ABSTRACT

Title of Document:	SENSITIVITY ANALYSIS OF STRUCTURAL PARAMETERS TO MEASUREMENT NONINVARIANCE: A BAYESIAN APPROACH
	Yoon Jeong Kang, Doctor of Philosophy, 2015
Directed By:	Professor Gregory R. Hancock Department of Human Development and Quantitative Methodology

Most previous studies have argued that the validity of group comparisons of structural parameters is dependent on the extent to which measurement invariance is met. Although some researchers have supported the concept of partial invariance, there is still no clear-cut partial invariance level which is needed to make valid group comparisons. In addition, relatively little attention has been paid to the implications of failing measurement invariance (e.g., partial measurement invariance) on group comparison on the underlying latent constructs in the multiple-group confirmatory factor analysis (MGCFA) framework. Given this, the purpose of the current study was to examine the extent to which measurement noninvariance affects structural parameter comparisons across populations in the MGCFA framework. Particularly, this study takes a Bayesian approach to investigate the sensitivity of the posterior distribution of structural parameter difference to varying types and magnitudes of noninvariance across two populations. A Monte Carlo simulation was performed to empirically investigate the sensitivity of structural parameters to varying types and magnitudes of noninvariant measurement models across two populations from a Bayesian approach. In order to assess the sensitivity of noninvariance conditions, three outcome variables were evaluated: (1) accuracy of statistical conclusion on structural parameter difference, (2) precision of the estimated structural parameter difference, and (3) bias in the posterior mean of structural parameter difference. Inconsistent with findings of previous studies, the results of this study showed that the three outcome variables were not sensitive to varying types and magnitudes of noninvariance across all conditions. Instead, the three outcome variables were sensitive to sample size, factor loading size, and prior distribution. These results indicate that even under a large magnitude of measurement noninvariance, accurate conclusions and inferences on structural parameter differences across populations could be obtained in the MGCFA framework. Implications for practice are discussed for applied researchers who wish to conduct group comparisons of structural parameters across populations under measurement noninvariance.

SENSITIVITY ANALYSIS OF STRUCTURAL PARAMETERS TO MEASUREMENT NONINVARIANCE: A BAYESIAN APPROACH

By

Yoon Jeong Kang

Doctoral Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Gregory R. Hancock, Chair
Professor Jeffrey R. Harring
Professor Laura M. Stapleton
Professor Ji Seung Yang
Professor Robert G. Croninger

Dedication

This dissertation is dedicated to

my parents Hakwon Kang and Younghee Cho.

Acknowledgements

I would like to express my deepest appreciation to my academic chair, Dr. Gregory R. Hancock, for his endless guidance and support throughout the dissertation process and my entire doctoral study. He inspires me in methodological research in Structural Equation Modeling and motivates me to participate in a variety of research activities. I really enjoy collaborating with him in preparing conference presentations and journal publications. His mentorship and encouragement were paramount in helping me obtain research experiences. I also would like to give thanks to my dissertation committee members, Dr. Jeffrey R. Harring, Dr. Laura M. Stapleton, Dr. Ji Seung Yang, and Dr. Robert G. Croninger for their extensive and insightful feedback on my dissertation.

Last and importantly, completing my dissertation would not have been possible without the love, patience, and support of my family. Special thanks go to my parents, for always being so supportive to me in countless ways; to my husband, Sangmin, for his support, encouragement, and love; to my sweet daughters, Michelle and Anna, for being healthy, being happy, and understanding for all the missed playtime while I studied.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

In the social and behavioral sciences, many processes are regarded as structural processes that conceptualize unobserved attributes (e.g., self-efficacy, quality of life). Often group comparisons on such unobserved attributes are at the heart of the research questions addressed by researchers. To illustrate, a researcher may investigate whether the self-efficacy of male students differs from that of female students. Given the fact that the attributes are not directly observed, they must be inferred from the observed variables using factor analytic models such as confirmatory factor analysis (CFA) where the unobserved attributes are often referred to as *latent constructs* or *factors*.

The validity of group comparisons on latent constructs has been a critical issue in social and behavioral studies. The validation of latent constructs for group comparison can be performed in the framework of construct validity, particularly related to the concept of construct *equivalence* or *comparability* in CFA (Little, 1997; Vandenberg & Lance, 2000; Wu, Li, & Zumbo, 2007). Although the concepts of construct equivalence and comparability slightly differ, they share the same basic idea of the conceptual equivalence of the latent constructs across groups. That is, it concerns whether the latent constructs inferred by a set of items (i.e., measurement instrument) have the same meaning in different populations. If different latent constructs are captured by a measurement instrument in different populations, group comparisons involving the latent constructs would be meaningless and invalid. As Vandenberg and Lance stated, if a set of items does not mean the same thing to different groups, group comparison on the latent constructs “may be tantamount to comparing apples and spark plugs” (Vandenberg & Lance, 2000, p. 9).

Construct equivalence is a conceptual notion and is related to theoretical validity (van de Vijver, 1998; van de Vijver & Tanzer, 2004). Thus, construct equivalence cannot be statistically tested and often rests on substantial theories or strong beliefs by researchers. Despite this, one statistical procedure, measurement invariance testing, has been used to collect evidence of construct equivalence. From a statistical standpoint, the measurement invariance test involves assessing equality of psychometric properties of a measurement instrument as well as equality of theoretical structures of latent constructs across populations. In the multi-group confirmatory factor analysis (MGCFA) framework, the theoretical structures of latent constructs are considered to be equal when items or tests load on the same latent construct across populations; this is often referred to as *pattern invariance* or *configural invariance* in measurement invariance literature (Horn & McArdle, 1992). The psychometric properties of a measurement instrument can be defined by three measurement parameters – factor loadings, intercepts, and error variances – in the MGCFA framework. Equality of psychometric properties of a measurement instrument can be achieved when measurement parameter estimates are considered to be equal across populations. In general, previous studies described three types of invariance for the measurement model in the MGCFA framework: weak (i.e., factor loading invariance), strong (i.e., factor loading and intercept invariance), and strict measurement invariance (i.e., factor loading, intercept, and error variance invariance) (Meredith, 1993; Meredith & Teresi, 2006).

