TESTING THE EFFECTIVENESS OF VARIOUS COMMONLY USED FIT INDICES FOR DETECTING MISSPECIFICATIONS IN MULTILEVEL STRUCTURAL EQUATION MODELS

A Dissertation

by

HSIEN-YUAN HSU

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

DOCTOR OF PHILOSOPHY

December 2009

Major Subject: Educational Psychology

TESTING THE EFFECTIVENESS OF VARIOUS COMMONLY USED FIT INDICES FOR DETECTING MISSPECIFICATIONS IN MULTILEVEL STRUCTURAL EQUATION MODELS

A Dissertation

by

HSIEN-YUAN HSU

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

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Oi-Man Kwok
	Victor Willson
Committee Members,	Bruce Thompson
	Michael Speed
Head of Department,	Victor Willson

December 2009

Major Subject: Educational Psychology

ABSTRACT

Testing the Effectiveness of Various Commonly Used Fit Indices for Detecting Misspecifications in Multilevel Structural Equation Models. (December 2009)

Hsien-Yuan Hsu, B.B.A., Nan-Hua University; M.S., National Taiwan University of Science and Technology

Co-Chairs of Advisory Committee: Dr. Oi-Man Kwok
Dr. Victor Willson

Two Monte Carlo studies were conducted to investigate the sensitivity of fit indices in detecting model misspecification in multilevel structural equation models (MSEM) with normally distributed or dichotomous outcome variables separately under various conditions. Simulation results showed that RMSEA and CFI only reflected within-model fit. In addition, SRMR for within-model (SRMR-W) was more sensitive to within-model misspecifications in factor covariances than pattern coefficients regardless of the impact of other design factors. Researchers should use SRMR-W in combination with RMSEA and CFI to evaluate the within-mode. On the other hand, SRMR for between-model (SRMR-B) was less likely to detect between-model misspecifications when ICC decreased. Lastly, the performance of WRMR was dominated by the misfit of within-model. In addition, WRMR was less likely to detect the misspecified between-models when ICC was relative low. Therefore, WRMR can be used to evaluate the between-model fit when the within-models were correctly specified and the ICC was not too small.

DEDICATION

To my mom, Mrs. Chin-Yueh Huang,

and those who helped me throughout my research work

ACKNOWLEDGEMENTS

I would like to express profound gratitude to my committee co-chair and advisor, Dr. Oi-man Kwok, for his encouragement and supervision throughout my dissertation study. His invaluable support accounts for more than 70% of my accomplishment (i.e., a huge effect size).

I would like to extend very special thanks to my co-chair, Dr. Victor Willson, and committee members, Dr. Bruce Thompson and Dr. Michael Speed, for their continuous guidance throughout the course of this research. Also, I wish to express my appreciation to Dr. Michael Benz and Dr. Dalun Zhang who, as my project advisors, helped me in the job search process.

A very special thanks to Mark's angels, Sandra and Susan, and Tze-li in Florida State University, who always stood by me. Many thanks to my friends, Bizarre, Chimmy, Eunju, Eric, Eun Sook, Gary, Goody, Jerry, John, June, Karen, Malena, Minjung, Myunghee, Peter, Ross, Rosy, Russell, Stephen, Vivian, Yan, and Yu-lin, for making my time at Texas A&M University a great experience.

Finally, my sincere thanks go to my family, especially my mom, for their support. My mom has devoted her whole life to helping me become the person I am today. I would never have obtained my doctoral degree without her endless love. I love you, Mom.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLE Page

CHAPTER I

INTRODUCTION

The methodological issue of analyzing hierarchical (or multilevel) structure data have prompted the growing development of multilevel modeling techniques over the past two decades (Heck, 2001). In educational and psychological research, two multilevel modeling techniques, namely hierarchical linear modeling (HLM) and multilevel structural equation modeling (MSEM), are widely adopted to analyze hierarchical structure data. HLM is referred to as an approach which facilitates the specification of univariate models (i.e., models with one outcome variable), while MSEM can be used to investigate “a wide range of multilevel, multivariate models” (Heck & Thomas, 2008, p. 100).

HLM primarily concerns the decomposition of variance in an univariate outcome variable into its within-group (e.g., student) and between-group (e.g., school) components and investigates the explained variances with sets of predictor variables existing in within-group or between-group levels (Heck & Thomas, 2008; Raudenbush & Bryk, 2002). However, HLM is restricted by its intrinsic inflexibility. First, HLM does not allow multivariate outcome variables at two or more levels to be explained by predictor variables. Second, HLM does not allow any latent variable underling the

This dissertation follows the style of *Educational & Psychological Measurement*.

observed variables to be included in the analysis (Heck & Thomas, 2008). In other words, the variables of interest in HLM are assumed to have no measurement errors (i.e., all variables are perfectly reliable). Such limitations put bounds to the application of HLM in educational and psychological research.

In comparison to HLM, MSEM refers to a more flexible approach in a general latent variable framework (Muthén & Asparouhov, 2009). The superiority of MSEM over HLM is showed as follows. First, MSEM incorporates measurement error in defining constructs through their observed indicators. More accurate estimates of the structural relationship between variables can be obtained after we correct for unreliability (Rowe, 2003). Second, a wide variety of theoretical multivariate models including latent and measured variables can be specified, tested, and compared (Heck & Thomas, 2008). Third, it is possible to evaluate the direct, indirect, and total effects operating among outcome variable(s) and predictor variable(s) simultaneously (Rowe, 2003). Because of the advantages of MSEM, it has been widely applied by many researchers across educational and psychological disciplines (e.g., Branum-Martin, et al., 2006; Cheung & Au, 2005; Duncan, Alpert, & Duncan, 1998; Dyer, Hanges, & Hall, 2005; Everson & Millsap, 2004; Gottfredson, Panter, Daye, Allen, & Wightman, 2009; Heck, 2001).

Unlike HLM, MSEM requires researchers to address how well the hypothesized model can reproduce the relations found in the sample data. MSEM is a confirmatory method that requires one to hypothesize a within-model and a between-model (i.e., model specification), and most importantly, the model fit needs to be justified (i.e., model evaluation). In MSEM, besides the overall model chi-square test ($\mathcal{X}^2$), RMSEA, CFI, SRMR, and WRMR are commonly reported in studies and are available in many standard statistical programs. RMSEA, CFI, and WRMR are global fit indices to reflect the degree of misfit for both within-model and between-model jointly (i.e., entire model). On the other hand, SRMR can be computed separately for the within-model (SRMR-W) and the between-model (SRMR-B). Up to now, SRMR-W and SRMR-B are the only fit indices that can be used to assess the within-model and between-model, respectively. In MSEM, most researchers rely mainly on traditional fit indices (e.g., RMSEA, CFI, SRMR, and WRMR), along with commonly used cutoff values proposed by Hu and Bentler (1999) or Yu (2002) as guidelines to justify the adequacy of hypothesized models.

Several potential problems associated with the application of RMSEA, CFI, SRMR-W, SRMR-B and WRMR in MSEM arise. First, as global fit indices, RMSEA and CFI might be more sensitive to misspecified within-models rather than misspecified between-models. Both RMSEA and CFI are a function of $\mathcal{X}^2$. However, the value of $\mathcal{X}^2$ is weighted differentially depending on the sample size at within-model and between-model. Generally, the sample size of the within-model (i.e., total sample size minus the number of groups) is a lot larger than the sample size of the between-model (i.e., the number of groups). Thus, the value of $\mathcal{X}^2$ as well as RMSEA and CFI are expected to be dominated by the within-model (Hox, 2002; Ryu & West, in press). Very few empirical studies have been conducted to evaluate whether RMSEA and CFI are sensitive to the entire model or the within-model only. This problem is in urgent need of investigation because researchers have treated RMSEA and CFI as global fit indices and believed these fit indices can indicate the degree of entire model fit in MSEM.

Second, the findings and cutoff values suggested in the previous studies regarding SRMR were based on simulated single-level data (i.e., data with independent observations) in conventional SEM and may not be generalized directly to MSEM (i.e., data with non-independent observations). Previous studies have shown that SRMR is more sensitive to misspecifications in the factor covariances and less sensitive to misspecifications in the pattern coefficients (e.g., Fan & Sivo, 2005; Hu & Bentler, 1998, 1999). However, in MSEM, SRMR is not a global fit index but can be computed separately for the within-model (SRMR-W) and the between-model (SRMR-B). No empirical study has been conducted to investigate the previous findings and recommended cutoff values are still applicable for SRMR-W and SRMR-B. Therefore, one cannot assume that SRMR-W and SRMR-B perform in a manner consistent with SRMR in conventional SEM. The performances of SRMR-W and SRMR-B need to be examined using simulated hierarchical structure data.

Third, the WRMR has been used in MSEM without a clear understanding of how it performs. WRMR is suitable to evaluate models with non-normally distributed outcomes (Muthén & Muthén, 1998-2007). There appears to be only one study (Yu, 2002) which evaluated the effectiveness of WRMR. However, in Yu’s (2002) study, the WRMR was assessed based on simulated single-level data with independent observations. It seems that the WRMR has not been extensively investigated in MSEM. Even though, some researchers have applied the WRMR cutoff values proposed by Yu (2002) to evaluate their hypothesized multilevel structural models with non-normally distributed outcome variables (e.g., Gottfredson, et al., 2009). To my knowledge, no study has comprehensively examined the effectiveness of WRMR in detecting misspecification in MSEM. The investigation of WRMR in MSEM is urgent to provide empirical implications when MSEM with non-normally distributed outcome variables is used.

Clearly, limited effort has been made to evaluate the effectiveness of various commonly used model fit indices for detecting misspecification in MSEM. Whether these fit indices and the corresponding cutoff values are still applicable for evaluating multilevel models is questionable. The purpose of my dissertation is to investigate the sensitivity of commonly used fit indices (i.e., RMSEA, CFI, SRMR-W, SRMR-B and WRMR) in detecting model misspecifications in two-level models with normally distributed or dichotomous (non-normally distributed) outcome variables separately under different conditions, including: number of groups in between-models, group size, and Intra-class Correlation (ICC), and model misspecification.

The dissertation is organized as followed. In Chapter II, the rationale of MSEM with normally distributed and dichotomous outcome variables, and some commonly used fit indices (i.e., RMSEA, CFI, SRMR and WRMR) were reviewed. Chapter III introduces two studies (i.e., Study 1 and 2) as investigations of some commonly used fit indices' sensitivity to model misspecifications in MSEM. Chapter IV and Chapter V containe the method, analysis, results and discussion of Study 1 and Study 2, respectively. Finally the conclusion was made in Chapter VI.

CHAPTER II

LITERATURE REVIEW

Multilevel structural equation modeling (MSEM) has been widely adopted for analyzing hierarchical structure data (i.e., data with non-independent observations) (Muthén & Asparouhov, 2009). Muthén (1994) discussed a disaggregated multilevel covariance structure approach, which is a common way to analyze hierarchical structure data by specifying a between-group model (or between-model) and a within-group model (or within-model) simultaneously (Muthén, 1994; Rabe-Hesketh, Skrondal, & Zheng, 2007). Many standard SEM programs such as LISREL (Jöreskog & Sörbom, 1996) and MPLUS (Muthén & Muthén, 1998-2007) have specific routines that use this approach to analyze hierarchical structure data. In this chapter, the rationale of multilevel structural equation modeling with normally distributed and dichotomous outcome variables will be presented, followed by the review of some commonly used fit indices (i.e., RMSEA, CFI, SRMR, and WRMR).

MSEM WITH NORMALLY DISTRIBUTED OUTCOME VARIABLES

For simplicity, we consider a single-factor model for the two-level data with a total of N individuals (i) nested within G groups (g) (i = 1…N individuals, and g = 1…G groups):

$$Y_{ig} = \upsilon + \lambda \eta_{ig} + \epsilon_{ig},$$

where υ is a measurement intercept vector, λ is a pattern coefficient vector, η is the factor score, and ϵ is the residual vector (Muthén, 1994). Notice that in MSEM, the factor means of η_{ig} should be viewed as a random effect and can be specified as following:

$$\eta_{ig} = \alpha + \eta_{Bg} + \eta_{Wgi},$$

where α is the overall expected value for η_{ig}, η_{Bg} is a random factor component capturing the variation across groups, and η_{Wgi} is a random factor presenting the variation over individuals within their respective groups. The expectation of both η_{Bg} and η_{Wgi} equal to zero (Muthén, 1994). The total factor variance can be decomposed into a between-group variance and a within-group variance:

$$V(\eta_{ig}) = \Psi_T = \Psi_B + \Psi_W.$$

The proportion of total variance that lies between groups can be described by an "intra-class correlation" (ICC):

$$\Psi_B \,/\, (\Psi_B + \Psi_W).$$

ICC is the degree of similarity of the observations within the same group (Muthén & Satorra, 1995). If the ICC is close to zero, the groups are slightly different from each other and a simple regression analysis conducted at the micro level (e.g., student) would be adequate. On the other hand, if the ICC is away from zero, it suggests that students within groups are more homogeneous and groups are more different from each other. In this case, multilevel modeling approaches are needed to analyze the data (Heck & Thomas, 2008). Muthén and Satorra (1995) proposed using design effect which is also a function of ICC as a guideline of whether using MSEM to analyze multilevel data:

Design effect = 1+ (averaged cluster size-1)*ICC

If the ICC was 0 and/or averaged cluster size was 1, then the design effect would be 1. In this case, no variance appears in the between-school level. Design effects larger than 2 implied that the variance in between-group level accounts for a significant amount of the total variance. Thus the standard errors would be underestimated if multilevel modeling approaches are not used (Muthén & Satorra, 1995). However, Roberts (2007) argued that the inclusion of predictor variables into the null model could “create” noticeable sample dependence even though the ICC is small in the null model, and the use multilevel modeling depends on the sample dependence introduced by certain predictor variables rather than solely on the magnitudes of the ICC.

In this manner, the residual variance of ϵ_{ig} can be also decomposed into a between-group component and a within-group component,

$$V(\epsilon_{ig}) = \Theta_B + \Theta_W.$$

Thus, the multilevel covariance structure can be produced (Hox, 2002; Muthén, 1994; Muthén & Satorra, 1995),

$$V(Y_{ig}) = \Sigma_T = \Sigma_B + \Sigma_W,$$

where Σ_B is the corresponding population between-group covariance matrix,

$$\Sigma_B = \Lambda_B \Psi_B \Lambda_B{}' + \Theta_B,$$

and Σ_w is the population within-group covariance matrix,

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

Similarly, the same decomposition can be applied to sample data directly by dividing the sample covariance matrix (S_T) into a sample between-group covariance matrix (S_B) and a sample within-group covariance matrix (S_W) (Heck & Thomas, 2000; Hox & Maas, 2004; Muthén, 1994):

$$S_T = S_B + S_W.$$

Muthén (1989, 1994) proposed a slightly different version of equation:

$$S_T = S_B + S_{PW},$$

where S_T is the sample total covariance matrix, S_B is the sample between-group covariance matrix, and S_{PW} is the sample pooled within-group covariance matrix. The equations for the three sample covariance matrices are:

$$S_T = (N\text{-}1)^{-1}\sum_{g=1}^{G}\sum_{i=1}^{Ng}(y_{gi}-\overline{y})(y_{gi}-\overline{y})',$$

$$S_B = (G\text{-}1)^{-1}\sum_{g=1}^{G}N_g(\overline{y_g}-\overline{y})(\overline{y_g}-\overline{y})',$$

$$S_{PW} = (N\text{-}G)^{-1}\sum_{g=1}^{G}\sum_{i=1}^{Ng}(y_{gi}-\overline{y_g})(y_{gi}-\overline{y_g})',$$

where G is the number of groups, N_g is the corresponding group size for the g-th group, and N is the total sample size.

Muthén (1994) indicated that the sample total covariance matrix (S_T) is an unbiased maximum likelihood (ML) estimator of the population total covariance matrix (Σ_T). The sample pooled-within covariance matrix (S_{PW}) is also an unbiased ML estimator of the population Σ_W with sample size *N*-*G* (Hox & Maas, 2004; Muthén & Satorra, 1995), while the sample between-group matrix (S_B) is an unbiased ML estimator of ($\Sigma_W + c\Sigma_B$)

with sample size G, and c is the average-like group size under the unbalanced design condition:

$$c = \left[N^2 - \sum_{g=1}^{G} N_g^2 \right] [N\,(G\text{-}1)]^{-1}.$$

Estimation for MSEM with Normally Distributed Outcome Variables

Full information maximum Likelihood (FIML) estimation "maximizes a likelihood fitting function that is the sum of *n* casewise likelihood functions" (Enders, 2001, p. 714). In MSEM, FIML estimation via expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977) facilitates the analysis of multilevel structure data including both continuous and categorical outcome variables (Heck & Thomas, 2008; Raudenbush & Bryk, 2002). Moreover, random slopes and intercepts in structural equation models can be estimated (Kaplan, 2008). Recently, several EM algorithm-based FIML estimators have been adopted in some statistical packages such as MPLUS (Muthén & Muthén, 1998-2007).

