

	Data AllReps_MSEM1;
		set AllReps_MSEM;
			Population_ME = &pme;
			if Population_ME = 1 then P_IDE = 0;
			if Population_ME = 2 then P_IDE = 0.03;
			if Population_ME = 3 then P_IDE = 0.12;
			if Population_ME = 4 then P_IDE = 0.24;
			MSEM_bias = MSEM_IDE - P_IDE;
			MSEM_CI_width = MSEM_HL - MSEM_LL;
			if (MSEM_HL = .) or (MSEM_LL = .) then delete;
			MSEM_reject = 0;
			if (MSEM_HL < 0 ) and (MSEM_LL < 0) then MSEM_reject = 1;
			if (MSEM_HL > 0 ) and (MSEM_LL > 0) then MSEM_reject = 1;
			if MSEMstatus = 1 then delete;   * Delete the non convergence replication;
	* Sumarize the outcome from replications;
	* nObs - the convergence rate: = nObs/N_rep;
	* reject - the rejection rate of null hypothesis, which is no mediation effect - For Type I error and Power analysis;
	* bias - the avarage bias of replications in this condition;
	* CI_width - the avarage width of Confidence Interval;
```

```
Proc SQL;
                    Create table MSEM_result as select count(*) as MSEM_nObs,

          mean(MSEM_IDE) as MSEM_IDE,

          mean(MSEM_bias) as MSEM_bias,

          mean(MSEM_CI_width) as MSEM_CI_width,

          mean(MSEM_reject) as MSEM_reject,
                                                                                                                        mean(P_IDE) as
P_IDE
                                                                                                                        from
AllReps_MSEM1;
          Quit;
data Final;
          set MSEM_result;
          N_Rep = 500;
          N_Bootstrap = 1000;
          Population_ME = &pme;
          Population_shape = &Pshape;
          Population_grp = &size2;
          Population_cell = &size1;
          Population_ICC = &Picc;
Proc print data = final HEADING = H;run;
```

Mplus code to fit the MSEM model

```
TITLE: 2-1-1 mediation (MSEM)

DATA: FILE IS C:\Users\tvpham2\Desktop\Simdata.txt;

VARIABLE: NAMES ARE group x m y;

USEVARIABLES ARE group x m y;

BETWEEN IS x;

CLUSTER IS group;

ANALYSIS: TYPE IS TWOLEVEL RANDOM; ESTIMATOR=ML;

MODEL:

%WITHIN%

m y;

y ON m;

%BETWEEN%
```

```
x m y;
m ON x(a);
y ON m(b);
y ON x;
MODEL CONSTRAINT:
NEW(indb);
indb=a*b;
SAVEDATA: File = C:\Users\tvpham2\Desktop\FS.txt;
          SAVE =  FSCORES;
```