There is agreement in the measurement invariance literature that configural invariance is a necessary condition for ensuring construct equivalence (e.g., Millsap & Kwok, 2004; Wu et al., 2007). It has been agreed that the latent construct cannot be assumed to be equivalent unless groups have the same latent construct structure. However, the same agreement has not been achieved regarding the necessity of the three types of measurement parameters' invariance for ensuring construct equivalence. Particularly, some researchers have believed that factor loading invariance is essential evidence for construct equivalence (e.g., Meredith & Teresi, 2006). Because factor loadings indicate the relation between observed variables and latent constructs, those researchers argued that factor loading equivalence means that the latent constructs are inferred from the observed variables in the same way across populations. Thus, if there is no evidence of measurement invariance, particularly factor loading invariance, as stated by Horn and McArdle, "the basis for drawing scientific inference is severely lacking: findings of differences between individuals and groups cannot be unambiguously interpreted" (Horn & McArdle, 1992, p.117).Therefore, some researchers believe that a key procedure of collecting evidence of construct equivalence is to assess factor loading invariance (Cheung & Rensvold, 1999; Little, 1997; Meredith & Teresi, 2006) and thus that measurement invariance is a prerequisite for meaningful and valid group comparison on latent constructs.

In reality, however, the assumption of measurement invariance has been found to be hard to achieve (Schmitt & Kuljanin, 2008) and thus some researchers have proposed the idea of less stringent measurement invariance (i.e., partial measurement invariance; Byrne, Shavelson, & Muthén, 1989) which assumes that only a subset of measurement model parameters is invariant. These researchers have argued that partial measurement invariance would be sufficient to make valid group comparisons (Byrne et al., 1989; Millsap & Kwok, 2004; Steenkamp & Baumgartner, 1998) if researchers have strong beliefs, empirical results, or substantive theory on construct equivalence across populations. This more lenient view implicitly implies that the underlying latent constructs are assumed to be the same although measurement model parameters may differ across populations. It also implies that measurement invariance may not be a necessary condition for meaningful group comparison on latent constructs.

In fact, previous literature has found that although a latent construct measured by a set of items is equivalent, nonequivalence of measurement model parameters can still occur through, for example, translation errors, different response tendencies, and different degrees of familiarity with item format (Bolt, 2000; Taylor & Lee, 2012; van de Vijver & Tanzer, 2004). In addition, in the educational measurement field, detecting different measurement model parameters across populations has received great attention in the literature. It should be noted that testing of equality of measurement model parameters within item response theory (e.g., difficulty parameters, discrimination parameters) is based on the implicit assumption that the items measure the same underlying construct in all populations (Kim, Cohen, & Park, 1995). That is, the presence of noninvariant items in terms of measurement model parameters does not imply that the underlying construct measured by the test items is different. Instead, it may imply that items simply work differently for some reasons that are not related to construct equivalence (Reise, Smith, & Furr, 2001).

A viewpoint of this dissertation adopted here would be that differences in measurement model parameters across populations, particularly factor loading differences, do not necessarily indicate construct non-equivalence. Construct equivalence can be possible such that the same construct is measured but measurement parameters may not be the same across groups (van de Vijver & Tanzer, 2004). In addition, in some situations where minor differences in measurement model parameters across populations are expected across populations, a small degree of measurement noninvariance could be tolerable for group comparison on latent constructs. Under some situations where researchers have sufficient theories on latent construct equivalence across populations, thus, differences in measurement model parameters may still support the group comparison on latent constructs to be meaningful and valid. Given that many more studies present measurement models that do not exhibit invariance across populations, one of the most important issues is how researchers can get accurate and valid group comparison conclusions on underlying constructs particularly when measurement invariance does not hold. Although much work has been done in terms of the methods and procedures for detecting noninvariant model parameters, relatively little attention has been paid to the implications of failing measurement invariance (e.g., partial measurement invariance) on group comparison on the underlying latent constructs. In order to fill these gaps in the literature, this study particularly concerns the impact of noninvariance within the measurement model on group comparison of latent constructs.

This study adopts a Bayesian approach to investigate the impact of measurement noninvariance because it provides a more practical argument in tests of measurement invariance. With typical MGCFA with maximum likelihood, all measurement model parameters are tested for null hypotheses of exact equality in terms of model parameters, which is often neither realistic nor practical in tests of measurement invariance. Using a Bayesian analysis strategy has the potential to provide a more flexible approach to address these limitations than a frequentist approach in invariance testing. Therefore, the purpose of the current dissertation is to empirically investigate the extent to which measurement noninvariance affects group comparison on latent constructs across populations taking a Bayesian approach. A Monte Carlo simulation approach was conducted to investigate the sensitivity of the posterior distribution of two structural parameters (i.e., structural regression coefficient difference, factor mean difference) to varying degrees of noninvariant measurement models across populations.

The following chapter, Chapter 2, presents a review of existing research, providing a context and theoretical framework for Bayesian structural equation modeling, measurement invariance, and group comparison within CFA framework and motivation for the current study. Chapter 3 specifies the design of the current study, including the data generating model, the simulation design factors, and the data analysis procedures. Results are reported and described in Chapter 4. Chapter 5 presents a summary and discussion of results, implications for practice, and limitations and directions for future study.

Chapter 2: Literature Review

Theoretical Background of Bayesian Structural Equation Modeling

Bayesian structural equation modeling (BSEM) has been well recognized as an attractive approach to analyze a variety of structural equation models (SEMs) (Lee, 2007). The basic attractive feature of a Bayesian approach is that researchers can incorporate useful prior knowledge into statistical analysis, potentially yielding better results. Therefore, BSEM takes different statistical properties and procedures for analyzing SEMs from the traditional maximum likelihood SEM. This section describes a general BSEM approach in the context of CFA. The Bayesian approach to CFA will be introduced, including Bayesian inference, Bayes' theorem, Bayesian estimation, and model assessment. Further special issues that can occur specifically with Bayesian CFA are also outlined.