Maximum likelihood (ML) fitting function under normality for two-level covariance structural analysis (Muthén & Satorra, 1995):

$$G\{log\,|\Sigma_W + c\Sigma_B\,| + trace[(\Sigma_W + c\Sigma_B)^{-1}S_B] - log|S_B| - p\} + (N - G)\,\{log|\Sigma_W| + trace[\Sigma_W^{-1}] - log|S_{PW}| - p\},$$

where "| |" indicates the determinant of a matrix. "p" is the total number of observed variables in the model.

Since the sample total covariance matrix can be partitioned into a sample between-group covariance matrix (S_B) and a sample within-group covariance matrix (S_W), multilevel structure data can be analyzed by estimating both the between-group model and the within-group model simultaneously. However, we cannot construct and test a model with only Σ_B because the sample between-group covariance matrix S_B is not the unbiased ML estimator of the population Σ_B but the combination of both Σ_W and $c\Sigma_B$ (Muthén, 1994). Thus, both within-group and between-group models must be fitted jointly and simultaneously (Hox & Maas, 2004; Rabe-Hesketh, et al., 2007).

MSEM WITH DICHOTOMOUS OUTCOME VARIABLES

In MSEM with dichotomous outcome variables, the sample matrix for analysis is not a conventional covariance or correlation matrix but a tetrachoric correlation matrix (Muthén, 1993; West, Finch, & Curran, 1995). Thus, only the correlations between variables are considered and the degrees of freedom for p measured variables are the number of parameters in the unrestricted model (p (p-1) /2) minus the number of free parameters (Muthén, 1993).

For simplicity, we consider a two-level single-factor model. Let y_{pig} denote p^{th} dichotomous measured variable (i.e., latent variable indicator) for individual i nested within in g group ($p = 1 \ldots P$ dichotomous variables, $_i = 1 \ldots N$ individuals, and $g = 1 \ldots G$ groups),

$$y_{pig} = \begin{cases} 1, \text{ if } \tau < y^*_{pig} \\ \\ 0, \text{ if } y^*_{pig} \leq \tau. \end{cases}$$

The equation expresses a threshold model which assumes that underlying the measured dichotomous variable y_{pig} is a normally-distributed continuous latent variable y^*_{pig}, which can determine the category of the measured dichotomous variable by the threshold (τ) (Asparouhov & Muthén, 2007; Bollen, 2002). In other words, "the variables of interest are conceptualized as continuous, but the response format administrated allows respondents to answer only in a restrictive, dichotomous scale" (Bollen, 2002, p. 620). For example, if the i^{th} individual falls short of the threshold, the response of this individual would be "0". On the other hand, if the i^{th} individual passes this threshold, the response of this individual would be "1".

Thus, similar to the MSEM with normally distributed indicators, a two-level model with dichotomous indicators can be expressed,

$$y^*_{pig} = \nu + \lambda \eta_{ig} + \epsilon_{ig},$$

Where y^*_{pig} contains now the continuous outcome latent factors, ν is a measurement intercept vector, λ is a pattern coefficient vector, η is the factor score, and ϵ is the residual vector (Muthén, 1994). Notice that in MSEM, the factor means of η_{ig} should be viewed as a random effect and can be specified as following:

$$\eta_{ig} = \alpha + \eta_{Bg} + \eta_{Wgi},$$

where α is the overall expected value for η_{ig}, η_{Bg} is a random factor component capturing the variation across groups, and η_{Wgi} is a random factor presenting the variation over individuals within their respective groups. The expectation for both η_{Bg} and η_{Wgi} still equal to zero (Muthén, 1994). The total factor variance can then be decomposed into a between-group variance and a within-group variance:

$$\mathrm{V}(\eta_{ig}) = \Psi_T = \Psi_B + \Psi_W.$$

In the same manner, the residual variation of ϵ_{ig} can be also decomposed into a between-group component and a within-group component,

$$V(\epsilon_{ig}) = \Theta_B + \Theta_W.$$

Thus, the multilevel covariance structure can be produced as same as in the MSEM with continuous indicators (Hox, 2002; Muthén, 1994; Muthén & Satorra, 1995),

$$V(y^*_{pig}) = \Sigma_T = \Sigma_B + \Sigma_W,$$

where Σ_B is the corresponding population between-group covariance matrix,

$$\Sigma_B = \Lambda_B \Psi_B \Lambda_B' + \Theta_B,$$

and Σ_w is the population within-group covariance matrix,

$$\Sigma_W = \Lambda_W \Psi_W \Lambda_W' + \Theta_W.$$

Residual variances of the latent response variables (y^*_{pig}) underling the dichotomous indicators are fixed at one for identification purpose in line with the Theta parameterization (Asparouhov & Muthén, 2007; Muthén & Muthén, 1998-2007).

Estimation for MSEM with Dichotomous Outcome Variables

The estimation of multilevel models with dichotomous outcome variables via the EM algorithm is computationally demanding because all random effects have to be numerically integrated (Asparouhov & Muthén, 2007). Asparouhov and Muthén (2007) extended Muthén's (1984) weighted least squares (WLS) estimation and proposed a limited- information weighted least squares estimation method (WLSM) that can be used to estimate two-level structural equation models with dichotomous, ordered polytomous, censored, and continuous outcome variables as well as combinations of such variables. The WLSM method uses high dimensional integration by multiple simple models with one and two dimensional integration (Asparouhov & Muthén, 2007).

The WLSM estimator consists of three stages (Asparouhov & Muthén, 2007; Muthén, 1984). In the first stage, all the parameters of the p^{th} univariate model are estimated using the two-level maximum likelihood (ML) method except for the off diagonal estimates of Σ_W and Σ_B. In the second stage, parameters for every pair of bivariate models given first stage estimates are estimated. Lastly, the model parameters are estimated by minimization of the weighted least squares fitting function with p measured variables:

$$F_{WLS} = (s - s^*) \; W \; (s - s^*)',$$

where s is the $p*$ x 1vector of all parameter estimates of the unrestricted model, $s*$ is the corresponding $p*$ x 1 vector of all standardized model-implied estimates, and W is the

p^*x p^* weighted matrix. Here p^* is defined as $p\ (p+1)/2$. Note that matrix W is G^{-1} where G is the asymptotic covariance of s (Asparouhov & Muthén, 2007; West, et al., 1995).

Asparouhov and Muthén (2007) conducted a Monte Carlo study to compare the effectiveness of ML and WLSM estimators. Asparouhov and Muthén found that (a) the parameter estimates of WLSM estimator were more efficient and less biased, (b) model estimation with WLSM estimator is more likely to be converged, and (c) the Type I error rate of chi-square statistics equals to 6% which is reasonable. In other words, the WLSM estimator outperforms the ML estimator. The WLSM estimator should be used when non-normally distributed outcome variables are included in the multilevel models. The WLSM estimator is implemented in the MPLUS 5.2 (Muthén & Muthén, 1998-2007).

COMMONLY USED FIT INDICES IN MSEM

The goodness of model fit is one of the primary questions in MSEM. The chi-square test ($\mathcal{X}^2$) and fit indices (e.g., RMSEA, CFI, SRMR, and WRMR) are used to evaluate whether empirical data supports a theoretical model. Many studies on fit indices for evaluating structure equation models have been conducted since 1980's in order to create clear guidelines[1] for SEM analyses (e.g., Chen, Curran, Bollen, Kirby, & Paxton, 2008; Fan & Sivo, 2005, 2007; Fan, Thompson, & Wang, 1999; Hu & Bentler, 1998, 1999; Marsh, Balla, & Hau, 1996; Sivo, Fan, Witta, & Willse, 2006). Up to the present,

[1] Although model fit indices in conventional SEM has been an object of study for a long time, there is still some disagreement as to the cutoff values for fit indices (Marsh, Hau, & Grayson, 2005; Marsh, Hau, & Wen, 2004).

besides the overall model $\mathcal{X}^2$, two types of goodness of fit indices for evaluating structure equation models are commonly reported in studies and are available in many standard statistical programs: absolute fit indices (e.g., RMSEA and SRMR), and relative fit indices (e.g., CFI).

According to the survey by McDonald and Ho (2002), both RMSEA and SRMR are among the most reported absolute fit indices, while CFI is the most reported relative fit index. A similar trend of reporting fit indices was also found in studies using confirmatory factor analysis. Jackson, Gillaspy, and Purc-Stephenson (2009) found that other than the overall model chi-square test, both RMSEA and CFI are the most reported fit indices in their literature review.

Another fit index, namely WRMR, was proposed by Muthén & Muthén (1998-2007). The advantage of WRMR is that it can be used to assess models with non-normal continuous outcome variables or dichotomous outcome variables. Yu (2002) has investigated the effectiveness of WRMR with various single-level models. In the PsycINFO database through May 2009, we found that 22 journal articles analyzing single-level data and 1 journal article analyzing multilevel-level data had cited Yu' study and applied the WRMR cutoff values proposed by Yu (2002) to evaluate their hypothesized models with non-normal continuous outcome variables or dichotomous outcome variables.

Brief Review of Chi-Square Test, RMSEA, CFI, SRMR and WRMR

Chi-Square Test

The chi-square test ($\mathcal{X}^2$) is probably the most commonly used statistical model fit index (Heck & Thomas, 2008). The $\mathcal{X}^2$ in SEM is used to test the null hypothesis that the unrestricted population covariance matrix of the observed variables Σ is equal to the model implied covariance matrix $\Sigma(\theta)$ (i.e., $H_0: \Sigma = \Sigma(\theta)$) (Mueller, 1996). However, we do not know the elements of Σ but can estimate them by elements in sample covariance matrix S. Similarly, we do not know the model parameters in the vector θ and must estimate these coefficients (i.e., $\hat{\theta}$) (Mueller, 1996). Thus, we can test the null hypothesis $\mathrm{H}_0: \Sigma = \Sigma(\theta)$ given the $\mathcal{X}^2$ value and *df* of the specified model. If we fall to reject the null hypothesis, we can conclude that the specified model leads to a precise reproduction of the population covariance matrix of the observed variables (Bollen & Long, 1993; Steiger, 2007).

However, several deficiencies of $\mathcal{X}^2$ have been noted. First, $\mathcal{X}^2$ test relied on many assumptions (e.g., multivariate normality of observed data, large sample size), nevertheless, not all can be met completely in practical applications (Mueller, 1996). Second, the value of $\mathcal{X}^2$ is affected by sample size. Large samples resulted in a rejection of the null hypothesis even trivial deviations between the covariance matrix implied by a specific model and the population covariance matrix of the observed variables (Bollen & Long, 1993). Due to the limitations of $\mathcal{X}^2$ test, other fit indices have been developed to evaluate model fit.

Root Mean Square Error of Approximation (RMSEA)

RMSEA (Steiger & Lind, 1980, May) is a fit index based on the population noncentrality parameter:

$$\text{RMSEA} = \sqrt{max(\frac{[\mathcal{X}^2 \ for \ the \ target \ model \ /df \ for \ the \ target \ model \ -1]}{the \ number \ of \ observations \ -1}, 0)},$$

where $\mathcal{X}^2$ for the target model is related to the discrepancy between the observed variance-covariance matrix and the model-implied variance-covariance matrix (Kline, 2005). Hence, RMSEA can be viewed as a measure of the average discrepancy between the observed and model-implied variance-covariance matrices per degree of freedom with the model complexity taken into account (Browne & Cudeck, 1993; Steiger, 2007). RMSEA is bounded by zero and the smaller the RMSEA indicates a better fit of the model to the data. A RMSEA equal to zero indicates that the target model fits the data perfectly. RMSEA is a global fit index that reflects the degree of fit (or misfit) for the entire model. If RMSEA performs in a similar way to the single level SEM, any misspecification in either within-model or between-model should be reflected by the RMSEA (with a large value).

Comparative Fit Index (CFI)

Bentler (1990) proposed CFI which is strictly bounded by 0 and 1. CFI shows the relative goodness of fit of a particular hypothesized model compared with a baseline model in which all covariances/correlations between any pair of variables are set to zero (Bentler, 1990; Tanaka, 1993). By using the baseline model's chi-square test statistic (*T*0), the corresponding model's degree of freedom (*df*0), the tested model's chi-square test statistic (*T*1), and the tested model's degree of freedom (*df*1), we can obtain $\tilde{\lambda}_i$ (i.e., *T*0 – *df*0) for the baseline model and $\tilde{\lambda}_k$ (i.e., *T*1 – *df*1) for the tested model. Then, CFI can be computed as:

$$\text{CFI} = 1 - \frac{\hat{\lambda}_k}{\hat{\lambda}_i},$$

with $\hat{\lambda}_i = \max(\tilde{\lambda}_i, \tilde{\lambda}_k, 0)$ and $\hat{\lambda}_k = \max(\tilde{\lambda}_k, 0)$. Because $\hat{\lambda}_i \geq \hat{\lambda}_k \geq 0$, CFI ranges between .00 and 1.00 (Bentler, 1990; Tanaka, 1993). Larger value of CFI (e.g., larger than .95) indicates a good fit of the model. Similar to RMSEA, CFI is expected to reflect the misfit for the entire multilevel model.

Standardized Root Mean Square Residual (SRMR)

SRMR can be obtained using the following formula:

$$\text{SRMR} = \sqrt{\left\{2\sum_{i=1}^{p}\sum_{j=1}^{i}\left[\frac{(s_{ij}-\hat{\sigma}_{ij})}{s_{ii}s_{jj}}\right]^2\right\} / \, p(p+1)},$$

where s_{ij} is a sample covariance between variables *i* and *j*; $\hat{\sigma}_{ij}$ is the corresponding model-implied covariance between variables *i* and *j;* s_{ii} and s_{jj} are the sample standard deviations for the variables *i* and *j*, respectively; and *p* is the total number of variables in the model for analysis (Bentler, 1995). The discrepancy between the sample covariance and the corresponding model-implied covariance $(s_{ij} - \hat{\sigma}_{ij})$ indicates the degree of fit (or misfit). In MSEMs, the covariance matrices for the within-model and the between-model are computed separately. Unlike RMSEA and CFI, which are global fit index measures, SRMR can be computed separately for the within-model (SRMR-W) and the between-model (SRMR-B) in MSEM and can be used for evaluating the plausible misspecification at each level. Generally, smaller SRMR (e.g., less than .08) indicates a good fit of the model.

Weighted Root Mean Square Residual (WRMR)

WRMR is defined as

$$\text{WRMR} = \sqrt{\sum_r^e \frac{(s_r - \hat{\sigma}_r)^2}{\upsilon_r} / e},$$

where e is the number of sample statistics, s_r and $\hat{\sigma}_r$ are elements of the sample statistics vector and model-implied vector, respectively and υ_r is an estimate of the asymptotic variance of s_r. “WRMR is suitable for models where sample statistics have widely varying variances, and when sample statistics are on different scales such as in models with mean and/or threshold structures. WRMR is also suitable with non-normal continuous outcomes” (Muthén, 1998-2004, p. 24). WRMR is expected to reflect the misfit for the entire multilevel model. Smaller value of WRMR (e.g., smaller than .90) indicates a good fit of the model.

The Traditionally Recommended Cutoff Values

Cutoff values have been suggested for the four commonly reported fit indices, RMSEA, CFI, SRMR, and WRMR. For example, Browne and Cudeck (1993) recommended that RMSEA equal to or less than 0.05 indicates a model with adequate fit, while Hu and Bentler (1999) suggest that RMSEA equal to or less than 0.06 is needed to conclude a well-fitting model. Chen and her colleagues (2008) evaluated the performance of RMSEA alone versus that of using it jointly with its related confidence interval given a fixed cutoff point 0.05. Little evidence was found that the use of 0.05 or any other value as universal cutoff values can determine the fitness of a model. From their results, a cutoff value 0.10 is too liberal to be used.

Hu and Bentler (1999) also recommend that CFI equal to or larger than 0.95 is an indication of a good fit model. On the other hand, the recommended SRMR cutoff value for a good fit model by Hu and Bentler (1999) is equal or less than 0.08. Cutoff values for WRMR were only studied by Yu (2002). Yu suggested a cutoff value of 1.00 for models with normal and non-normal continuous outcomes. Furthermore, WRMR cutoff values close to 0.95 or 1.00 was suggested for models with dichotomous outcomes. In Muthén’s (1998-2004) technical report, Muthén indicated a WRMR smaller than 0.90 was recommended for good models with continuous as well as with categorical outcomes.

All these cutoff values are based on simulation studies with independent observations (i.e., single-level data). Whether these cutoff values are still applicable to assessing the model goodness of fit in multilevel structural equation models with non-independent observations is questionable, however.

CHAPTER III

INVESTIGATION OF FIT INDICES' SENSITIVITY IN MSEM

The current investigations (i.e., Study 1 and 2) were motivated by the limited effort has been made to evaluate the effectiveness of various commonly used model fit indices for detecting misspecification in MSEM. In Study 1, fit indices (i.e., RMSEA, CFI, SRMR-W and SRMR-B) were examined under multilevel confirmatory factor analysis (MCFA) models with normally distributed outcome variables (i.e., indicators). The design factors included (a) number of groups in between-models (150, 200, and 250), (b) group size (15 and 30), (c) Intra-class Correlation (high and low), and (d) misspecification type (true, misspecifications in factor covariance and in pattern coefficient).