Bayesian Inference

Bayesian inference differs from frequentist inference (e.g., maximum likelihood estimation) in two distinct ways. The first key difference is a way of viewing unknown parameters (Fox, 2010; Kaplan & Depaoli, 2012; Lee, 2007; Levy & Choi, 2013). In the frequentist approach, an unknown parameter is assumed to be fixed and hence the principle is to find *parameter estimates* to make inferences about that fixed parameter. For example, maximum likelihood (ML) estimation is one of the most commonly used estimation method in SEM and serves as a default in most SEM computer programs (e.g., LISREL, EQS, and Mplus). In the context of SEM, ML estimation seeks to find the best model point parameter estimates that yield the maximum likelihood of the observed data.

The frequentist approach actually focuses on the parameter estimate (Levy & Choi, 2013). In Bayesian inference, however, an unknown parameter is assumed to be random with a distribution and hence the unknown parameter is estimated as a form of distribution for the parameter (i.e., posterior distribution). The posterior distribution for the parameter is constructed by combining observed data with prior knowledge or information, which is a second key difference. In a Bayesian analysis, researchers need to assign a prior distribution for each model parameter to reflect the researchers' prior knowledge, belief, and/or substantive theory and combine these prior distributions with data at hand for making inferences about parameters. Accurate results can be obtained by incorporating appropriate prior distributions into an analysis.

In a Bayesian analysis, inferences about parameters are drawn directly from the posterior distributions of the parameters of interest. Following the description given by Muthén and Asparouhov (2012, p. 315) and Kaplan and Depaoli (2012, p. 651), a joint probability distribution of events A and B can be written using conditional and marginal probabilities:

$$P(A,B) = P(A|B)P(B) = P(B|A)P(A) \tag{1}$$

where $P(A,B)$ denotes a joint probability distribution of events A and B, $P(A|B)$ represents a conditional probability distribution of event A given event B, $P(B|A)$ represents a conditional probability distribution of event B given event A. $P(A)$ and $P(B)$ represent marginal distributions of event A and B, respectively. Equation 1 can be rewritten as:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}, \tag{2}$$

which is Bayes' theorem. Bayesian inference computes the posterior probability, which is $P(B|A)$ in Equation 2, according to Bayes' theorem. With an unknown parameter (θ) given data (y), the posterior probability distribution of the unknown parameter (θ) given data (y) is

$$P(\theta \mid y) = \frac{P(y \mid \theta)P(\theta)}{P(y)}. \tag{3}$$

In Equation 3, $P(\theta \mid y)$ is a posterior probability distribution of unknown parameter (θ) given data (y), $P(y \mid \theta)$ is the conditional distribution of the data (y) given the parameter (θ), $P(\theta)$ is the prior distribution of the parameter (θ), and $P(y)$ is the marginal distribution of the data (y). Equation 3 states that the posterior distribution is a product of the conditional distribution of the data (y) given the parameter (θ) and the prior distribution of the parameter (θ), normalized by the marginal distribution of the data (y), so that the posterior distribution integrate to one. Because the marginal distribution of the data (y) does not involve the parameter (θ), dropping the marginal distribution of the data (y) yields the unnormalized posterior distribution, which is expressed as:

$$P(\theta \mid y) \propto P(y \mid \theta)P(\theta). \tag{4}$$

Although the posterior distribution is a probability in Equation 4, the area of posterior distribution is no longer 1, being posterior distribution proportional to the conditional distribution of data given the parameter times the prior distribution. Given that the conditional distribution of the data given the parameter is equivalent to the likelihood distribution of the parameters given the data (i.e., $P(y \mid \theta) = L(\theta \mid y)$), Equation 4 is equivalent to

$$P(\theta \mid y) \propto L(\theta \mid y)P(\theta)\,. \tag{5}$$

Equation 5 implies that a Bayesian inference is drawn from the posterior distribution that is constructed by combining the likelihood distribution of data with the parameter's prior distribution (Kaplan & Depaoli, 2012; Lee, 2007).

Bayesian Approach to the CFA Model

The traditional form of a CFA measurement model with *p* indicators of *m* latent variables can be expressed as:

$$\mathbf{Y} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\delta}\,, \tag{6}$$

where $\mathbf{Y}$ is the $p\times1$ vector of observed scores, $\boldsymbol{\tau}$ is the $p\times1$ vector of intercepts, $\boldsymbol{\Lambda}$ is the $p\times m$ factor loading matrix, $\boldsymbol{\xi}$ is the $m\times1$ vector of theoretical latent variable scores, and $\boldsymbol{\delta}$ is the $p\times1$ vector of error terms. It is typically assumed that latent variables are not correlated with error terms and hence the covariance structure of observed variables can be written as:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}\,, \tag{7}$$

where $\boldsymbol{\Sigma}$ is the model-implied covariance matrix, $\boldsymbol{\Phi}$ is the matrix of latent variables' variances and covariances, and $\boldsymbol{\Theta}$ is the matrix of error variances and covariances (if any). Further, assuming that means of error terms are zero, the expectation of the observed variables $\mathbf{Y}$ is:

$$E(\mathbf{Y}) = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\kappa}\,, \tag{8}$$

where $\boldsymbol{\kappa}$ is the vector of latent variables' means.

Unlike the traditional form of a CFA measurement model, a Bayesian approach to a CFA measurement aims to find posterior distributions for the unknown parameters. For a Bayesian analysis of one factor model, for instance, the posterior distributions of unknown parameters such as intercepts, factor loadings, error variances, latent variables' variances, and latent variables' means can be expressed as follows:

$$P(\boldsymbol{\tau},\boldsymbol{\Lambda},\boldsymbol{\Theta},\boldsymbol{\Phi},\boldsymbol{\kappa} \mid \mathbf{Y}) = \frac{P(\mathbf{Y} \mid \boldsymbol{\tau},\boldsymbol{\Lambda},\boldsymbol{\Theta},\boldsymbol{\Phi},\boldsymbol{\kappa})P(\boldsymbol{\tau})P(\boldsymbol{\Lambda})P(\boldsymbol{\Theta})P(\boldsymbol{\Phi})P(\boldsymbol{\kappa})}{P(\mathbf{Y})}, \tag{9}$$

where $P(\mathbf{Y} \mid \boldsymbol{\tau},\boldsymbol{\Lambda},\boldsymbol{\Theta},\boldsymbol{\Phi},\boldsymbol{\kappa})$ represents the conditional distribution of data given the unknown parameters ($\boldsymbol{\tau},\boldsymbol{\Lambda},\boldsymbol{\Theta},\boldsymbol{\Phi},$ and $\boldsymbol{\kappa}$), and $P(\boldsymbol{\tau})$, $P(\boldsymbol{\Lambda})$, $P(\boldsymbol{\Theta})$, $P(\boldsymbol{\Phi})$, and $P(\boldsymbol{\kappa})$ are the prior distributions for intercepts, factor loadings, error variances, latent variables' variances, and latent variables' means, respectively. It is assumed that the prior distribution of a specific parameter is independent of the prior distribution of the other parameters. It should be noted that in a Bayesian analysis, all unknown parameters are assigned prior distributions. Literature has stated that specification of correct prior distributions plays an important role in the Bayesian inference on the unknown parameters in CFA (Fox, 2010; Kaplan & Depaoli, 2012; Levy & Choi, 2013; MacCallum, Edwards, & Cai, 2012; Muthén & Asparouhov, 2012). Some Bayesian SEM resources (Kaplan & Depaoli, 2012; Palomo, Dunson, & Bollen, 2007; Lee, 2007; Levy & Choi, 2013) have suggested that the choice of prior distributions should be based on substantive theory or previous empirical results about the parameters. When researchers have strong prior knowledge about model parameters, the information can be added into the model to estimate posterior distribution of the model parameters. In this case, such prior distributions are referred to as *informative priors*. In Bayesian SEM textbooks, often *conjugate* prior distributions are recommended as informative priors (Fox, 2010; Kaplan & Depaoli, 2012; Lee, 2007; Levy & Choi, 2013). A conjugate prior distribution describes a prior distribution that results in the posterior distribution following the same distributional form as the prior distribution (Gelman, Carlin, Stern, & Rubin, 2004). As an example of using the univariate binominal model (e.g., success vs. fail), success follows the binominal distribution and the probability of getting y success in n trials can expressed as:

$$P(y \mid \theta) = \binom{n}{y} \theta^{y} (1-\theta)^{n-y} . \tag{10}$$

Suppose that a beta distribution is regarded as a prior distribution. The probability density function of the beta distribution follows as:

$$P(\theta) = k * \theta^{\alpha-1} (1-\theta)^{\beta-1} , \tag{11}$$

where k is constant. When the beta prior distribution is combined with the likelihood function, the posterior distribution follows the beta distribution, similar to the prior distribution.

$$\begin{aligned} P(\theta \mid y) &\propto P(y \mid \theta) P(\theta) \\ &\propto \binom{n}{y} \theta^{y} (1-\theta)^{n-y} \, k * \theta^{\alpha-1} (1-\theta)^{\beta-1} \\ &\propto \binom{n}{y} k \, \theta^{y} (1-\theta)^{n-y} \theta^{\alpha-1} (1-\theta)^{\beta-1} \\ &\propto \theta^{y+\alpha-1} (1-\theta)^{n-y+\beta-1} \end{aligned} \tag{12}$$

In this example, the beta distribution is called a conjugate prior distribution. In a Bayesian approach, choice of conjugate prior distribution is a convenient feature because use of conjugate prior distributions yields a posterior distribution of known form that is analytically to solve (Kaplan & Depaoli, 2012; Levy & Choi, 2013). If a prior distribution is not a conjugate prior distribution, the posterior distribution is often not a known form and thus the estimation of the posterior distribution may not be analytically solved. In this case, a special estimation, the Markov chain Monte Carlo estimation, can be used to obtain the posterior distribution (Kaplan & Depaoli, 2012). In a typical SEM analysis, a normal distribution with small variance is often used as a prior distribution for a factor loading, intercept, or mean of a latent variable because the normal prior distribution is a conjugate prior distribution for the parameter. For the same reason, an inverse-gamma distribution is used as a prior distribution for an error variance and inverse-Wishart distribution is typically used as a prior distribution for a variance or covariance of a latent variable. Following the description given by Levy and Choi (2013), these conjugate prior distributions for the unknown parameters in Equation 9 can be expressed as:

$$\begin{aligned}
&\lambda \sim N(\mu_\lambda, \sigma_\lambda^2), \\
&\tau \sim N(\mu_\tau, \sigma_\tau^2), \\
&\theta \sim G^{-1}(\alpha_\theta, \beta_\theta), \\
&\kappa \sim N(\mu_\kappa, \sigma_\kappa^2), \\
&\Phi \sim W^{-1}(d\Phi_0, d)
\end{aligned} \tag{13}$$

As seen in Equation 13, these conjugate prior distributions have their own parameters, which are referred to as *hyperparameters*. Existing Bayesian textbooks (e.g., Kaplan & Depaoli, 2012; Lee, 2007; Levy & Choi, 2013) recommend using fixed known values for the hyperparameters in that they reduce computational complexity. In this case, the values of hyperparameters μ_λ, σ_λ^2, μ_τ, σ_τ^2, α_θ, β_θ, μ_κ, σ_κ^2, $d\Phi_0$, and d should be chosen to be consistent with researchers' prior knowledge (Lee, 2007). If the hyperparameters in the conjugate prior distributions are not known, researchers may use ML estimates for the parameters obtained from part of the data (Lee, 2007). It should be noted that by specifying the values of hyperparameters by researchers, it is implicitly assumed that the prior knowledge on model parameters is known without any uncertainty. When prior knowledge on model parameters is based on the estimates from previous data analyses, however, the prior knowledge is subject to uncertainty. In this case, the hyperparameters can be treated as unknown parameters and hence have their own prior distributions, resulting in a fully Bayesian approach. The advantage of the fully Bayesian approach is that all uncertainties are fully accounted for in the analyses and hence yield estimates with realistic standard errors (Bernardinelli & Montomoli, 1992; Carriquiry, & Pawlovich, 2004).