Study 2 primarily investigated the sensitivity of a commonly used fit index, namely WRMR, in MSEM. WRMR was widely used to evaluate the degree of model fit when non-normal outcome variables were included in the models. In Study 2, WRMR as well as RMSEA, CFI, SRMR-W and SRMR-B were examined under MCFA models with dichotomous outcome variables (a type of non-normal outcome variables). The design factors included (a) number of groups in between-models (150, 200, and 250), (b) group size (15 and 30), (c) Intra-class Correlation (high and low), (d) threshold (0 and 1), and (e) model misspecification (true, misspecifications in factor structure).

CHAPTER IV

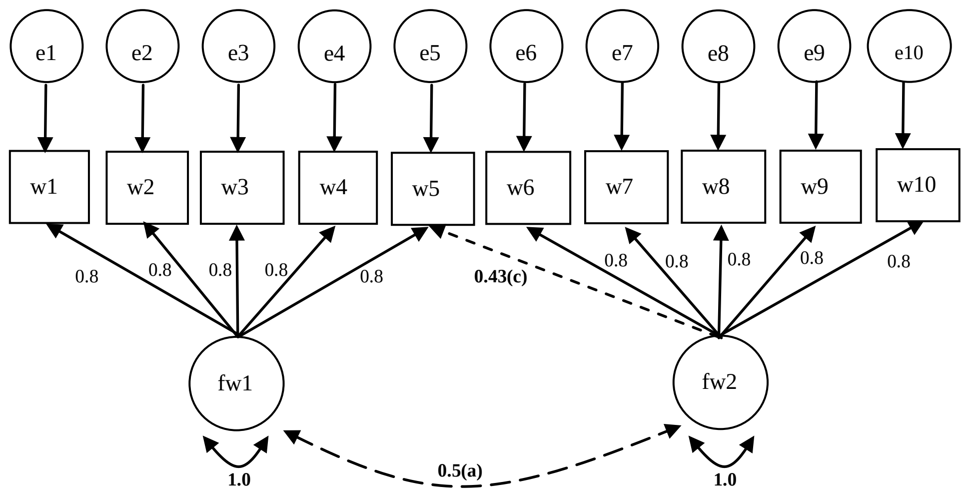
STUDY 1: THE SENSITIVITY OF FIT INDICES IN MSEM WITH NORMALLY DISTRIBUTED OUTCOME VARIABLES

As presented before, the goal of Study 1 was to examine the sensitivity of fit indices (i.e., RMSEA, CFI, SRMR-W and SRMR-B) under multilevel confirmatory factor analysis models with normally distributed outcome variables (i.e., indicators). The method used in Study 1 was introduced first, followed by the results and discussion.

METHOD

A Monte Carlo study was conducted using Mplus 5.1 (Muthén & Muthén, 1998-2007) to investigate the sensitivity of four commonly used fit indices (i.e., RMSEA, CFI, SRMR-W and SRMR-B) for detecting different types of model misspecifications in multilevel SEMs under various conditions. Confirmatory factor analysis (CFA) models, or measurement models, are commonly used in SEM related simulation studies (e.g., Hu & Bentler, 1998, 1999; Muthén & Muthén, 2002; Yuan & Bentler, 2002). Therefore, a multilevel confirmatory factor analysis (MCFA) model was employed for data generation in my study. As presented in Figures 4.1 and 4.2, both within- and between-models were specified as having the same factor structure with ten observed indicators loaded on two latent factors.

Within-model



Between-model

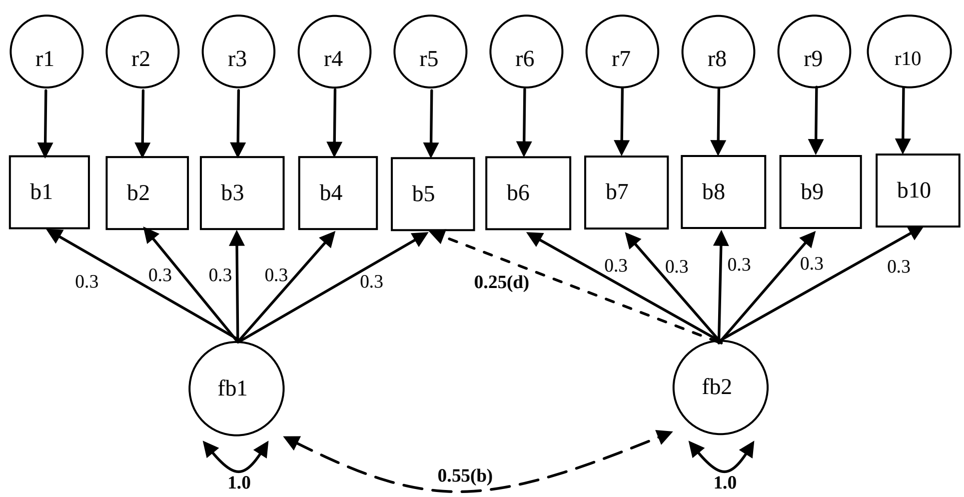
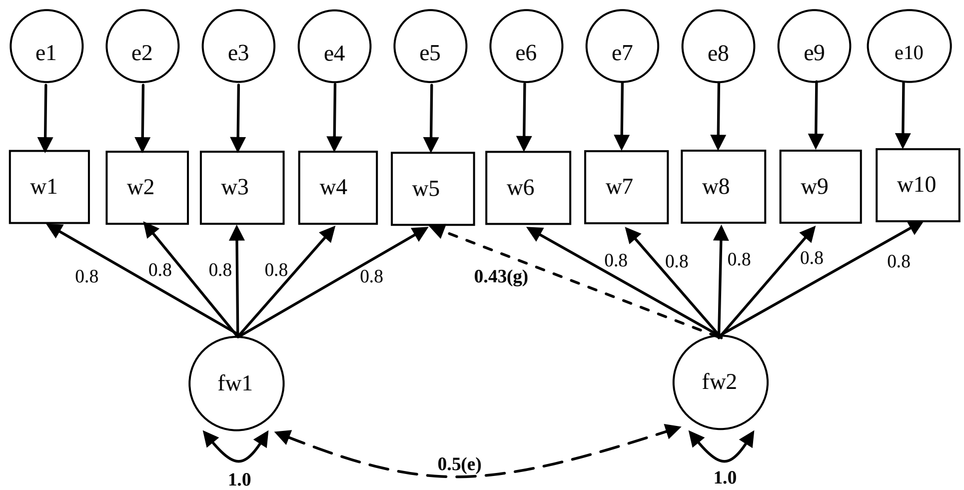


Figure 4.1 Multilevel CFA Model for Data Generation (Low ICC) in Study 1

Note. The dashed paths in the figures are omitted in the analysis to create the misspecified models.

Within-model



Between-model

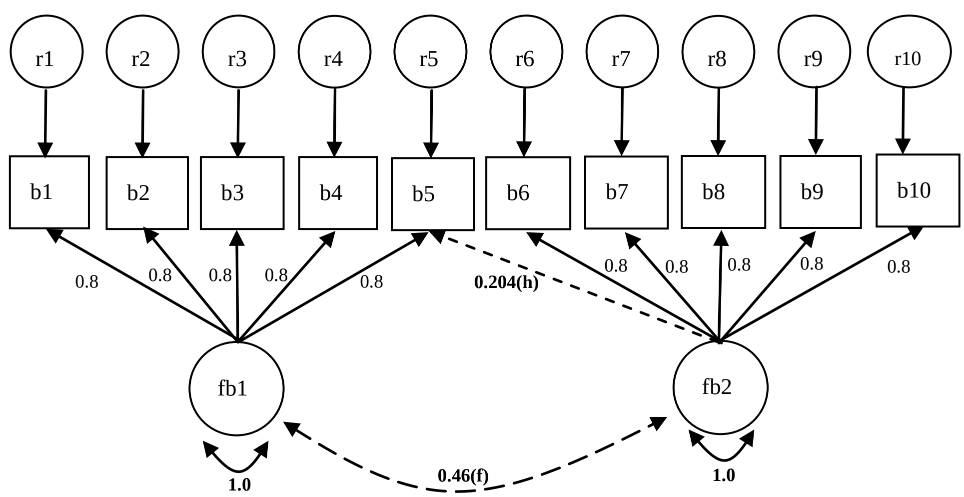


Figure 4.2 Multilevel CFA Model for Data Generation (High ICC) in Study 1

Note. The dashed paths in the figures are omitted in the analysis to create the misspecified models.

In DiStefano and Hess' (2005) empirical review of psychological assessment literature, the minimum number of indicators per latent variable was 4.2 and the maximum number was 6.9, so five indicators per latent variable were used here. Four design factors were considered in this study: number of groups (NG) at the between level, group size (GS), intra-class correlation (ICC), and the misspecification type (MT).

Number of Groups (NG). Number of groups relates to the estimation of the between-model. Hox and Maas (2001) concluded that large NGs (i.e., larger than 50 groups) were needed for acceptable estimates for a between-model with low ICC conditions. On the other hand, in SEM literature, two hundred observations seem to be the minimum required sample size for obtaining unbiased and consistent estimates when using the maximum likelihood estimation method (Boomsma, 1987; Loehlin, 2004). In the present study, three different NGs were adopted (i.e., 150, 200, and 250 groups) to evaluate whether NG level would affect fit indices performance in detecting model misspecification.

Group size (GS). Whereas size was problematic with NG, Hox and Maas (2001) concluded that small group size (GS) (i.e., 10 observations per group) could produce admissible within-model estimates . Thus, we used two levels of GS (i.e., 15 and 30 observations per group).

Intra-Class Correlation (ICC). The low ICC condition was created by giving lower values (0.3) for the pattern coefficients in the between-model (see Figure 4.1) while the high ICC condition was created by substituting higher values (0.8) for pattern coefficients in the between-model (see Figure 4.2). The pattern coefficients of the between-model considered in this study lead to two levels of ICC: 0.18 (low) and 0.40 (high).

Misspecification Types (MT). Two types of misspecification, the same employed by Hu and Bentler (1998, 1999), were adopted to examine fit index sensitivity to varying forms of model misspecification. The two MTs were under-parameterized factor covariance (i.e., simple misspecification) and under-parameterized pattern coefficient (i.e., complex misspecification). In simple misspecification, the covariance between two latent factors was constrained to 0 when the true covariance parameter was not equal to 0. Figure 4.1 demonstrates how the covariance parameter (i.e., *a*) in the within-model would be constrained to 0 in the simple misspecification condition. For complex misspecification, a pattern coefficient was constrained to 0 when the true pattern coefficient was not equal to 0. Pattern coefficient *c* in the within-model would be constrained to 0 for the complex misspecification.

As suggested by Fan and Sivo (2005, 2007) and Marsh, Hau, and Wen (2004), the severity of all the misspecification conditions was also taken into account. Fan and Sivo (2005, 2007) proposed that the statistical power for rejecting a misspecified model could be used as an indicator of the severity of the corresponding model misspecification. The magnitudes of different misspecification conditions (i.e., paths *a* to *h* in Figures 4.1 and 4.2) in the present study were adjusted based on Fan and Sivo's (2005, 2007) approach. The statistical powers of all the misspecification conditions were controlled to be close to 0.70.

Four different sets of simulation for misspecified models were conducted to evaluate the sensitivity of RMSEA, CFI, SRMR-W and SRMR-B: (a) no model misspecification (true model, M_T; see Figure 4.1 and 4.2), (b) model misspecification in the within-model only (M_W), (c) model misspecification in the between-model only (M_B), and (d) model misspecification in both within- and between-models (M_{WB}). The maximum likelihood estimation method with robustness to non-normality and non-independence of observations (MLR) was employed for all the analyses (Muthén & Muthén, 1998-2007).

ANALYSIS

Four design factors were included in Study 1: numbers of groups in between levels (NG; 150, 200 and 250), group size (GS; 15 and 30), intra-class correlation (ICC; low and high), and misspecification types (MT; simple, and complex). For the M_T model, factors integrated into 12 conditions (3 NG x 2 GS x 2 ICC) whereas for the M_W, M_B, and M_{WB} models, factors were integrated into 24 conditions (3 NG x 2 GS x 2 ICC x 2 MT). For each condition, 1,000 replications were generated and analyzed.

Replications with convergence problems or improper solutions (e.g., negative unique variances) were first excluded. Then, means and standard deviations (SDs) of each fit index were reported for the M_T model replications. A series of ANOVAs were conducted to determine the impact of the design factors on the effectiveness of the four targeted fit indices (i.e., RMSEA, CFI, SRMR-W, and SRMR-B) in identifying the true model. Similarly, for the M_W, M_B, and M_{WB} model replications, means and SDs of each fit index were reported for evaluation of the sensitivity of the four fit indices for detecting misspecifications. A series of ANOVAs were then conducted to examine whether the target fit indices were equally sensitive to different types of misspecification regardless of other design factors. Finally, the statistical powers based on the traditionally recommended cutoff values (i.e., RMSEA < 0.06, CFI > 0.95, SRMR < 0.08) were computed from the simulated data.

RESULTS

The purpose of this study was to evaluate the effectiveness of fit indices in detecting misspecified multilevel SEMs. Under M_T condition, theoretically, all the fit indices were expected to show that the specified models perfectly fitted the data regardless of the design factors (i.e., NG, GS or ICC). Under M_W, M_B and M_{WB} conditions, it was expected that all the fit indices should be sensitive to the model misspecifications regardless the design factors. In the end, the statistical power in detecting the misspecified models was examined using the traditionally recommended cutoff values of the four fit indices.

Convergence Failure and Improper Solutions

Table 4.1 presents the percentage of replications with convergence problems or improper solutions (e.g., negative unique variances) across all (six) sample size conditions. Results show that the number of problematic replications was generally small (less than 1%) in the simulation studies. All the problematic replications were not included in the analyses presented below.

Table 4.1 Percentage of Replications with Nonconvergence and Improper Solutions in Study 1

	Number of Groups x Group Size					
Misspecification	150X15	150X30	200X15	200X30	250X15	250X30
Simple						
Nonconvergence	0.016	0.000	0.010	0.000	0.005	0.000
Improper Solutions	0.019	0.000	0.011	0.000	0.006	0.000
Complex						
Nonconvergence	0.000	0.000	0.000	0.000	0.000	0.000
Improper Solutions	0.027	0.002	0.021	0.000	0.009	0.000

Performance of the Target Fit Indices on the Correctly Specified Models

Table 4.2 presents the means and SDs of the chi-square statistics ($\mathcal{X}^2$), RMSEA, CFI, SRMR-W and SRMR-B when fitting the correctly specified (or true) model (M_T). When the model was correctly specified, the mean of $\mathcal{X}^2$ would be close to the degrees of freedom of the model (*df* = 66). The mean (67.724) of $\mathcal{X}^2$ shown in Table 4.2 indicated that the simulation study was correctly conducted.

If the specified model fits the data perfectly, RMSEA, SRMR-W and SRMR-B are equal to 0 while CFI is equal to 1. The means of RMSEA and SRMR-W were close to 0 with trivial SDs, which indicated that both RMSEA and SRMR-W perform well under M_T condition. Similarly, CFI (with mean and SD equal to 1.000 and 0.000, respectively) performed extremely well under M_T condition. On the other hand, the performance of SRMR-B under M_T condition was not as good as other fit indices with a relatively large mean value and standard deviation (i.e., mean = 0.032; SD = 0.014).

Table 4.2 Descriptive Statistics of the Fit Indices of the True Models (M_T) in Study 1

Fit Index	Mean	SD
Chi-square	67.724	11.823
RMSEA	0.003	0.004
CFI	1.000	0.000
SRMR-W	0.007	0.002
SRMR-B	0.032	0.014

Note. n=12,000. Degrees of freedom of the hypothesized model were 66. RMSEA = root-mean-square error of approximation. CFI = comparative fit index. SRMR-W = standardized root mean square residual for within-model. SRMR-B = standardized root mean square residual for between-model.

Also examined was the possible impact of the three design factors (i.e., NG, GS, and ICC) on the performance of the four target fit indices separately under M_T condition using ANOVA. The results are presented in Table 4.3. The total sum of square (SOS) of each fit index showed the variability of the corresponding fit index across all true model replications, while eta-squared (η^2) demonstrated the proportion of the variance accounted for by a particular design factor or the interaction effect term. Note that η^2 was obtained by dividing the Type III sum of squares of a particular predictor or the interaction effect by the corrected total sum of squares.

As shown in Table 4.3, all the design factors (e.g., NG, GS, & ICC) only accounted for a trivial proportion of the total SOS of the model χ^2. Moreover, the total SOS's of RMSEA, CFI and SRMR-W were very small and the corresponding η^2 of the design factors was negligible for all three. On the other hand, compared with the other fit indices, the total SOS of SRMR-B was relatively large (2.352) and the ICC accounted for 70.25% of the total SOS of SRMR-B. The means of SRMR-B under low and high ICC conditions were 0.044 and 0.021, respectively.