When there is no existing substantive theory, knowledge, or empirical results about parameters, noninformative or *diffuse* prior distributions can be assigned to parameters instead of informative prior distributions. Either uniform distributions or conjugate prior distributions with very large variance are commonly recommended as noninformative prior distributions in the literature (Lee, 2007). For example, default prior distributions used in Mplus are normal distributions with a means of zero and variances of 10^{10} for the factor loadings and intercepts. It should be noted that the current Bayesian textbooks do not recommend the use of the default prior distributions used in some software programs (e.g., Mplus). The default prior distribution for the error variance used in Mplus, for instance, is an inverse gamma distribution with hyperparameters -1 and 0. Given that this prior is uniformly 1 on the entire line from minus infinity to plus infinity, an inverse gamma distribution with hyperparameters -1 and 0 can be an *improper* prior distribution and hence is not recommended to be used in other software programs (e.g., WinBUGS). Although the Mplus developers argued that such an improper prior distribution was selected based on extensive simulation studies and has little effect on estimation (Mplus team, email communication, July 11, 2014), more thorough investigation might be required to assess the impact of such an improper prior distortion on Bayesian estimation.

Markov Chain Monte Carlo Estimation

In a Bayesian analysis, the posterior distribution is commonly obtained through Markov chain Monte Carlo (MCMC) estimation. MCMC estimation in Bayesian analyses is an algorithm that approximates the posterior distributions by repeatedly drawing a series of values of unknown parameters from approximate distributions (Gelman et al., 2004). For MCMC estimation, there are several algorithms to estimate actual posterior distributions such as Gibbs sampling, Metropolis-Hastings sampling, and Metropolis sampling.

Among them, Gibbs sampling is one of the popular algorithms in Bayesian SEM (Fong & Ho, 2013; Golay, Reverte, Rossier, Favez, & Lecerf, 2012; Kaplan & Depaoli, 2012; Muthén & Asparouhov, 2012). Following the description given in Levy and Choi (2013), Gibbs sampling begins with initial values for the parameters, denoted as $\theta_1^{(0)}$, $\theta_2^{(0)}, \ldots, \theta_R^{(0)}$, where $\theta_r^{(t)}$ denotes the value of model parameter *r* at iteration *t*. Given the starting point, values for parameter θ are repeatedly drawn from its full conditional distribution given the observed data and the current values of all other model parameters. In other words, for each parameter θ_r, we obtain the $t+1^{\text{st}}$ iteration value of the chain by drawing from $P(\theta_r \mid Y, \theta_1^{(t+1)}, \ldots, \theta_{r-1}^{(t+1)}, \theta_{r+1}^{(t)}, \ldots, \theta_R^{(t)})$. One cycle is given by sequentially drawing values from

$$\theta_1^{(t+1)} \sim P(\theta_1 \,|\, Y, \theta_2^{(t)}, \ldots, \theta_R^{(t)})$$

$$\theta_2^{(t+1)} \sim P(\theta_2 \,|\, Y, \theta_1^{(t+1)}, \theta_3^{(t)}, \ldots, \theta_R^{(t)})$$

$$.$$

$$.$$

$$\theta_R^{(t+1)} \sim P(\theta_R \,|\, Y, \theta_1^{(t+1)}, \ldots, \theta_{R-1}^{(t+1)}) .$$

This step is repeated for a large number *t* iterations until the posterior distributions are considered to be converged according to some convergence criterion. Although the Gibbs sampling method has been commonly used in Bayesian software programs (e.g., WinBUGS), it should be noted that the Gibbs sampling method tends to get stuck, leading to slow convergence when there is a high level of posterior correlations between parameters or when posterior distributions are bimodal (Justel & Pena, 1996; Raftery & Lewis, 1992a; Smith & Roberts, 1993).

Typically, multiple MCMC chains and large number of iterations are commonly used for the determination of convergence of the MCMC process. The multiple chains are in parallel and independent in that each chain has different staring values and different random seeds for the random draws of values for unknown parameters (Muthén & Asparouhov, 2012). Convergence of the MCMC process can also be obtained from one single chain with a considerable larger number of iterations. Regarding the number of chains necessary for Bayesian inferences and diagnostics, some researchers recommend to use one chain with a large number of iterations because use of multiple chains has few benefits and one chain with a large number iterations performs equally well in most standard statistical models (Geyer, 1991; Raftery & Lewis, 1992b). However, it should be noted that use of one chain in Bayesian analysis may lead to slow convergence if a random starting value is poorly chosen. With a single chain, thus, a starting value should be carefully chosen based on preliminary experimentation (Raftery & Lewis, 1992b). For the multiple chain situations, convergence of the MCMC process can be investigated by a formal comparison of the between-chain variance with within-chain variance, referred to as the *potential scale reduction factor* (PSRF) (Gelman et al., 2004), is commonly recommended (Fong & Ho, 2013; Gelman et al., 2004; Golay et al., 2013; Kaplan & Depaoli, 2012; Muthén & Asparouhov, 2012). The PSRF can be calculated as (e.g., Muthén & Asparouhov, 2012)

$$PSRF = \sqrt{\frac{Var_{within} + Var_{between}}{Var_{within}}}\,, \tag{14}$$

$$Var_{within} = \frac{1}{c}\sum_{j=1}^{c}\frac{1}{t}\sum_{i=1}^{t}(\theta_{ij} - \bar{\theta}_{.j})^2\,,$$

$$Var_{between} = \frac{1}{c-1}\sum_{j=1}^{c}(\bar{\theta}_{.j} - \bar{\theta}_{..})^2\,,$$

where Var_{within} is a within-chain variance, $Var_{between}$ is a between-chain variance, c is the number of chains, t is the number of iterations, and θ_{ij} denotes the value of parameter θ in the t^{th} iteration of chain j. With a single chain, PSRF can be calculated using the third and the fourth quarters of the chain (Muthén & Muthén, 1998-2012). Equation 11 states that when PSRF is near 1, between-chain variance is relatively small compared to within-chain variance and hence implies the convergence. Previous literature has suggested that if PSRF values are between 1 and 1.1, satisfactory convergence is achieved (Fong & Ho, 2013; Gelman et al., 2004, Golay et al., 2013; Kaplan & Depaoli, 2012; Muthén & Asparouhov, 2012).

Model Fit Assessment and Significance Tests

In Bayesian SEM, model fit assessment is commonly conducted using posterior predictive model checking (PPMC). PPMC involves generating a posterior predicted distribution and a posterior predicted dataset. Conceptually, the posterior predicted distribution is the distribution of future observations from both the observed data and the model (Levy & Choi, 2013). Following the descriptions given by Levy and Choi, first the posterior predicted distribution is constructed via simulation by taking random draws from the posterior distributions. Letting random draws of vector of model parameters denoted as, $\theta^{(1)}$, $\theta^{(2)}$,…, $\theta^{(t)}$, where $\theta^{(t)}$ denotes the vector of model parameters at t^{th} draw. Using those drawn parameters, a posterior predicted dataset is generated conditional on the observed data and the model. Then, the discrepancy measure is evaluated between the random draws for the model parameters and the posterior predicted datasets at each draw. PPMC works by comparing discrepancy between model parameters and the observed data with model parameters and the posterior predicted dataset. This can be done using posterior predicted p-value (PPP) and is defined as (Kaplan & Depaoli, 2012),

$$PPP = p(D(y_{predicted} - \theta \mid y) \geq D(y - \theta \mid y), \tag{15}$$

where $D(y_{predicted} - \theta \mid y)$ is a discrepancy measure between model parameters and the posterior predicted data and $D(y - \theta \mid y)$ is a discrepancy measure between model parameters and the observed data. Perfect model fit is expected to have a PPP of .5 (Lee, 2007).

In practice, researchers are often interested in testing the statistical significance associated with the estimates (e.g., factor loading, factor mean) as well as model fit assessment. In the Bayesian approach, inference about parameters is based on their posterior distributions. The posterior distributions are often summarized into point estimates (e.g., posterior means, posterior median, posterior standard deviations) or interval estimates (e.g., 95% credibility intervals) for Bayesian hypothesis testing. In general, the point estimates based on the posterior distributions are related to the loss function. The posterior mean is associated with the square loss function, which finds parameter estimates by minimizing the mean squared error between the true parameters and corresponding parameter estimates. The posterior median is linked to the absolute deviation loss function, which finds parameter estimates by minimizing the absolute error between the true parameters and corresponding parameter estimates. When the true parameter is continuous, the posterior mean has been commonly recommended and used for Bayesian hypothesis testing (Kaplan & Depaoli, 2012). With continuous indicator variables, the mean of the posterior distribution can be calculated as (Kaplan & Depaoli, 2012):

$$E(\theta \setminus y) = \int_{-\infty}^{+\infty} \theta \, p(\theta \mid y) d\theta \quad . \tag{16}$$

The posterior mean is often referred to as the *expected a posteriori* or EAP estimate. Similarly, the variance of θ can be written as,

$$var(\theta \mid y) = E(\theta^2 \mid y) - E(\theta \mid y)^2 \; . \tag{17}$$

In addition, 95% central credibility intervals can be constructed by computing the 2.5th and 97.5th percentiles of the posterior distribution. Unlike the frequentist approach, the 95% central credibility interval might not be symmetric.

Issues in Bayesian SEM

There are some issues to consider when employing Bayesian analysis. First of all, the choice of prior distributions can have a great impact on both the results on posterior distribution of parameters and the performance of the MCMC estimation (Kaplan & Depaoli, 2012; MacCallum et al., 2012), and thus researchers should be careful when selecting prior distributions for parameters. In general, prior distributions should be chosen based on substantive knowledge about the parameters of interests (Kaplan & Depaoli, 2012; Lee, 2007; Levy & Choi, 2013). When researchers have strong knowledge about parameters such as factor loadings or intercepts, this knowledge can be incorporated into the model. In such cases, informative prior distributions can be assigned to parameters. For the factor loadings, for example, normal distributions can be used as informative prior distributions and their means and standard deviations can be specifically chosen to reflect a researcher's prior knowledge. As the strength of researchers' knowledge increases for parameters, the variance of normal distribution would decrease to reflect confidence of precision in the prior distribution. On the other hand, researchers might not have prior knowledge about parameters and in this case noninformative prior distributions can be used. As noninformative priors, normal distribution with mean and large variance or uniform distribution can be assigned to factor loadings to reflect little prior knowledge about those parameters. As Levy and Choi (2013) mentioned, the prior distributions do not need to follow the same prior distributions for parameters. In other words, based on substantive knowledge about the parameters, the prior distribution for each factor loading can differ by specifying different prior means and variances for different factor loadings or take different distributional forms. Similar approaches are applied to the prior distributions for intercepts, error variances, and factor variances. Although Bayesian SEM resources (Kaplan & Depaoli, 2012; Levy & Choi, 2013; Palomo et al., 2007) provide these general guidelines for prior distributions, it may not be easy to choose prior distributions. As MacCallum, Edwards, and Cai (2012) mentioned, when choosing prior distributions about parameters, researchers need to make decisions on the values of hyperparameters governing selected distributions (i.e., prior distributions) as well as the forms of distributions. The problem is that applied researchers do not often have such detailed knowledge about parameters in reality. Therefore, it is recommended to perform sensitivity analysis of different choice of prior distribution on parameter estimates before making inference on parameters (e.g., MacCallum et al., 2012).

Theoretical Framework for Measurement Invariance

Measurement Invariance

Measurement invariance conceptually expresses the idea that a measurement instrument designed to capture underlying latent constructs operates in the same way over times or in different populations (Horn & McArdle, 1992; Millsap, 2011; Vandenberg & Lance, 2000). When the same instrument is administered over repeated occasions, for instance, measurement invariance holds if the latent construct assesses in the same way across time points, which is often referred to as longitudinal invariance. More often, measurement invariance is a research of interest to many applied researchers who wish to compare latent construct across groups. In this case, measurement invariance holds if the latent construct operates in the same way across different populations. This conceptual definition of measurement invariance across different populations is also expressed as probability terms in the previous literature (Millsap, 2011; Wu et al., 2007) such as

$$P(Y\,|\,F,G) = P(Y\,|\,F), \tag{18}$$

where Y is the observed score from a measurement instrument, F is the theoretical latent construct score, and G represents group membership. Equation 18 states that the conditional probability of attaining the observed scores of Y given the theoretical latent construct score, F, is independent of the group membership. In measurement invariance testing, groups are often categorized as a reference group and a focal group. In general, the majority group is referred to as a *reference group* while the minority group is referred to as a *focal group*. The reference group becomes a basis of reference for making comparison and the focal group becomes a typical concern for investigating measurement invariance. This conceptual probabilistic definition of measurement invariance is defined as the equality of measurement model parameters across groups from the statistical standpoint (e.g., MGCFA) and will be further discussed in the next section.