Table 4.3 Eta-Squares (η^2) for the Fit Indices of the True Models (M_T) in Study 1

Sources	Fit Index				
	Chi-square	RMSEA	CFI	SRMR-W	SRMR-B
Total SOS	1677332.821	0.153	0.002	0.042	2.352
Overall η^2	0.22%	4.19%	7.32%	55.79%	79.02%
NG	0.14%	1.46%	2.40%	13.98%	6.00%
GS	0.03%	2.68%	4.61%	41.34%	1.77%
ICC	0.00%	0.00%	0.06%	0.00%	70.25%
NG*GS	0.01%	0.03%	0.25%	0.47%	0.03%
NG*ICC	0.01%	0.01%	0.00%	0.00%	0.87%
GS*ICC	0.00%	0.00%	0.00%	0.00%	0.09%
NG*GS*ICC	0.03%	0.01%	0.00%	0.00%	0.01%

Note. *n*=12,000. Degrees of freedom of the hypothesized model were 66. RMSEA = root-mean-square error of approximation. CFI = comparative fit index. SRMR-W = standardized root mean square residual for within-model. SRMR-B = standardized root mean square residual for between-model. SOS = sum of squares. NG = number of group. GS = group size. ICC = intra-class correlation.

Performance of the Target Fit Indices on the Misspecified Models

Table 4.4 presents the means and SDs of the $\mathcal{X}^2$ and the four target fit indices under models with different misspecification types across various simulation conditions. $\mathcal{X}^2$ had substantially large means and SDs under M_W or M_{WB} conditions. For example, under M_W condition, means and SDs of $\mathcal{X}^2$ under models with simple (mean=1022.868; SD=393.811) and complex (mean=989.711; SD=381.208) misspecifications were found to be considerably larger. RMSEA deviated from 0 with trivial SDs under M_W or M_{WB} conditions. For example, under M_W condition, means of RMSEA in simple and complex misspecifications were 0.056 and 0.055, respectively. CFI shrank from 1 with small SDs under M_W or M_{WB} conditions. For example, under M_W condition, means of CFI in simple and complex misspecifications were 0.967 and 0.968, respectively.

SRMR-W also inflated from 0 with small SDs under M_W or M_{WB} conditions. Note that, the means of SRMR-W were noticeably different between simple and complex misspecifications. For example, under M_W condition, means of SRMR-W in simple and complex misspecifications were 0.203 and 0.052, respectively. On the contrary, under M_B condition, only SRMR-B showed a relatively large value (or differed from 0) to indicate a potential model misspecification, especially under simple misspecification (0.216) where the covariance between the latent factors was omitted from the model.

Table 4.4 Descriptive Statistics of the Fit Indices of the Misspecified Models in Study 1

Misspeci-fication	Fit Index	M_W (*n=11,929*) Mean	SD	M_B (*n=12,000*) Mean	SD	M_{WB} (*n=12,000*) Mean	SD
Simple	Chi-square	1022.868	393.811	101.855	19.435	1065.716	399.491
	RMSEA	0.056	0.002	0.011	0.004	0.057	0.002
	CFI	0.967	0.003	0.999	0.001	0.965	0.002
	SRMR-W	0.203	0.007	0.007	0.002	0.203	0.006
	SRMR-B	0.047	0.025	0.216	0.054	0.215	0.055
		M_W (*n=11,996*) Mean	SD	M_B (*n=11,887*) Mean	SD	M_{WB} (*n=11,999*) Mean	SD
Complex	Chi-square	989.711	381.208	97.100	19.304	1038.621	388.056
	RMSEA	0.055	0.002	0.010	0.004	0.056	0.002
	CFI	0.968	0.003	0.999	0.001	0.966	0.003
	SRMR-W	0.052	0.002	0.007	0.002	0.052	0.002
	SRMR-B	0.034	0.015	0.056	0.014	0.057	0.016

Note. All the means and SDs were calculated after replications with convergence problems or improper solutions were excluded. RMSEA = root-mean-square error of approximation. CFI = comparative fit index. SRMR-W = standardized root mean square residual for within-model. SRMR-B = standardized root mean square residual for between-model.

Sensitivity of the Target Fit Indices for Various Types of Misspecification

In this section, whether the target fit indices are equally sensitive to various types of misspecification were evaluated by controlling for other design factors. The misspecification type (MT) contained two levels of misspecification: simple and complex misspecifications. Tables 4.5, 4.6, and 4.7 present the total SOS's for fit indices, and the η^2 of each factor and interaction term under M_W, M_B, and M_{WB} conditions, respectively. Note that the total SOS's of RMSEA and CFI were very small (ranging from 0.110 to 0.363) across M_W, M_B, and M_{WB} conditions. Hence the corresponding η^2s of the design factors were negligible.

On the other hand, the total SOS's for SRMR-W and SRMR-B were in general larger than CFI and RMSEA. SRMR-W had large total SOS's under M_W (135.236) and M_{WB} (134.983) conditions, but a relatively small total SOS under M_B (0.089) condition, while SRMR-B had large total SOS's under M_B (192.262) and M_{WB} (188.827) conditions, but a relatively small total SOS under M_W (11.291) condition.

Moreover, SRMR-W was found to be sensitive to the various types of misspecification (MT) under M_W (η^2 = 99.57%) and M_{WB} (η^2 = 99.56%) conditions. On the other hand, SRMR-B was found to be sensitive to the various types of misspecification under M_B (η^2 = 80.41%) and M_{WB} (η^2 = 79.01%) conditions. As Table 4.4 shows, both SRMR-W and SRMR-B were in general more sensitive to simple misspecification (with larger mean values) than complex misspecification.

Additionally, total SOS's of SRMR-B were accounted for by the interaction effect ICC*MT under M_B and M_{WB} conditions (η^2 = 8.42% and 9.38%, respectively). The results showed that the effects of MT on the performance of SRMR-B were substantially moderated by ICC.

Table 4.5 Eta-Squares (η^2) for the Fit Indices of the Model Misspecification in the Within-model Only (M_W) in Study 1

	Fit Index				
Sources	Chi-square	RMSEA	CFI	SRMR-W	SRMR-B
Total SOS	3604237807.173	0.110	0.172	135.236	11.291
Overall η^2	97.35%	9.93%	31.47%	99.57%	92.66%
NG	23.53%	0.23%	0.07%	0.00%	1.80%
GS	70.73%	4.71%	6.44%	0.00%	7.56%
ICC	0.00%	0.17%	19.23%	0.00%	63.50%
MT	0.16%	4.71%	4.10%	99.57%	8.98%
NG*GS	2.90%	0.01%	0.00%	0.00%	0.00%
NG*ICC	0.00%	0.04%	0.03%	0.00%	0.19%
NG*MT	0.01%	0.01%	0.01%	0.00%	0.06%
GS*ICC	0.00%	0.03%	1.58%	0.00%	3.17%
GS*MT	0.02%	0.00%	0.00%	0.00%	2.29%
ICC*MT	0.00%	0.01%	0.00%	0.00%	4.11%
NG*GS*ICC	0.00%	0.01%	0.01%	0.00%	0.00%
NG*GS*MT	0.00%	0.00%	0.00%	0.00%	0.00%
NG*ICC*MT	0.00%	0.00%	0.00%	0.00%	0.03%
GS*ICC*MT	0.00%	0.00%	0.00%	0.00%	0.96%
NG*GS*ICC*MT	0.00%	0.00%	0.00%	0.00%	0.01%

Note. SOS = sum of square. RMSEA = root-mean-square error of approximation. CFI = comparative fit index. SRMR-W = standardized root mean square residual for within-model. SRMR-B = standardized root mean square residual for between-model. SOS = sum of squares. NG = number of group. GS = group size. ICC = intra-class correlation. MT = misspecification type (simple or complex misspecifications).

Table 4.6 Eta-Squares (η^2) for the Fit Indices of the Model Misspecification in the Between-model Only (M_B) in Study 1

	Fit Index				
Sources	Chi-square	RMSEA	CFI	SRMR-W	SRMR-B
Total SOS	9136317.101	0.363	0.016	0.089	192.262
Overall η^2	28.04%	26.19%	27.52%	59.31%	92.21%
NG	9.21%	0.01%	0.21%	11.96%	0.03%
GS	2.21%	9.54%	14.50%	45.77%	0.02%
ICC	14.00%	13.32%	9.45%	0.43%	3.31%
MT	1.48%	1.40%	1.27%	0.25%	80.41%
NG*GS	0.09%	0.01%	0.03%	0.31%	0.00%
NG*ICC	0.54%	0.01%	0.01%	0.00%	0.01%
NG*MT	0.05%	0.00%	0.00%	0.00%	0.00%
GS*ICC	0.30%	1.73%	1.79%	0.18%	0.01%
GS*MT	0.00%	0.06%	0.14%	0.12%	0.00%
ICC*MT	0.01%	0.04%	0.00%	0.20%	8.42%
NG*GS*ICC	0.01%	0.01%	0.01%	0.01%	0.00%
NG*GS*MT	0.01%	0.01%	0.01%	0.00%	0.00%
NG*ICC*MT	0.00%	0.00%	0.01%	0.00%	0.00%
GS*ICC*MT	0.12%	0.05%	0.09%	0.07%	0.00%
NG*GS*ICC*MT	0.01%	0.00%	0.00%	0.01%	0.00%

Note. SOS = sum of square. RMSEA = root-mean-square error of approximation. CFI = comparative fit index. SRMR-W = standardized root mean square residual for within-model. SRMR-B = standardized root mean square residual for between-model. SOS = sum of squares. NG = number of group. GS = group size. ICC = intra-class correlation. MT = misspecification type (simple or complex misspecifications).

Table 4.7 Eta-Squares (η^2) for the Fit Indices of the Model Misspecification in Both Within- and Between-models (M_{WB}) in Study 1

	Fit Index				
Sources	Chi-square	RMSEA	CFI	SRMR-W	SRMR-B
Total SOS	3599634980.729	0.111	0.190	134.983	188.827
Overall η^2	97.05%	3.93%	26.17%	99.56%	92.05%
NG	25.58%	0.05%	0.01%	0.00%	0.03%
GS	68.49%	0.05%	0.08%	0.00%	0.03%
ICC	0.01%	0.38%	21.43%	0.00%	3.56%
MT	0.12%	3.18%	2.59%	99.56%	79.01%
NG*GS	2.82%	0.00%	0.00%	0.00%	0.00%
NG*ICC	0.00%	0.01%	0.01%	0.00%	0.00%
NG*MT	0.01%	0.02%	0.02%	0.00%	0.01%
GS*ICC	0.00%	0.05%	1.90%	0.00%	0.02%
GS*MT	0.02%	0.01%	0.02%	0.00%	0.01%
ICC*MT	0.00%	0.16%	0.10%	0.00%	9.38%
NG*GS*ICC	0.00%	0.01%	0.01%	0.00%	0.00%
NG*GS*MT	0.00%	0.00%	0.00%	0.00%	0.00%
NG*ICC*MT	0.00%	0.00%	0.00%	0.00%	0.00%
GS*ICC*MT	0.00%	0.00%	0.00%	0.00%	0.00%
NG*GS*ICC*MT	0.00%	0.01%	0.00%	0.00%	0.00%

Note. SOS = sum of square. RMSEA = root-mean-square error of approximation. CFI = comparative fit index. SRMR-W = standardized root mean square residual for within-model. SRMR-B = standardized root mean square residual for between-model. SOS = sum of squares. NG = number of group. GS = group size. ICC = intra-class correlation. MT = misspecification type (simple or complex misspecifications).

Empirical Type I Error Rate and Statistical Power

The Type I error (α) of the fit indices was evaluated based on the rejection rates obtained for the true models (Hu & Bentler, 1999). In this section, the empirical Type I error rate of RMSEA, CFI, SRMR-W and SRMR-B given the true model (i.e., M_T) was examined. Because SRMR-B was found to be sensitive to ICC under M_T condition, it was decided to evaluate SRMR-B by high ICC (0.40) and low ICC (0.18). On the other hand, the statistical power of the fit indices was evaluated based on the rejection rates obtained for the misspecified models (i.e., M_W, M_B, and M_{WB}) (Hu & Bentler, 1999).

Table 4.8 presents the magnitudes of RMSEA, CFI, SRMR-W and SRMR-B for three different Type I error rates: 10%, 5% and 1%. Note that the smaller the magnitudes of RMSEA, SRMR-W and SRMR-B are, the better the models fit the sample data. On the other hand, larger CFI indicated a better model fit. In order to obtain reasonable Type I error rates (i.e., no larger than 5%), RMSEA and SRMR-W were at least 0.010, while the upper bound of CFI was 0.999. SRMR-B was at least 0.030 and 0.061 for high ICC and low ICC models, respectively.

Table 4.9 presents the corresponding values of RMSEA, CFI, SRMR-W and SRMR-B in terms of different levels of statistical power. To obtain a statistical power equal or larger than 0.80, the upper bounds of RMSEA and SRMR-W were 0.054, and 0.052, while CFI needed to be at least 0.970. The upper bounds of SRMR-B were 0.044 and 0.060 for high ICC and low ICC models, respectively.

Table 4.8 Cutoff Values of RMSEA, CFI, SRMR-W and SRMR-B in Terms of Empirical Type I Error Rates in Study 1

	Type I Error (α) Rates		
Fit Index	10%	5%	1%
RMSEA	0.008	0.010	0.013
CFI	0.999	0.999	0.998
SRMR-W	0.009	0.010	0.012
SRMR-B with High ICC	0.028	0.030	0.036
SRMR-B with Low ICC	0.057	0.061	0.071

Note. RMSEA = root-mean-square error of approximation. CFI = comparative fit index. SRMR-W = standardized root mean square residual for within-model. SRMR-B = standardized root mean square residual for between-model. ICC = intra-class correlation.

Table 4.9 Cutoff Values of WRMR, RMSEA, CFI, SRMR-W and SRMR-B in Terms of Various Levels of Statistical Power for Rejecting Misspecifications in the Within-model (M_W) and Between-model (M_B) in Study 1

	Statistical Power							
	70%		75%		80%		85%	
Fit Index	M_W	M_B	M_W	M_B	M_W	M_B	M_W	M_B
RMSEA	0.055	-	0.054	-	0.054	-	0.053	-
CFI	0.969	-	0.969	-	0.970	-	0.970	-
SRMR-W	0.053	-	0.052	-	0.052	-	0.051	-
SRMR-B with High ICC	-	0.048	-	0.046	-	0.044	-	0.042
SRMR-B with Low ICC	-	0.067	-	0.064	-	0.060	-	0.058

Note. RMSEA = root-mean-square error of approximation. CFI = comparative fit index. SRMR-W = standardized root mean square residual for within-model. SRMR-B = standardized root mean square residual for between-model.

The performance of the traditionally recommended cutoff values for the four target fit indices were also compared with the simulated data and the results are presented in Table 4.10. The statistical power of these fit indices was generally low (ranging from 0.00% to 54.33%) when the traditional cutoff values of RMSEA (0.06), CFI (0.95), and SRMR (0.08) were applied to evaluate the misspecifications in MSEM.

Table 4.10 Statistical Power when Traditional Cutoff Values of RMSEA (0.06), CFI (0.95), and SRMR (0.08) are Used for Evaluating Misspecified Models in Study 1

	Statistical Power	
Fit Index	M_W	M_B
RMSEA	1.77%	-
CFI	0.00%	-
SRMR-W	49.85%	-
SRMR-B	-	53.22%

DISCUSSION

In the present study, the effectiveness of four commonly used fit indices (i.e., RMSEA, CFI, SRMR-W and SRMR-B) on detecting the misspecifications in multilevel SEMs were evaluated using Fan and Sivo's (2005, 2007) approach to control the severity of the model misspecification.

RMSEA and CFI

RMSEA and CFI can effectively detect the model misspecifications in a within-model without being confounded by other design factors (i.e., NG, GS, and ICC). However, these two fit indices are not sensitive to the misspecifications occurring in a between-model. A possible explanation of this finding is that both RMSEA and CFI are a function of the overall model chi-square test statistics ($\mathcal{X}^2$), which is asymptotically distributed as a noncentral $\mathcal{X}^2$ distribution with the noncentrality parameter equal to λ. Indeed, λ is a function of the sample size and the severity of the model misspecification (Saris & Satorra, 1993). In multilevel SEMs, the sample size of the within-model was the total sample size minus the number of groups (i.e., N-G), while the sample size of the between-model was the number of groups (i.e., G). Generally, N-G is a lot larger than G. Therefore, compared with the between-model (with a relatively smaller sample size), the within-model contributed/weighed in with much more information to the overall model $\mathcal{X}^2$ value (Hox, 2002). This is also reflected in the simulation Study 1n which was found substantially larger model $\mathcal{X}^2$ values in the misspecified within-models than the misspecified between-models as presented in Table 4.4.

In addition, the χ^2 values for the misspecified within-models were close to the χ^2 values for the misspecified both within- and between-models (see Table 4.4). This result provided another support concerning the fact that misspecifications in the between-model cannot be fully reflected in the overall model χ^2 given that the between-model misspecifications only contribute a relatively small amount of information to the overall model χ^2 value. As indicated previously, both RMSEA and CFI are a function of the overall model χ^2 value and the within-model contributes most of the information for the overall model χ^2 value. Thus, RMSEA and CFI are far more sensitive to the within-model misspecifications and these two fit indices can be effectively used for evaluating plausible within-model misspecifications but not between-model misspecifications. Hox (2002) also drew a similar conclusion where both RMSEA and CFI are not equally weighted by the within- and between-models and these two fit indices may be more sensitive to misspecification in the within-model than the between-model.

SRMR-W and SRMR-B

SRMR-W is a designated fit index for detecting within-model misspecifications and this simulation study has shown that SRMR-W can be used for identifying misspecifications in a within-model regardless of the impact of other design factors (i.e., NG, GS, and ICC). Nevertheless, SRMR-W is more sensitive to simple misspecification (i.e., misspecification in the factor covariances) than complex misspecification (i.e., misspecification in the pattern coefficients), which is consistent with previous studies (e.g., Fan & Sivo, 2005; Hu & Bentler, 1998; Hu & Bentler, 1999). An explanation for this finding is that SRMR-W reflects the average discrepancy between the observed within-covariance matrix and the model-implied within-covariance matrix, with misspecifications in the factor covariances resulting in more covariances being constrained to zero in the model-implied covariance matrix. This in turn increases the discrepancy between the observed and model-implied covariance matrices and results in a large SRMR-W (Fan & Sivo, 2005). Similar to SRMR-W, SRMR-B is a designated fit index for detecting misspecifications in a between-model. A similar pattern of results was also found for SRMR-B; that is, SRMR-B is more sensitive to simple misspecification than complex misspecification. However, SRMR-B is more sensitive to misspecification in high ICC models than low ICC models.

Statistical Power of the Traditionally Recommended Cutoff Values

We also examined the statistical power of the traditionally recommended cutoff values of the four target fit indices based on the simulated data. These cutoff values in general resulted in very low statistical powers (below 0.55) for rejecting misspecified models. To maintain a reasonably high statistical power (i.e., statistical power equal to 0.80 or higher), lower cutoff values for RMSEA (0.054), SRMR-W (0.052), and SRMR-B (0.044 for high ICC models and 0.060 for low ICC models), are required while a higher cutoff value for CFI (0.970) is needed.

Implications and Recommendations

In this study, the effectiveness of four commonly used fit indices (RMSEA, CFI, SRMR-W and SRMR-B) in detecting model misspecifications in multilevel SEMs was evaluated. The simulation results showed that the RMSEA, CFI, and SRMR-W are generally only sensitive to within-model misspecifications but not to between-model misspecifications. Thus, a model with low RMSEA and SRMR-W and high CFI does imply that a within-model fits the data adequately, but not necessarily implies that a between-model also fits the data well. Researchers should interpret the goodness-of-fit with caution because low RMSEA, SRMR-W and high CFI may simply be the result of the insensitivity of these fit indices to between-model misspecifications.

On the other hand, SRMR-B was the only commonly used fit index that was sensitive to misspecified between-models. Although SRMR-B can be used for detecting between-model misspecifications, the sensitivity of this fit index is a function of the misspecification type and the ICC of the model. That is, SRMR-B is more sensitive to simple misspecifications and a high ICC model. Further development on more sensitive fit indices, especially for between-model misspecifications, is needed.

In addition, both SRMR-W and SRMR-B are more sensitive to simple misspecification than complex misspecification. For example, SRMR-W and SRMR-B are more sensitive to misspecifications in covariances, and the high value of these indices can be viewed as a signal of possible misspecifications in the covariance part of a model. Following the general recommendation by Hu and Bentler (1999), researchers should use these fit indices in combination rather than just relying on a single fit index to evaluate a model.

Finally, the traditionally recommended cutoff values did not perform well in our simulation study (i.e., resulted in low statistical power for rejecting the misspecified models). Thus, these cutoff values should only be used with caution when evaluating multilevel SEMs. According to the simulation results, more strict cutoff values for these fit indices are required for maintaining a reasonable statistical power.

CHAPTER V

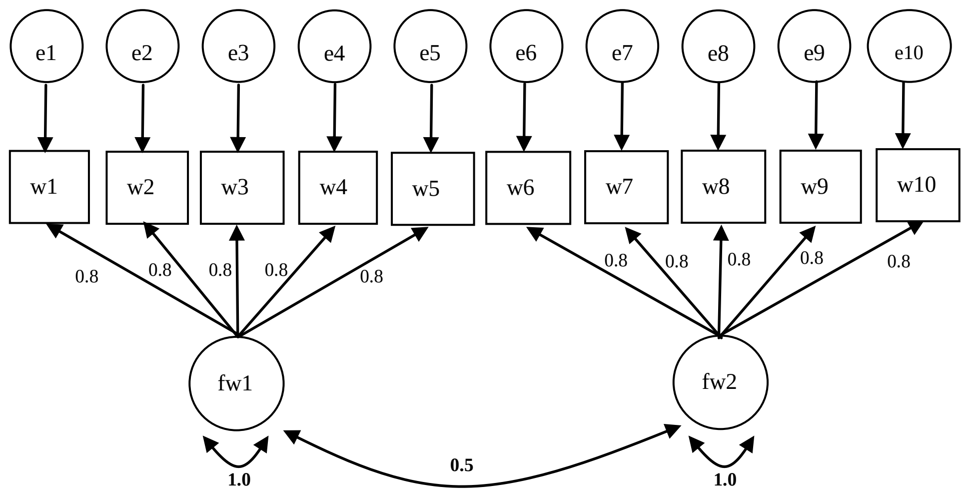
STUDY 2: THE SENSITIVITY OF FIT INDICES IN MSEM WITH DICHOTOMOUS OUTCOME VARIABLES

The goal of Study 2 was to examine the sensitivity of fit indices (i.e., WRMR, RMSEA, CFI, SRMR-W and SRMR-B) in multilevel confirmatory factor analysis models with dichotomous outcome variables (i.e., indicators). This chapter introduces the method employed in my second study. It includes analysis, results, and discussion.

METHOD

A Monte Carlo study was conducted using Mplus 5.2 (Muthén & Muthén, 1998-2007) to investigate the sensitivity of the five commonly used fit indices (i.e., WRMR, RMSEA, CFI, SRMR-W and SRMR-B) for detecting model misspecifications in two-level models under various conditions. A multilevel confirmatory factor analysis (MCFA) model was used for data generation. Figures 5.1 and 5.2 present both within- and between-models, which were specified as having the same factor structure with ten dichotomous indicators loaded on two latent factors. Five design factors were considered in the present study: number of groups (NG) at the between level, group size (GS), intra-class correlation (ICC), threshold (TH), and the model misspecification (MM). The descriptions for design factors are presented as follows.

Within-model



Between-model

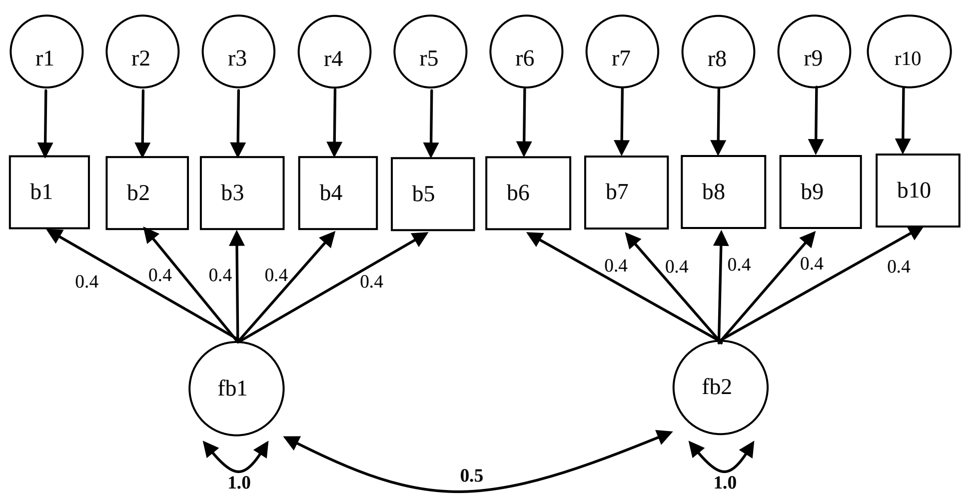
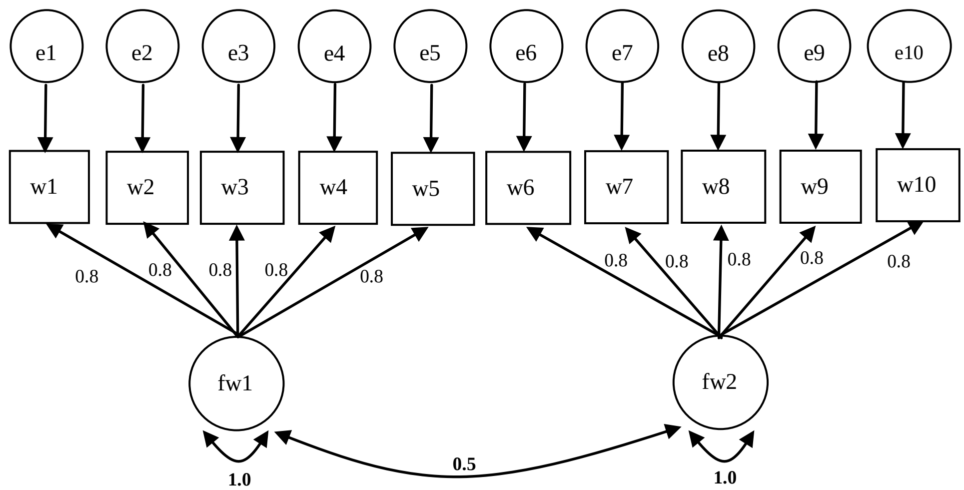


Figure 5.1 Multilevel CFA Model for Data Generation (Low ICC) in Study 2

Within-model



Between-model

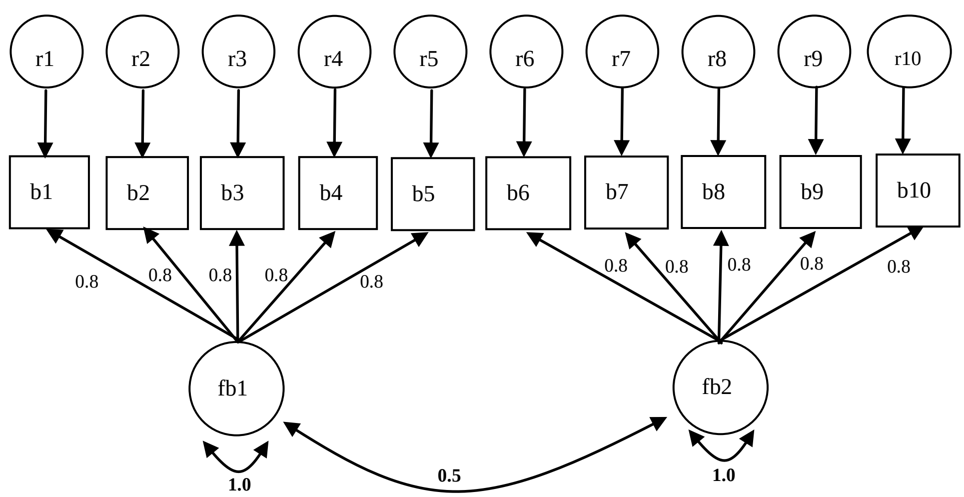


Figure 5.2 Multilevel CFA Model for Data Generation (High ICC) in Study 2

Number of Groups (NG). Number of groups relates to the estimation of the between-model. Hox and Maas (2001) concluded that large NGs (i.e., larger than 50 groups) were needed for acceptable estimates for a between-model with low ICC conditions. On the other hand, in SEM literature, two hundred observations seem to be the minimum required sample size for obtaining unbiased and consistent estimates when using the maximum likelihood estimation method (Boomsma, 1987; Loehlin, 2004). In the present study, three different NGs were adopted (i.e., 150, 200, and 250 groups) to evaluate whether NG level would affect fit indices performance in detecting model misspecification.

Group size (GS). Whereas size was problematic with NG, Hox and Maas (2001) concluded that small group size (GS) (i.e., 10 observations per group) could produce admissible within-model estimates . Thus, we used two levels of GS (i.e., 15 and 30 observations per group).

Intra-Class Correlation (ICC). The low ICC condition was created by giving lower values (0.4) for the pattern coefficients in the between-model (see Figure 5.1) while the high ICC condition was created by substituting higher values (0.8) for pattern coefficients in the between-model (see Figure 5.2). The pattern coefficients of the between-model considered in this study lead to two levels of ICC: .16 (low) and .29 (high).

Threshold (TH). Two levels of TH were adopted: 0 and 1. Threshold can determine the category of the measured dichotomous variable (Asparouhov & Muthén, 2007; Bollen, 2002). For example, falling short of the threshold, would result in a response of

"0". Conversely, passing this threshold, would result in a reponse of “1”. Thresholds of the dichotomous outcome variables were set to be 0 or 1 in the present study. When thresholds were equalto 0, the proportion of response “0” to “1” was 50% : 50% (balanced condition), whereas when thresholds equaled 1, the proportion of response “0” to “1” was 75% : 25% (skewed condition).

Model Misspecification (MM). For examining fit index sensitivity to model misspecification, the same model misspecification condition employed by Ryu and West (in press) was adopted. This model, misspecification in factor structure, consists of a two-factor model misspecified as a single-factor model. For example, a correctly specified (or true) model with low ICC is shown in Figure 5.1. Both within- and between-models were specified as having a two-factor structure with ten dichotomous indicators loaded on two latent factors. Within-model misspecification (M_W) indicated that the within-model was misspecified as a single-factor model, while the between-model was correctly specified as a two-factor model. Between-model misspecification (M_B) indicated that the between-model was misspecified as a single-factor model, while the within-model was correctly specified as a two-factor model. Conversely, within- and between-models misspecification (M_{WB}) indicated that the within- and between-models were misspecified as a single-factor model simultaneously.

In the current study, four models with different model misspecifications were specified to evaluate the sensitivity of WRMR, RMSEA, CFI, SRMR-W, and SRMR-B. These models were the following: (a) no model misspecification (true model, M_T; see

Figure 5.1 and 5.2), (b) model misspecification in the within-model only (M_W), (c) model misspecification in the between-model only (M_B), and (d) model misspecification in both within- and between-models (M_{WB}). The weighted least squares estimation method (WLSM) was used for all the analyses (Muthén & Muthén, 1998-2007).

ANALYSIS

Four design factors within each of the four models (M_T, M_W, M_B, and M_{WB}) were employed: (a) numbers of groups in between levels (NG; 150, 200 and 250), (b) group size (GS; 15 and 30), (c) intra-class correlation (ICC; low and high), and (d) threshold (TH; balanced and skewed). These factors were integrated into 96 conditions (4 models x 3 NG x 2 GS x 2 ICC x 2 TH). For each condition, 200 replications were generated and analyzed.

Convergence rates were first reported. Then, new replications were generated to replace replications with convergence problems or improper solutions (e.g., negative unique variances) in order to make 200 convergent solutions for each condition. For the M_T model replications, means and standard deviations (SDs) of each fit index were reported. A series of ANOVAs were conducted to determine the impact of the design factors on the effectiveness of the five targeted fit indices (i.e., WRMR, RMSEA, CFI, SRMR-W, and SRMR-B) in identifying the correctly specified models. Similarly, for the M_W, M_B, and M_{WB} model replications, means and SDs of each fit index were first reported for evaluating the sensitivity of the five fit indices for detecting model misspecifications. A series of ANOVAs were then conducted to examine whether the target fit indices were sensitive to model misspecification regardless of other design factors. Finally, the statistical powers based on the traditionally recommended cutoff values (i.e., RMSEA < 0.06, CFI > 0.95, SRMR < 0.08, WRMR < 0.90) were computed from the simulated data.

RESULTS

The purpose of this study was to evaluate the effectiveness of fit indices in detecting model misspecification in multilevel SEM with dichotomous outcome variables. First, the convergence failure and improper solution of our Monte Carlo study are reported. Also investigated are the performances of the five fit indices (i.e., WRMR, RMSEA, CFI, SRMR-W, and SRMR-B) in the correctly specified (or true) models (M_T). Theoretically, all the fit indices were expected to show that the specified models perfectly fitted the data regardless of the NG, GS, ICC and TH. The effectiveness of the fit indices in detecting the misspecified models (M_W, M_B, and M_{WB}) was examined to determine the impact of the design factors on the performance of fit indices. It was expected that all the fit indices should be sensitive to the model misspecifications regardless the design factors. In the end, the statistical power in detecting the misspecified models was examined using the traditionally recommended cutoff values of the five fit indices.

Convergence Failure and Improper Solutions

The percentage of replications with convergence problems or improper solutions (e.g., negative unique variances) across all combinations of sample size conditions ranged from 0.227% to 0.341%. Results show that the number of problematic replications was generally small (less than 1%) in the simulation studies. All the problematic replications were not included and replaced with new replications in the analyses presented below.

Performance of the Target Fit Indices on the Correctly Specified Models

If the specified model fits the data perfectly, RMSEA, SRMR-W and SRMR-B are equal to 0.00 while CFI is equal to 1.00. Conversely, the magnitude of WRMR varies from model to model. Table 5.1 presents the means and SDs of the chi-square values ($\mathcal{X}^2$), WRMR, RMSEA, CFI, SRMR-W and SRMR-B when fitting the correctly specified (or true) model (M_T). When the model was correctly specified, the mean of $\mathcal{X}^2$ would be close to the degrees of freedom of the model (*df* = 68). The mean (67.671) of $\mathcal{X}^2$ shown in Table 5.1 indicated that the simulation study was correctly conducted.