<u>Measurement Invariance in MGCFA Model</u>

In the SEM framework, the tests of measurement invariance are typically conducted through the multi-group confirmatory factor analysis (MGCFA) model. The following equation describes the MGCFA model with continuous observed variables for multiple groups. The measurement model with *p* observed variables and *m* common factors in MGCFA is specified as:

$$\mathbf{Y_G} = \boldsymbol{\tau}_\mathbf{G} + \boldsymbol{\Lambda}_\mathbf{G}\boldsymbol{\xi}_\mathbf{G} + \boldsymbol{\delta}_\mathbf{G}, \tag{19}$$

where G indicates group membership. For group G, $\mathbf{Y_G}$ is the $p\times1$ vector of observed scores, $\boldsymbol{\tau}_\mathbf{G}$ is the $p\times1$ vector of intercepts, $\boldsymbol{\Lambda}_\mathbf{G}$ is the $p\times m$ matrix of factor loadings, $\boldsymbol{\xi}_\mathbf{G}$ is the m×1 vector of theoretical latent construct scores, and $\boldsymbol{\delta}_{\mathbf{G}}$ is the p×1 vector of error terms. It is typically assumed that latent construct scores are not correlated with error variances and hence the covariance structure of observed variables, $\mathbf{Y}$, in group G can be written as,

$$\boldsymbol{\Sigma}_{\mathbf{G}} = \boldsymbol{\Lambda}_{\mathbf{G}}\boldsymbol{\Phi}_{\mathbf{G}}\boldsymbol{\Lambda}'_{\mathbf{G}} + \boldsymbol{\Theta}_{\mathbf{G}}, \tag{20}$$

where, for group G, $\boldsymbol{\Sigma}_{\mathbf{G}}$ represents the matrix of the variance and covariance of observed variables , $\boldsymbol{\Phi}_{\mathbf{G}}$ is the matrix of the variances and covariances of the latent constructs, and $\boldsymbol{\Theta}_{\mathbf{G}}$ is the matrix of error variances and covariances. Equation 18 states that the variances and covariances of observed variables are functions of three types of model parameters, which are factor loadings, variances, and covariances of latent constructs, and error variances. Equation 20 suggests that if the factor loadings and error variances for observed variables are equivalent across groups (i.e., $\boldsymbol{\Lambda}_{1} = \boldsymbol{\Lambda}_{2} = \ldots = \boldsymbol{\Lambda}_{\mathbf{G}}$), the difference in variances and covariances for the observed variables across groups can be a true representation of the difference in variances and covariances for the latent construct. Further, under the assumption of zero means for error terms, the expectation of observed variables in group G is written as,

$$E(\mathbf{Y}_{\mathbf{G}}) = \boldsymbol{\tau}_{\mathbf{G}} + \boldsymbol{\Lambda}_{\mathbf{G}}\boldsymbol{\kappa}_{\mathbf{G}}, \tag{21}$$

where for group G, $\boldsymbol{\kappa}_{\mathbf{G}}$ is the mean of the latent constructs. Equation 21 states that the means of observed variables are functions of three parameters: intercepts, factor loadings, and means of latent constructs. Equation 21 also suggests that if the factor loadings and intercepts for observed variables are equivalent across groups, the difference in means for the observed variables is a direct reflection of differences in latent construct means.

Meredith (1993) and Meredith and Teresi (2006) described three types of invariance for the measurement model in the MGCFA framework: weak, strong, and strict measurement invariance. *Weak invariance* has factor loading invariance across populations ($\mathbf{\Lambda}_1 = \mathbf{\Lambda}_2 = \ldots = \mathbf{\Lambda}_G$) while intercepts and error variances may vary across populations. *Strong measurement invariance* indicates that both the factor loadings and intercepts of observed variables are the same across populations ($\mathbf{\Lambda}_1 = \mathbf{\Lambda}_2 = \ldots = \mathbf{\Lambda}_G$ and $\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \ldots = \boldsymbol{\tau}_G$) while error variances may still vary across populations. Finally, with *strict measurement invariance*, error variances as well as factor loadings and intercepts of measures are assumed to be the same across populations ($\mathbf{\Lambda}_1 = \mathbf{\Lambda}_2 = \ldots = \mathbf{\Lambda}_G$, $\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \ldots = \boldsymbol{\tau}_G$, and $\mathbf{\Theta}_1 = \mathbf{\Theta}_2 = \ldots = \mathbf{\Theta}_G$).

Frequentist Approach to Tests of Measurement Invariance

Historically, tests of measurement invariance have been primarily conducted through the frequentist approach in both applied studies and methodological studies. Often, maximum likelihood estimation for continuous indicators and weighted least squares estimation for ordered categorical indicators have been commonly used in tests of measurement invariance in the frequentist approach. In the frequentist approach, measurement invariance tests involve assessing whether model parameters are exactly the same across populations.