The mean of WRMR was 0.559 with SD equal to 0.070. Even though the SD of WRMR was relatively large compared with other fit indices, WRMR's mean was considerably smaller than the recommended cutoff value 0.90. The results showed WRMR accurately indicated the model was correctly specified. Similarly, the mean of RMSEA was close to 0 with trivial SD, which indicated that RMSEA performed well under M_T. Also, CFI (with mean and SD equal to 0.999 and 0.002, respectively) performed extremely well when models were correctly specified. Conversely, the performance of SRMR-W and SRMR-B on correctly specified models were not as good as RMSEA and CFI, which had relatively large means (SRMR-W = 0.020 and SRMR-B = 0.045) and SDs (SRMR-W = 0.005 and SRMR-B = 0.020).

Table 5.1 Descriptive Statistics of the Fit Indices of the True Models (M_T) in Study 2

Fit Index	Mean	SD
Chi-square	67.671	14.467
WRMR	0.559	0.070
RMSEA	0.003	0.004
CFI	0.999	0.002
SRMR-W	0.020	0.005
SRMR-B	0.045	0.020

Note. n=4,800. Degrees of freedom of the hypothesized model were 68. WRMR = weighted root mean square residual. RMSEA = root-mean-square error of approximation. CFI = comparative fit index. SRMR-W = standardized root mean square residual for within-model. SRMR-B = standardized root mean square residual for between-model.

The possible impact of the four design factors (i.e., NG, GS, ICC, and TH) on the performance of the five target fit indices separately on the correctly specified models using ANOVA was examined also. Table 5.2 presents the results. The total sum of squares (SOS) of each fit index showed the variability of the corresponding fit index across all true model replications, while eta-squared (η^2) demonstrated the proportion of the variance accounted for by a particular design factor or the interaction effect term. Note that η^2 was obtained by dividing the Type III sum of squares of a particular predictor or interaction effect by the corrected total sum of squares.

As shown in Table 5.2, all the design factors (e.g., NG, GS, ICC, and TH) only accounted for a trivial proportion of the total SOS of the model $\mathcal{X}^2$. Moreover, the total SOS's of RMSEA, CFI and SRMR-W were very small (ranged from 0.01 to 0.13). The corresponding η^2 of the design factors was negligible for all three fit indices. Compared with the other fit indices, the total SOS of WRMR was quite large (23.53) with the ICC accounting for 24.52% of the total SOS. In contrast, the total SOS of SRMR-B was relatively large (1.95) with the ICC accounting for 57.33% of the total SOS, followed by GS (12.78%), and NG (6.38%).

Table 5.2 Eta-Squares (η^2) for the Fit Indices of the True Models (M_T) in Study 2

Sources	Fit Index					
	Chi-square	WRMR	RMSEA	CFI	SRMR-W	SRMR-B
Total SOS	1004396.954	23.525	0.065	0.013	0.131	1.947
Overall η^2	0.77%	26.68%	3.29%	6.12%	75.03%	82.34%
NG	0.02%	0.02%	0.61%	1.09%	15.56%	6.38%
GS	0.03%	1.13%	2.07%	3.33%	46.37%	12.78%
ICC	0.09%	24.52%	0.00%	0.01%	1.44%	57.33%
TH	0.04%	0.57%	0.06%	1.02%	10.43%	1.60%
NG*GS	0.04%	0.03%	0.04%	0.10%	0.46%	0.27%
NG*ICC	0.09%	0.08%	0.04%	0.01%	0.01%	0.60%
NG*TH	0.13%	0.10%	0.12%	0.06%	0.08%	0.03%
GS*ICC	0.04%	0.00%	0.04%	0.00%	0.08%	2.55%
GS*TH	0.04%	0.05%	0.06%	0.28%	0.46%	0.19%
ICC*TH	0.01%	0.06%	0.00%	0.04%	0.10%	0.40%
NG*GS*ICC	0.03%	0.01%	0.05%	0.02%	0.00%	0.06%
NG*GS*TH	0.01%	0.00%	0.01%	0.02%	0.01%	0.00%
NG*ICC*TH	0.08%	0.02%	0.07%	0.08%	0.01%	0.03%
GS*ICC*TH	0.00%	0.00%	0.02%	0.03%	0.00%	0.11%
NG*GS*ICC*TH	0.12%	0.09%	0.10%	0.03%	0.02%	0.01%

Note. *n*=4,800. Degrees of freedom of the hypothesized model were 68. WRMR = weighted root mean square residual. RMSEA = root-mean-square error of approximation. CFI = comparative fit index. SRMR-W = standardized root mean square residual for within-model. SRMR-B = standardized root mean square residual for between-model. SOS = sum of squares. NG = number of group. GS = group size. ICC = intra-class correlation. TH= threshold.

Performance of the Target Fit Indices on the Misspecified Models

Table 5.3 presents the means and SDs of the $\mathcal{X}^2$ and the five target fit indices under models with different misspecification types across various simulation conditions. Generally speaking, four global model fit indices, namely $\mathcal{X}^2$, WRMR, RMSEA, and CFI, showed more sensitivity under M_W and M_{WB} conditions than under M_B condition. $\mathcal{X}^2$ had substantially large means and SDs in either M_W or M_{WB} conditions. In contrast with mean and SD of $\mathcal{X}^2$ under M_B (mean=321.452; SD=165.976), means and SDs under M_W (mean=1284.166; SD=587.723) and M_{WB} (mean=1515.964; SD=661.123) were found to be considerably larger. WRMR showed large deviation from 0.90 with relatively large SDs in M_W (mean=2.382; SD=0.525) and M_{WB} (mean=2.613; SD=0.532) conditions. In M_B condition, mean and SD of WRMR were 1.183 and 0.267, respectively. RMSEA deviated from 0 with trivial SDs in either the M_W or the M_{WB} conditions. Means of RMSEA under M_W and M_{WB} conditions were 0.061 and 0.067, respectively. CFI shrank from 1.0 with small SDs under M_W and M_{WB} conditions. Means of CFI in M_W and M_{WB} conditions were 0.838 and 0.804, respectively.

SRMR-W and SRMR-B were sensitive to within-models and between-models, respectively. SRMR-W inflated from 0 with small SDs under M_W (mean=0.090; SD=0.005) and M_{WB} (mean=0.090; SD=0.005) conditions and closed to 0 under M_B (mean=0.020; SD=0.005) condition. In contrast, SRMR-B showed relatively large values (or differed from zero) under M_B (mean=0.176; SD=0.035) and M_{WB} (mean=0.177; SD=0.035) conditions but also deviated from 0 under M_W (mean=0.046; SD=0.020) condition.

Table 5.3 Descriptive Statistics for the Fit Indices of the Misspecified Models in Study 2

	Misspecified Model					
	M_W		M_B		M_{WB}	
Fit Index	Mean	SD	Mean	SD	Mean	SD
Chi-square	1284.166	587.723	321.452	165.976	1515.964	661.123
WRMR	2.382	0.525	1.183	0.267	2.613	0.532
RMSEA	0.061	0.007	0.028	0.010	0.067	0.008
CFI	0.838	0.024	0.964	0.023	0.804	0.030
SRMR-W	0.090	0.005	0.020	0.005	0.090	0.005
SRMR-B	0.046	0.020	0.176	0.035	0.177	0.035

Note. n=4,800 for each misspecified model. WRMR = weighted root mean square residual. RMSEA = root-mean-square error of approximation. CFI = comparative fit index. SRMR-W = standardized root mean square residual for within-model. SRMR-B = standardized root mean square residual for between-model.

Sensitivity of the Target Fit Indices for Various Types of Misspecification

In this section, the target fit indices were evaluated to determine whether they are equally sensitive to model misspecifications by taking into account other design factors. Tables 5.4, 5.5, and 5.6 present the total SOS for fit indices, and the η^2 of each factor and interaction term under M_W, M_B, and M_{WB} conditions, respectively. Note that the total SOS's of RMSEA, CFI, and SRMR-W were small (ranging from 0.131 to 4.459) across M_W, M_B, and M_{WB} conditions. Hence the corresponding η^2s of the design factors were negligible. These results indicated the magnitudes of RMSEA, SRMR-W, and CFI were less sensitive to the design factors (NG, GS, ICC, and TH) included in our simulation.

The total SOS for WRMR was in general larger than the total SOS's for other fit indices. WRMR had large total SOS's under M_W condition (1323.965, see Table 5.4) and M_{WB} (1355.716, see Table 5.6), but a relatively small total SOS under M_B condition (341.674, see Table 5.5). For the M_W condition, the design factors GS, NG, TH, and ICC accounted for 58.59%, 19.82%, 11.06%, and 2.39% of total SOS, respectively, while the interaction effects only accounted for a total of 1.76% of total SOS. For the M_B condition, the design factors ICC, NG, and GS accounted for 52.54%, 12.26%, and 6.30% of total SOS, respectively. Also, the interaction effects only accounted for a total of 1.62% of total SOS. For the M_{WB} condition, the design factors GS, NG, and TH accounted for 56.70%, 23.71%, and 10.97% of total SOS, respectively. ICC and the interaction effects only accounted for a total of 1.76% of total SOS.

SRMR-B had a relatively large total SOS under M_B (5.835, see Table 5.5) and M_{WB} (5.898, see Table 5.6) conditions. For the M_B condition, the design factors ICC accounted for 49.24% of total SOS, while other design factors and the interaction effects only accounted for a total of 1.87% of total SOS. Similarly, for the M_{WB} condition, the design factors ICC dominated most of the total SOS (49.29%). The design factor ICC showed substantial influence on the performance of SRMR-B under M_B and M_{WB} conditions.

Table 5.4 Eta-Squares (η^2) for the Fit Indices of the Model Misspecification in the Within-model Only (M_W) in Study 2

	Fit Index					
Sources	Chi-square	WRMR	RMSEA	CFI	SRMR-W	SRMR-B
Total SOS	1657664919.686	1323.965	0.234	2.665	0.142	1.938
Overall η^2	93.53%	93.62%	65.57%	25.77%	3.77%	82.31%
NG	18.40%	19.82%	0.03%	0.05%	0.68%	6.34%
GS	57.49%	58.59%	3.01%	14.87%	1.96%	12.83%
ICC	0.86%	2.39%	4.62%	3.24%	0.13%	57.31%
TH	11.89%	11.06%	56.93%	5.97%	0.46%	1.64%
NG*GS	2.59%	0.82%	0.01%	0.08%	0.18%	0.31%
NG*ICC	0.06%	0.01%	0.02%	0.01%	0.02%	0.61%
NG*TH	0.48%	0.13%	0.00%	0.01%	0.00%	0.02%
GS*ICC	0.05%	0.08%	0.07%	1.37%	0.03%	2.47%
GS*TH	1.49%	0.38%	0.00%	0.00%	0.06%	0.19%
ICC*TH	0.11%	0.30%	0.79%	0.00%	0.05%	0.40%
NG*GS*ICC	0.01%	0.00%	0.01%	0.01%	0.00%	0.06%
NG*GS*TH	0.07%	0.01%	0.01%	0.03%	0.02%	0.00%
NG*ICC*TH	0.01%	0.01%	0.04%	0.08%	0.08%	0.01%
GS*ICC*TH	0.02%	0.01%	0.00%	0.00%	0.01%	0.11%
NG*GS*ICC*TH	0.00%	0.01%	0.03%	0.05%	0.09%	0.01%

Note. n=4,800. WRMR = weighted root mean square residual. RMSEA = root-mean-square error of approximation. CFI = comparative fit index. SRMR-W = standardized root mean square residual for within-model. SRMR-B = standardized root mean square residual for between-model. SOS = sum of squares. NG = number of group. GS = group size. ICC = intra-class correlation. TH= threshold.

Table 5.5 Eta-Squares (η^2) for the Fit Indices of the Model Misspecification in the Between-model Only (M_B) in Study 2

	Fit Index					
Sources	Chi-square	WRMR	RMSEA	CFI	SRMR-W	SRMR-B
Total SOS	132202782.677	341.674	0.437	2.499	0.131	5.835
Overall η^2	82.85%	73.92%	81.76%	75.53%	75.69%	51.11%
NG	10.24%	12.26%	0.02%	0.01%	15.72%	0.44%
GS	6.04%	6.30%	5.72%	5.97%	47.06%	0.73%
ICC	61.17%	52.54%	71.59%	65.81%	1.46%	49.24%
TH	1.44%	1.20%	1.56%	0.46%	10.19%	0.03%
NG*GS	0.34%	0.19%	0.00%	0.00%	0.47%	0.09%
NG*ICC	2.93%	1.29%	0.01%	0.00%	0.01%	0.13%
NG*TH	0.08%	0.04%	0.00%	0.00%	0.10%	0.02%
GS*ICC	0.29%	0.01%	2.74%	3.19%	0.11%	0.27%
GS*TH	0.01%	0.00%	0.06%	0.00%	0.46%	0.03%
ICC*TH	0.22%	0.05%	0.02%	0.06%	0.06%	0.03%
NG*GS*ICC	0.06%	0.02%	0.02%	0.01%	0.00%	0.04%
NG*GS*TH	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%
NG*ICC*TH	0.02%	0.01%	0.01%	0.01%	0.01%	0.02%
GS*ICC*TH	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%
NG*GS*ICC*TH	0.01%	0.01%	0.01%	0.01%	0.02%	0.03%

Note. n=4,800. WRMR = weighted root mean square residual. RMSEA = root-mean-square error of approximation. CFI = comparative fit index. SRMR-W = standardized root mean square residual for within-model. SRMR-B = standardized root mean square residual for between-model. SOS = sum of squares. NG = number of group. GS = group size. ICC = intra-class correlation. TH= threshold.

Table 5.6 Eta-Squares (η^2) for the Fit Indices of the Model Misspecification in Both Within- and Between-models (M_{WB}) in Study 2

	Fit Index					
Sources	Chi-square	WRMR	RMSEA	CFI	SRMR-W	SRMR-B
Total SOS	2097563316.083	1355.716	0.291	4.459	0.143	5.898
Overall η^2	95.04%	93.42%	77.59%	48.01%	3.68%	51.20%
NG	20.71%	23.71%	0.02%	0.05%	0.61%	0.51%
GS	52.45%	56.70%	0.00%	1.20%	1.87%	0.69%
ICC	6.25%	0.20%	29.98%	44.57%	0.14%	49.29%
TH	11.00%	10.97%	44.97%	1.80%	0.45%	0.04%
NG*GS	2.42%	0.81%	0.01%	0.04%	0.21%	0.07%
NG*ICC	0.33%	0.01%	0.01%	0.00%	0.03%	0.18%
NG*TH	0.46%	0.13%	0.00%	0.00%	0.00%	0.01%
GS*ICC	0.06%	0.24%	1.87%	0.22%	0.03%	0.26%
GS*TH	1.20%	0.32%	0.00%	0.00%	0.07%	0.03%
ICC*TH	0.05%	0.29%	0.68%	0.04%	0.05%	0.04%
NG*GS*ICC	0.02%	0.00%	0.01%	0.02%	0.03%	0.05%
NG*GS*TH	0.06%	0.01%	0.00%	0.01%	0.01%	0.00%
NG*ICC*TH	0.00%	0.01%	0.02%	0.02%	0.08%	0.01%
GS*ICC*TH	0.02%	0.01%	0.00%	0.01%	0.02%	0.01%
NG*GS*ICC*TH	0.01%	0.01%	0.02%	0.03%	0.08%	0.01%

Note. *n*=4,800. WRMR = weighted root mean square residual. RMSEA = root-mean-square error of approximation. CFI = comparative fit index. SRMR-W = standardized root mean square residual for within-model. SRMR-B = standardized root mean square residual for between-model. SOS = sum of squares. NG = number of group. GS = group size. ICC = intra-class correlation. TH= threshold.

Empirical Type I Error Rate and Statistical Power

The empirical Type I error (α) of the fit indices was evaluated based on the rejection rates obtained for the true models (Hu & Bentler, 1999). In this section, the empirical Type I error rate of WRMR, RMSEA, CFI, SRMR-W and SRMR-B given the true model (i.e., M_T) was examined. On the other hand, the statistical power of the fit indices was evaluated based on the rejection rates obtained for the misspecified models (i.e., M_W, M_B, and M_{WB}) (Hu & Bentler, 1999).

Table 5.7 presents the magnitudes of WRMR, RMSEA, CFI, SRMR-W and SRMR-B for three different Type I error rates: 10%, 5% and 1%. Note that the smaller the magnitudes of WRMR, RMSEA, SRMR-W and SRMR-B are, the better the models fit the sample data. On the other hand, larger CFI indicated a better model fit. In order to obtain reasonable Type I error rates (i.e., no larger than 5%), WRMR, RMSEA, SRMR-W and SRMR-B were at least 0.677, 0.010, 0.029, and 0.083, while the upper bound of CFI was 0.996.

Table 5.8 presents the corresponding values of WRMR, RMSEA, CFI, SRMR-W and SRMR-B in terms of different levels of statistical power. To obtain a statistical power equal or larger than 0.80, the upper bounds of WRMR, RMSEA, SRMR-W, and SRMR-B were1.898, 0.055, 0.086, and 0.114, while CFI needed to be at least 0.858.

Table 5.7 Cutoff Values of WRMR (0.90), RMSEA (0.06), CFI (0.95), SRMR (0.08) in Terms of Empirical Type I Error Rates in Study 2

	Type I Error (α) Rates		
Fit Index	1%	5%	10%
WRMR	0.725	0.677	0.653
RMSEA	0.013	0.010	0.008
CFI	0.993	0.996	0.997
SRMR-W	0.033	0.029	0.027
SRMR-B	0.101	0.083	0.075

Note. n=4,800 for each misspecified model. WRMR = weighted root mean square residual. RMSEA = root-mean-square error of approximation. CFI = comparative fit index. SRMR-W = standardized root mean square residual for within-model. SRMR-B = standardized root mean square residual for between-model.

Table 5.8 Cutoff Values of WRMR, RMSEA, CFI, SRMR-W and SRMR-B in Terms of Various Levels of Statistical Power for Rejecting Misspecifications in the Within-model (M_W) and Between-model (M_B) in Study 2

	Statistical Power							
	70%		75%		80%		85%	
Fit Index	M_W	M_B	M_W	M_B	M_W	M_B	M_W	M_B
WRMR	2.030	1.007	1.967	0.971	1.898	0.937	1.814	0.899
RMSEA	0.057	0.021	0.056	0.020	0.055	0.019	0.054	0.018
CFI	0.850	0.981	0.853	0.983	0.858	0.985	0.862	0.986
SRMR-W	0.088	-	0.087	-	0.086	-	0.085	-
SRMR-B	-	0.154	-	0.149	-	0.144	-	0.138

Note. n=4,800 for each misspecified model. WRMR = weighted root mean square residual. RMSEA = root-mean-square error of approximation. CFI = comparative fit index. SRMR-W = standardized root mean square residual for within-model. SRMR-B = standardized root mean square residual for between-model.

Most researchers have relied on traditional fit indices (e.g., RMSEA, CFI, SRMR, and WRMR), along with commonly used cutoff values proposed by Hu and Bentler (1999) or Yu (2002) as guidelines to justify the adequacy of hypothesized models. In this section, the performance of the traditionally recommended cutoff values for the five target fit indices were also compared with the simulated data and the results are presented in Table 5.9. WRMR showed reasonable statistical power (equal or larger than 80%) under M_W (100.00%) and M_B (84.83%) conditions. Moreover, RMSEA showed low statistical power under M_W (56.71%) condition and 0.00% statistical power under M_B condition given a cutoff value equal to 0.06. The statistical power of CFI was satisfied under M_W (100.00%) condition but not acceptable disappointed under M_B (28.52%) condition. Lastly, SRMR-W showed reasonable statistical power under M_W (97.23%) condition and SRMR-B perform well under M_B (100.00%) condition when the traditional cutoff values of SRMR (0.08) were applied.

Table 5.9 Statistical Power when Traditional Cutoff Values of WRMR (0.90), RMSEA (0.06), CFI (0.95), and SRMR (0.08) are Used for Evaluating Misspecified Models in Study 2

	Statistical Power	
Fit Index	M_W	M_B
WRMR	100.00%	84.83%
RMSEA	56.71%	0.00%
CFI	100.00%	28.52%
SRMR-W	97.23%	-
SRMR-B	-	100.00%

Note. n=4,800 for each misspecified model. WRMR = weighted root mean square residual. RMSEA = root-mean-square error of approximation. CFI = comparative fit index. SRMR-W = standardized root mean square residual for within-model. SRMR-B = standardized root mean square residual for between-model.

DISCUSSION

In this study evaluated the effectiveness of five commonly used fit indices (i.e., WRMR, RMSEA, CFI, SRMR-W and SRMR-B) to detect model misspecifications in MSEM with dichotomous outcome variables. The performance of WRMR under different model misspecification conditions was first presented. Following the section on WRMR, the other fit indices are discussed.

WRMR

In this section, I compared the differential performances of WRMR under three conditions: (a) M_T condition, (b) M_W and M_{WB} conditions and (c) M_B condition. First of all, WRMR was found to be sensitive to different degrees of ICC (η^2=24.52%) under M_T

condition, however, WRMR did not overreject the correctly specified models based on simulation results. Figure 5.3 shows the box plot for WRMR by ICC under M_T condition. The means of WRMR for high and low ICC were 0.525 and 0.594, respectively. In other words, WRMR tended to be higher (i.e., more likely to reject correctly specified models) when ICC increased. This finding was not a concern because the magnitudes of WRMR for all replications under M_T condition (*n*=4,800) did not exceed the recommended cutoff value of 0.90.

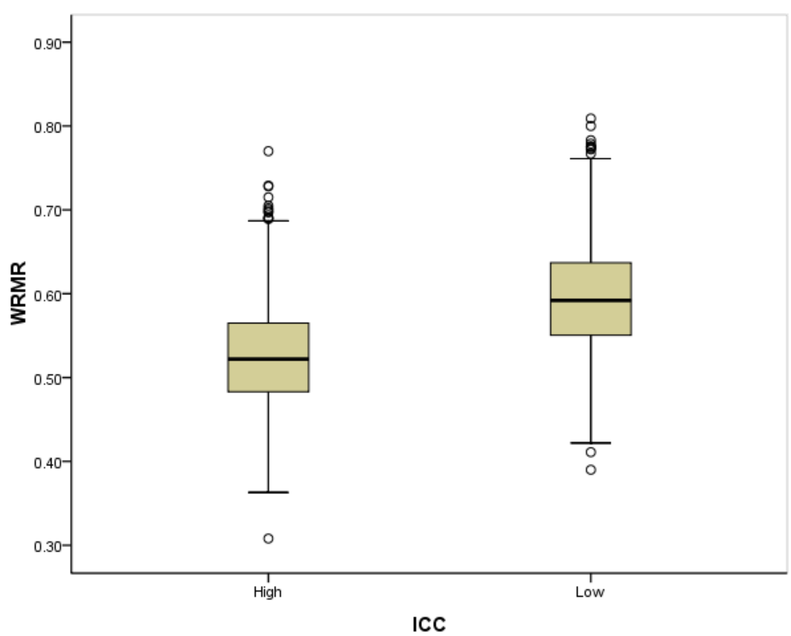
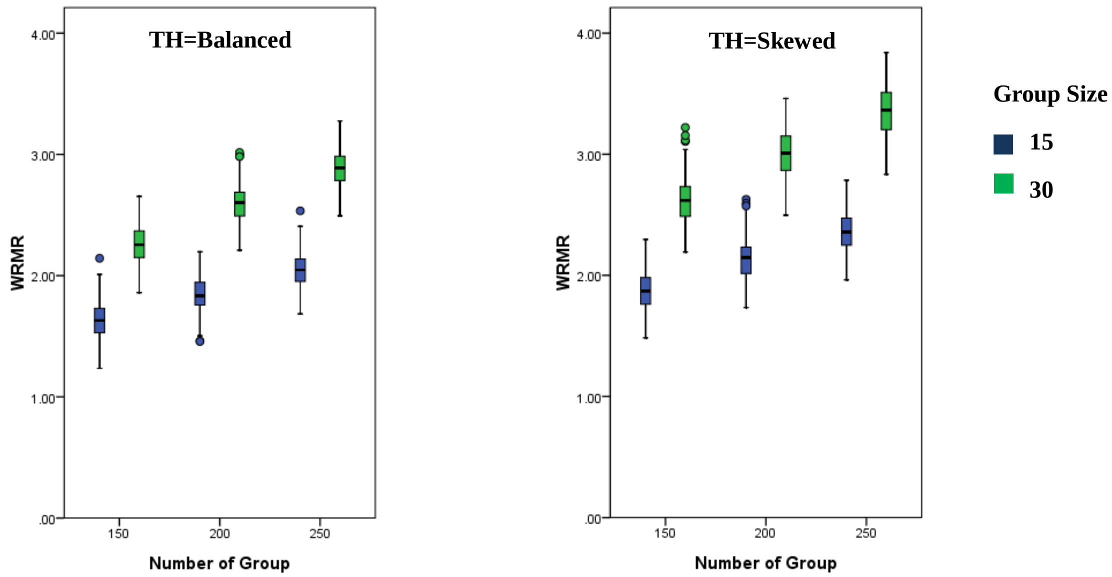


Figure 5.3 Box Plot for Weighted Root Mean Square Residual (WRMR) under M_T Condition

Second, WRMR performed similarly under M_W and M_{WB} conditions. WRMR reacted in the same pattern to misspecified within-models regardless of the accuracy of the corresponding between-models specified in the MSEM. For example, total SOS's were substantially large (ranged from 1323.96 to 1355.72) under M_W and M_{WB} conditions, compared to total SOS under M_B condition (341.67). Additionally, the design factor GS dominated more than 52% of total SOS, followed by NG (about 20%) and TH (about 10%). Thus, almost 80% of total SOS's were accounted for by GS and NG. As demonstrated by the boxplots in Figure 5.4, WRMR was more sensitive to within-model misspecifications when the total sample size (NG x GS) increased. This result was expected because the finding that larger sample size was associated with higher statistical power of fit index was also found in other related SEM simulation studies (e.g., Hancock & Freeman, 2001).

In addition, WRMR was sensitive to the design factor TH (i.e., misspecified within-models with either balanced or skewed outcome variables). Results showed that the magnitudes of WRMR tended to be a little lower (i.e., less likely to detect the model misspecifications) when outcome variables were balanced; however, generally the sensitivity of WRMR was acceptable under balanced and skewed outcome variables conditions.

M_W Condition



M_{WB} Condition

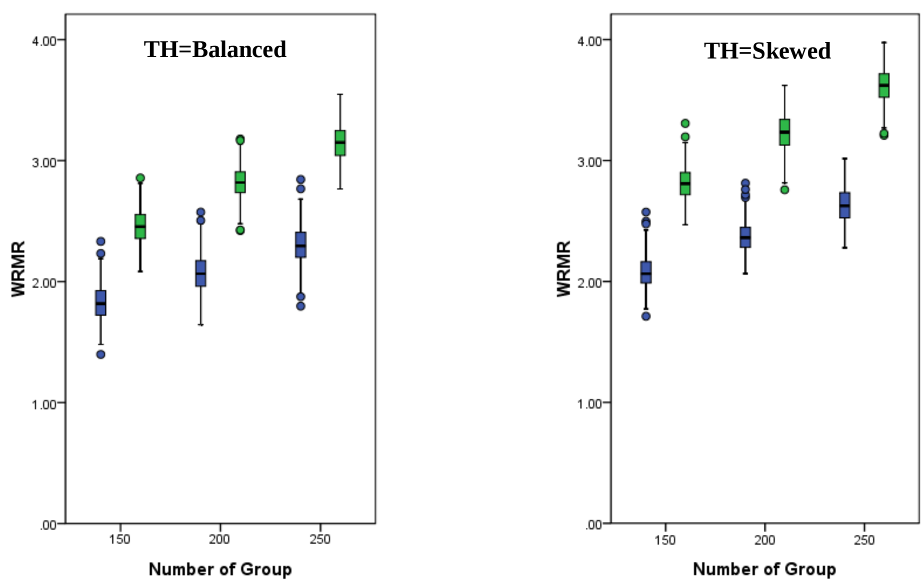


Figure 5.4 Box Plots for Weighted Root Mean Square Residual (WRMR) under M_W and M_{WB} Conditions

Lastly, under M_B condition, one concern was whether WRMR could detect model misspecifications in between-models. WRMR was found to be sensitive to different degrees of ICC (η^2 = 52.54%), followed by NG (12.26%) and GS (6.30%). Figure 5.5 showed box plots for WRMR under M_B condition by ICC, NG, and GS. Overall, the design factor ICC showed extreme influence on the performance of WRMR. Figure 5.5 indicated that WRMR was less likely to detect the between-model misspecifications (i.e., low statistical power) when ICC was relatively lower. Lower ICC implied that the between-model accounted for less proportion of the total variance (Hox, 2002). Thus, the less variance misspecified between-models carried, the fewer model misspecifications contributed to the WRMR. In other words, the information of model misspecification would be “weighted” by the proportion of total variance accounted for by the between-model.

In addition, WRMR was more sensitive to model misspecification when the total sample size increased. This finding was consistent with the findings under M_W and M_{WB} conditions. Note that such findings could be applied only when the within-model was correctly specified. Once the within-model was misspecified, the performance of WRMR would be dominated by the misfit of within-model (i.e., not sensitive to misspecified between-models anymore). Put simply, WRMR can be used to evaluate the model fit of between-models only when the within-models are correctly specified and the ICC is not too small.

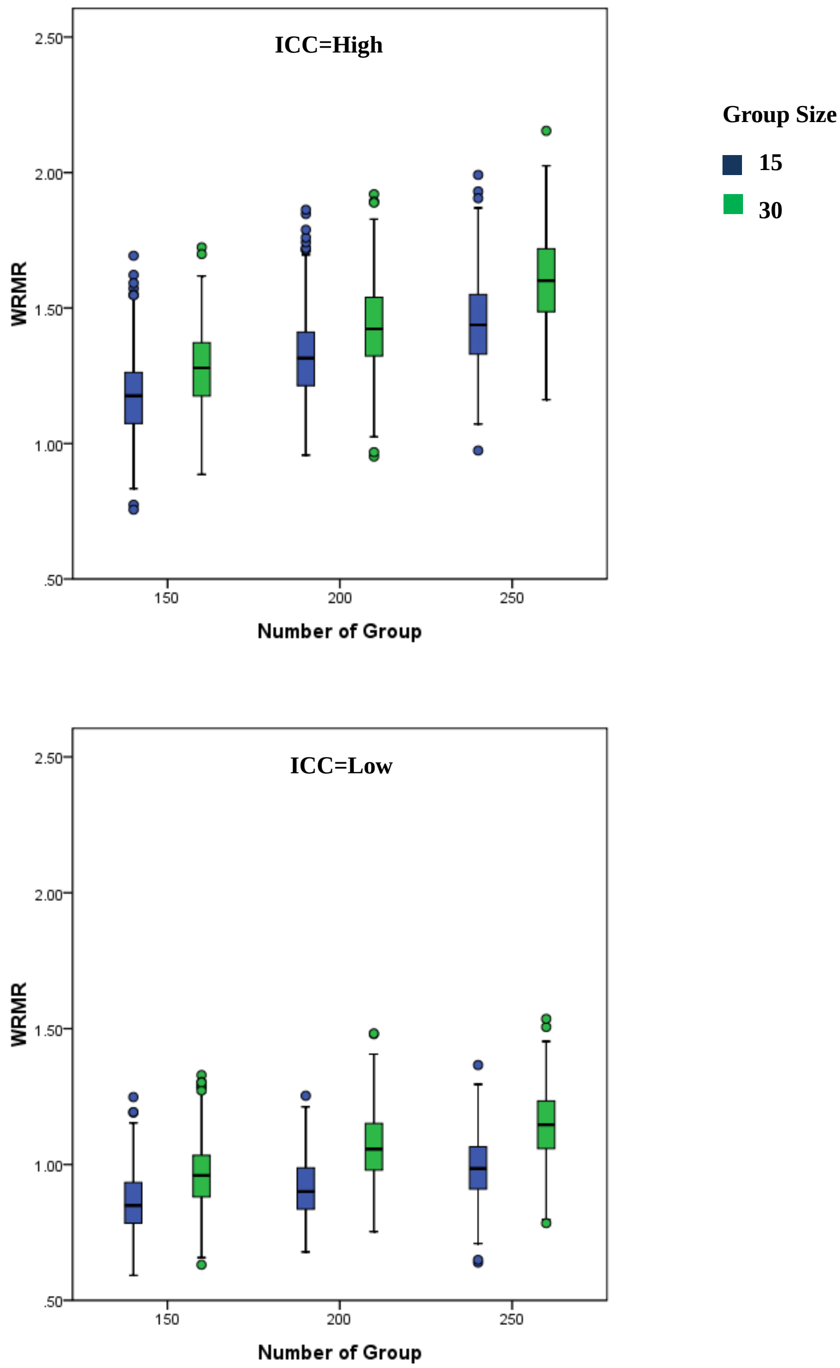


Figure 5.5 Box Plot for Weighted Root Mean Square Residual (WRMR) under M_B Condition

RMSEA and CFI

RMSEA and CFI can effectively detect the model misspecifications in a within-model without being confounded by other design factors (i.e., NG, GS, ICC, and TH). However, these two fit indices were not sensitive to the misspecifications occurring in a between-model. A possible explanation for this finding was that both RMSEA and CFI were a function of the overall model chi-square statistics ($\mathcal{X}^2$). Chi-square statistics was asymptotically distributed as an noncentral $\mathcal{X}^2$ distribution with the noncentrality parameter equal to λ. Indeed, λ was a function of the sample size and the severity of the model misspecification (Saris & Satorra, 1993). In multilevel SEMs, within-model sample size was the total sample size minus the number of groups (i.e., N-G), while between-model sample size was the number of groups (i.e., G). Generally, N-G was much larger than G. Therefore, compared with the between-model (with a relatively smaller sample size), the within-model contributed/weighed in with much more information to the overall model $\mathcal{X}^2$ value (Hox, 2002). This finding was also reflected in the simulation Study 1n which substantially larger model $\mathcal{X}^2$ values were found in the misspecified within-models than in the misspecified between-models as presented in Table 5.3. Hox (2002) also drew a similar conclusion where both RMSEA and CFI were not equally weighted by the within- and between-models. The author concluded that these two fit indices may be more sensitive to misspecification in the within-model than the between-model.

SRMR-W and SRMR-B

SRMR-W was a designated fit index for detecting within-model misspecifications. This simulation showed that SRMR-W can be used for identifying misspecifications in a within-model regardless of the impact of other design factors (i.e., NG, GS, ICC, and TH). Similar to SRMR-W, SRMR-B was a designated fit index for detecting misspecifications in a between-model. However, a different pattern of results was found for SRMR-B. Findings indicated that SRMR-B was more sensitive to model misspecification in high ICC models than low ICC models. Figure 5.6 showed a box plot for SRMR-B under M_B condition. SRMR-B tended to be less likely to detect the between-model misspecifications when ICC was relatively lower. This finding was consistent with WRMR performance findings.

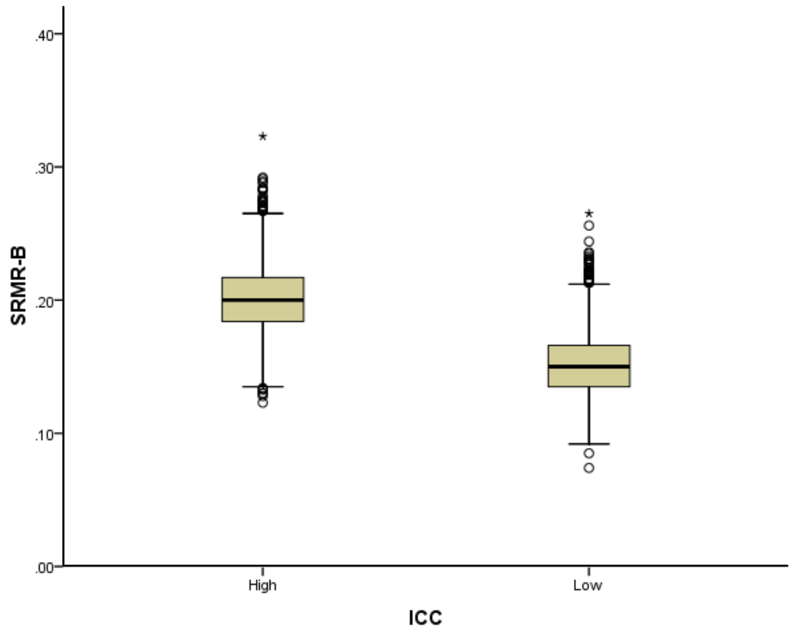


Figure 5.6 Box Plot for Standardized Root Mean Square Residual for Between-model (SRMR-B) under M_B Condition

Statistical Power of the Traditionally Recommended Cutoff Values

The statistical power of traditionally recommended cutoff values of the five target fit indices were also examined using simulated data. The WRMR cutoff worked well to obtain reasonably high statistical power (i.e., statistical power equal to 80% or higher). In contrast RMSEA was only sensitive under M_W condition, and the RMSEA cutoff value did result in reasonably high statistical power. Thus a lower cutoff value for RMSEA (0.055) is required. Like RMSEA, CFI was only sensitive under M_W condition. However, unlike RMSEA, the CFI cutoff value maintained reasonably high statistical power. Lastly, the cutoff values of SRMR-W and SRMR-B resulted in reasonably high statistical power.

Implications and Recommendations

The effectiveness of five commonly used fit indices (WRMR, RMSEA, CFI, SRMR-W and SRMR-B) in detecting MSEM model misspecifications with dichotomous outcome variables was evaluated in this study. First, simulation results showed that the RMSEA, CFI, and SRMR-W were generally only sensitive to within-model misspecifications, but they were not sensitive to between-model misspecifications. Thus, model fit for a model with low RMSEA and SRMR-W and high CFI implies adequate within-model fit but does not necessarily imply adequate between-model fit. Researchers should interpret the goodness-of-fit with caution because low RMSEA, SRMR-W and high CFI may simply be the result of the insensitivity of these fit indices to between-model misspecifications.

Second, SRMR-B, a commonly used fit index, was sensitive to misspecified between-models. Although SRMR-B can be used for detecting between-model misspecifications, the sensitivity of this fit index is a function of model ICC. Thus, SRMR-B is more sensitive to high ICC models. Even though SRMR-B performed well with both high and low ICC models in my study, results for SRMR-B showed a pattern of statistical power loss in detecting model misspecification when ICC decreased.

Third, WRMR was the fit index which was sensitive to misspecified within- and between-models. Researchers should first use RMSEA, CFI, and SRMR-W in combination to evaluate the within-model. Given a correctly specified within-model, researchers can apply WRMR to assess the accuracy of a between-model. However, researchers must take into account the impact of statistical power loss in detecting model misspecifications with decreasing ICC.

Finally, traditionally recommended cutoff values performed well in our simulation study (i.e., resulted in reasonably statistical power for rejecting the misspecified models) except for RMSEA. Thus, these cutoff values could be used as guidelines when evaluating models in MSEM.

CHAPTER VI

CONCLUSIONS

MSEM is a multivariate analytic approach for analyzing hierarchically structured data by specifying a within-model and a between-model separately and simultaneously. However, practices associated with model evaluation are problematic due to the lack of empirical research. Two Monte Carlo studies were conducted to investigate the sensitivity of fit indices in detecting model misspecification under different conditions. In Study 1, fit indices (i.e., RMSEA, CFI, SRMR-W and SRMR-B) were examined employing multilevel confirmatory factor analysis (MCFA) models with normally distributed outcome variables (i.e., indicators). In Study 2, WRMR as well as RMSEA, CFI, SRMR-W and SRMR-B were examined employing MCFA models with dichotomous outcome variables (a type of non-normal outcome variables).

Simulation results showed that the two global fit indices, RMSEA and CFI, only reflected within-model fit in MSEM with either normally distributed or dichotomous outcome variables. As shown in Table 4.4, RMSEA and CFI deviated from 0 and 1 (i.e., sensitive to model misspecifications), respectively under M_W condition but were very close to 0 and 1 (i.e., not sensitive to model misspecifications), respectively under M_B condition when the outcome variables were normally distributed. Similarly, RMSEA and CFI performed in a consistent manner when the outcome variables were dichotomous as shown in Table 5.3. Hence, researchers *cannot* treat RMSEA and CFI as global fit indices and believe that these fit indices can indicate the degree of entire model fit in

MSEM. A model with satisfactory RMSEA and CFI does imply that the within-model fits the data adequately but does not necessarily imply that the between-model fits the data well.

SRMR-W is another commonly reported fit index that is sensitive to misspecified within-models rather than misspecified between-models in MSEM with either normally distributed or dichotomous outcome variables. For this reason, researchers should use SRMR-W in combination with RMSEA and CFI to evaluate the within-model. As reflected in Table 4.4, SRMR-W deviated from 0 (i.e., sensitive to model misspecifications) under M_W condition but was very close to 0 (i.e., not sensitive to model misspecifications) under M_B condition when the outcome variables were normally distributed. Additionally, SRMR-W performed in the same pattern when the outcome variables were dichotomous as reflected in Table 5.3. Furthermore, simulation results also showed that SRMR-W was sensitive to different types of model misspecification (MT). As showed in Table 4.5, SRMR-W was differentially sensitive to simple and complex misspecifications in the within-model (η^2=99.57%) under M_W condition. Therefore, researchers could use SRMR-W, RMSEA, and CFI in combination to evaluate the within-models in MSEM. By using these fit indices, researchers should have no difficulty in determining the model fit of within-models.

On the other hand, simulation results showed that SRMR-B can be used *conditionally* to detect model misspecifications in the between-models in MSEM with either normally distributed or dichotomous outcome variables. Overall, the design factor ICC showed influence on the performance of SRMR-B. As shown in Table 4.6 and 5.5, the design factor ICC accounted for 3.31% and 49.24% of SRMR-B total SOS's under M_B conditions when the outcome variables were normally distributed and dichotomous, respectively. SRMR-B can show reasonable statistical power (i.e., statistical power equal to 80% or higher) in high ICC models. In other words, SRMR-B is less likely to detect between-model misspecifications when ICC decreases. Future studies should investigate the impact of statistical power loss in detecting between-model misspecifications with decreasing ICC.

The performance of WRMR was examined in MSEM with dichotomous outcome variables. The findings of WRMR were more complicated. First, once the within-model was misspecified, the performance of WRMR was dominated by the misfit of within-model (i.e., not sensitive to misspecified between-models anymore). WRMR was expected to have a significantly larger mean under M_{WB} condition than M_W condition. However, as shown in Table 5.3, WRMR had similar means (and SDs) across M_W (mean=2.382, SD=0.525) and M_{WB} (mean=2.613, SD=0.532) conditions. Tables 5.4 and 5.6 also showed that the effects of design factors on the performance of WRMR were quite close. Hence, WRMR failed to reflect misfit of the between-model if the within-model was misspecified.

Secondly, WRMR can be used to evaluate the model fit of between-models when the within-models are correctly specified and the ICC is not too small. As shown in Figure 5.5, WRMR performed well when ICC was relatively high (upper box plot). However, WRMR was less likely to detect the misspecified between-models when ICC was relative low (lower box plot).

Based on the findings of WRMR, the procedure for applying WRMR in MSEM with dichotomous outcome variables was presented as follows: (a) researchers should use WRMR, RMSEA, CFI, and SRMR-W in combination to evaluate a within-model; and (b) researchers should apply WRMR to evaluate the accuracy of a between-model only after achieving a correctly specified within-model. Note that, researchers must take into account the impact of ICC because WRMR might lose statistical power to detect between-model misspecifications when ICC decreases.

Only a limited number of design factors and parameter values have been considered in this simulation study. As Marsh et al. (2004) pointed out, researchers should not overly generalize from the results of Monte Carlo studies and should apply their results with caution. For a more complete picture of the effectiveness of commonly used fit indices in detecting model misspecifications in multilevel SEMs, further studies covering a broader range of conditions and parameter values are needed. My study only focused on misspecification in the covariance structure of a model. Future studies could examine the effectiveness of these fit indices in detecting misspecification in the mean structure of a model.

REFERENCES

Asparouhov, T., & Muthén, B. O. (2007). *Computationally efficient estimation of multilevel high-dimensional latent variable models.* Paper presented at the Joint Statistical Meeting, Salt Lake City, UT.

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

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Bollen, K. A. (2002). Latent variables in psychology and the social science. *Annual Review of Psychology, 53*, 605-634.

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

Boomsma, A. (1987). The robustness of maximum likelihood estimation in structural equation models. In P. Cuttance & R. Ecob (Eds.), *Structural modeling by example* (pp. 160-188). New York, NY: University of Cambridge.

Branum-Martin, L., Mehta, P. D., Fletcher, J. M., Carlson, C. D., Ortiz, A., Carlo, M., et al. (2006). Bilingual phonological awareness: Multilevel construct validation among Spanish-speaking kindergarteners in transitional bilingual education classrooms. *Journal of Educational Psychology, 98*(1), 170-181.

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

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research, 36*(4), 462-494.

Cheung, M. W.-L., & Au, K. (2005). Applications of multilevel structural equation modeling to cross-cultural research. *Structural Equation Modeling, 12*(4), 598–619.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*

Series B, 39, 1-38.

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

Duncan, T. E., Alpert, A., & Duncan, S. C. (1998). Multilevel covariance strcture analysis of sibling antisocial behavior. *Structural Equation Modeling, 5*(3), 211-228.

Dyer, N. G., Hanges, P. J., & Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly, 16*, 149-167.

Enders, C. K. (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement, 61*(5), 713.

Everson, H. T., & Millsap, R. E. (2004). Beyond individual differences: Exploring school effects on SAT scores. *Educational Psychologist, 39*(3), 157-172.

Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling, 12*(3), 343-367.

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, *42*(3), 509–529.

Fan, X., Thompson, B., & Wang, L. (1999). The effects of sample size, estimation methods, and model specification on Structural Equation Modeling fit indices. *Structural Equation Modeling, 6*(1), 56-83.

Gottfredson, N. C., Panter, A. T., Daye, C. E., Allen, W. F., & Wightman, L. F. (2009). The effects of educational diversity in a national sample of law students: Fitting multilevel latent variable models in data with categorical indicators. *Multivariate Behavioral Research, 44*(3), 305-331.

Hancock, G. R., & Freeman, M. J. (2001). Power and Sample Size for the Root Mean Square Error of Approximation Test of Not Close Fit in Structural Equation Modeling. *Educational and Psychological Measurement, 61*(5), 741-758.

Heck, R. H. (2001). New developments and techniques in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 89-127). Mahwah, NJ: Lawrence Erlbaum Assoc Inc.

Heck, R. H., & Thomas, S. L. (2008). *An introduction to multilevel modeling techniques* (2nd ed.). New York, NY: Routledge.

Hox, J. J. (2002). *Multilevel analysis techniques and applications*. Mahwah, NJ: Lawrence Erlabaum Associates.

Hox, J. J., & Maas, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling, 8*(2), 157-174.

Hox, J. J., & Maas, C. J. M. (2004). Multilevel structural equation models: The limited information approach and the multivariate multilevel approach. In K. v. Montfort, J. Oud & A. Satorra (Eds.), *Recent developments on structural equation models: Theory and applications*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424-453.

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

Jackson, D. L., Gillaspy, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods, 14*(1), 6-23.

Jöreskog, K. G., & Sörbom, D. (1996). *Lisrel 8: User's reference guild* (2nd ed.). Chicago: Scientific Software International Inc.

Kaplan, D. W. (2008). *Structural equation modeling: Foundations and extensions* (2nd ed.). Thousand Oaks, Calif: Sage Publications.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: The Guilford Press.

Loehlin, J. C. (2004). *Latent variable models: An introduction to factor, path, and structural equation analysis* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald. multivariate applications book series* (pp. 275-340). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320-341.

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*(1), 64-82.

Mueller, R. O. (1996). *Basic principles of structural equation modeling: An introduction to LISREL and EQS*. New York, NY: Springer.

Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*(1), 115-132.

Muthén, B. O. (1993). Goodness of fit with categorical and other nonnormal variables. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 205-234). Newbury Park, CA: SAGE Publications.

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

Muthén, B. O. (1998-2004). *Mplus technical appendices*. Los Angeles, CA: Muthén & Muthén.

Muthén, B. O., & Asparouhov, T. (2009). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis*. New York, NY: Routledge.

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

Muthén, L. K., & Muthén, B. O. (1998-2007). *Mplus user's guild* (5th ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*(4), 599-620.

Rabe-Hesketh, S., Skrondal, A., & Zheng, X. (2007). Multilevel structural equation modeling. In S.-Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 209-227). Amsterdam, The Netherlands: North-Holland publications.

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

Roberts, J. K. (2007). *Group dependency in the presence of small intraclass correlation coefficients: An argument in favor of not interpreting the ICC*. Paper presented at the American Educational Research Association.

Rowe, K. J. (2003). Estimating interdependent effects among multilevel composite variables in psychosocial research: An example of the application of multilevel structural equation modeling. In S. P. Reise & N. Duan (Eds.), *Multilevel modeling methodological advances, issues, and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ryu, E., & West, S. G. (in press). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling*.

Saris, W. E., & Satorra, A. (1993). Power evaluations in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 181-204). Newbury Park, CA: Sage.

Sivo, S. A., Fan, X., Witta, E. L., & Willse, J. T. (2006). The search for "optimal" cutoff properties: Fit index criteria in structural equation modeling. *Journal of Experimental Education*, *74*(3), 267-288.

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

Steiger, J. H., & Lind, J. C. (1980, May). *Statistically based tests for the number of factors.* Paper presented at the annual spring meeting of the Psychometric Society, Iowa City, IA.

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

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA: Sage Publications.

Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (Unpublished doctoral dissertation).

University of California, Los Angeles, CA.

Yuan, K.-H., & Bentler, P. M. (2002). On normal theory based inference for multilevel models with distributional violations. *Psychometrika, 67*(4), 539-562.

VITA

Hsien-Yuan Hsu
Email: hsuhy0914@gmail.com
Phone: (979) 661-6498
Address:
Department of Educational Psychology
Texas A&M University
704 Harrington Office Building.
College Station, TX 77843-4222

EDUCATION

2009	Ph.D. Educational Psychology (Research, Measurement, and Statistics) Texas A&M University, College Station, Texas
2004	M.E. Technological and Vocational Education (Educational Technology) National Taiwan University of Science and Technology Taipei, Taiwan
2001	B.B.A. Information Management Nan-Hua University Chiayi, Taiwan

RESEARCH INTERESTS

Structural Equation Modeling
Hierarchical Linear Modeling
Longitudinal Data Analysis
(Multilevel) Confirmatory / Exploratory Factor Analysis
Measurement Invariance
Test Theory
Bootstrapping
Questionnaire Design