In general, measurement invariance tests are conducted in hierarchical order: weak invariance, strong invariance, and strict invariance. In order to test each level of measurement invariance, two measurement models (i.e., unconstrained model vs. constrained model in terms of factor loadings, intercepts, and/or error variances) are fit to the same sample data and are compared, typically using chi-square statistics or goodness of fit indices (GOFs). Although the chi-square difference test (likelihood ratio test) is most frequently used, it has been found to be highly sensitive to sample size in invariance testing (Chen, 2007; Cheung & Rensvold, 2002; Meade, Johnson, & Braddy, 2008). For this reason, recent research has suggested to use alternative GOFs in invariance testing that are not sample size sensitive. Of these studies, Cheung and Rensvold's work is particularly important in that they specified four desirable properties of change of GOFs (ΔGOFs) used for testing measurement invariance. These include the following: (1) ΔGOFs should not be sensitive to the overall fit in the baseline model, (2) ΔGOFs should not be sensitive to model complexity, (3) ΔGOFs should not be redundant with other GOFs, and (4) ΔGOFs should not be sensitive to sample size. Following these four criteria, they examined 20 GOFs based on the minimum value of the fit function through the simulation works and found that ΔCFI, ΔGamma-hat, ΔMcDonald's NCI, ΔIFI, and ΔRNI have those desirable properties. Due to high correlation among ΔIFI, ΔCFI, and ΔRNI, they suggested reporting only one of these three indices. Given that CFI is a popular index in CFA, they recommended using ΔCFI, ΔGamma hat, and ΔMcDonald's NCI to assess measurement invariance. Further, they provided empirically derived cutoff values for ΔCFI, ΔGamma-hat, and ΔMcDonald's NCI that were .01, .001, and .02, respectively, at an α-level of .01 across all types of invariance tests. Chen (2007) and Meade and colleagues (2008) extended Cheung and Rensvold's (2002) study by further examining the performance of these ΔGOFs detecting a lack of measurement invariance. In Chen's study, it was found that ΔCFI, ΔRMSEA, ΔGamma Hat, and ΔMcDonald's NCI performed well in tests of measurement invariance in terms of Type I error and power. Hence, Chen provided empirically derived cutoff values for ΔCFI, ΔRMSEA, ΔGamma Hat, and ΔMcDonald's NCI that were .005, .01, .005, and .01, respectively, at an α-level of .01 across three types of invariance tests. Similar to Cheung and Rensvold's (2002) study, they recommended using one of the following four indices for tests of measurement invariance: ΔGamma-hat, ΔIFI, ΔRNI, and ΔCFI and ΔMcDonald's NCI. The same authors provided a common cutoff value for ΔCFI, .002, to assess either weak or strong factorial invariance while they provided empirically derived cutoff values for ΔMcDonald's NCI based on the number of factors and items. Although these studies suggested using common cutoff values for some ΔGOFs, however, use of the common cutoff values particularly for CFI should be done cautiously. According to Kang and Hancock's (2013) recent study, the cutoff values particularly for the ΔCFI were greatly influenced by measurement quality and sample size. Their simulation results showed that as sample size or factor loading size increased, the cutoff values for the ΔCFI was smaller, indicating that the use of a common cutoff value for the ΔCFI in measurement invariance testing may be inappropriate regardless of measurement quality. However, ΔMcDonald's NCI was not affected by sample size and measurement quality, indicating that the use of a common cutoff value for ΔMcDonald's NCI in measurement invariance testing may be appropriate regardless of measurement quality.

Although measurement invariance tests from the frequentist approach have been widely conducted in applied studies, there are several issues to consider when employing a frequentist approach to measurement invariance. First of all, the selection of a reference indicator is of great importance in measurement invariance testing. Given that latent variables have no defined metric, every latent variable must be assigned its own scale in order to make the model and the variables' implied characteristics identified. When the frequentist approach to measurement invariance tests is employed, such scaling is most commonly accomplished by linking the metric of the latent variable to one of its measured indicators by fixing the associated loading (i.e., reference indicator). For the remainder of the dissertation, the chosen indicator is referred to as the *reference indicator*. Linking based on a reference indicator could be problematic in measurement invariance testing when the true factor loading of the reference indicator differs across populations. Previous literature demonstrated that the noninvariance of the reference indicator is likely to cause severe problems because all parameter estimates are adjusted by the different scaling constants across populations and hence make invalid comparison of measurement parameters (Cheung & Rensvold, 1999; Hancock, Stapleton, & Arnold-Berkovits, 2009; Johnson, Meade, & DuVernet, 2009; Stark, Chernyshenko, & Drasgow, 2006). Despite this issue, in the reality of measurement invariance studies, a reference indicator has typically been selected with relatively little consideration or possibly even by default in a given software package (Schmitt & Kuljanin, 2008). Although methodological researchers have proposed several ways to identify an invariant indicator across groups for the reference indicator (e.g., Cheung & Rensvold, 1999; Yoon & Millsap, 2007), these methods are not commonly used because they can be labor intensive and/or ineffective. For example, Yoon and Millsap (2007) found that their proposed approach to find invariant factor loadings consistently failed to find noninvariant items when there are large number of noninvariant items, small samples, and small differences between parameters. For this reason, unfortunately the current frequentist approach to measurement invariance may produce misleading tests of measurement invariance.

Another issue is that the frequentist approach to measurement invariance test can be greatly inefficient when there are many groups to be tested. That is, with a large number of groups (e.g., 10 groups), it may not be feasible to find an invariant reference indicator across groups. Further, under measurement noninvariance, identifying noninvariant items across multiple groups may be very cumbersome particularly when for some groups a measurement parameter (e.g., factor loading) may be invariant, but for some other groups the measurement parameter may be noninvariant (Muthén & Asparouhov, 2013). Similarly, detecting noninvariant items can be inefficient when a measurement instrument contains more than two latent variables with a large number of items per each latent construct. In this sense, the current frequentist approach may provide inefficient tests of measurement invariance particularly with large number of groups, latent constructs, and/or items.

Lastly, measurement invariance tests with the frequentist approach may be too strict for measurement invariance testing. With typical MGCFA with the frequentist approach (e.g., maximum likelihood), all measurement model parameters are tested for null hypotheses of exact equality in terms of model parameter estimates, which are unlikely to hold in reality. From a practical point of view, small differences between model parameter estimates could be equally compatible with theory or the researchers' hypotheses and thus could ensure sufficient support to make valid comparisons on latent construct across populations. Unfortunately, the frequentist approach to measurement invariance test does not allow for even small differences in model parameters across groups.

Bayesian Approach to Tests of Measurement Invariance

Given these issues of measurement invariance tests with the frequentist approach, a Bayesian approach has emerged as a more flexible alternative to measurement invariance. A Bayesian approach to measurement invariance has recently been proposed by Muthén and Asparouhov (2013) and hence the Bayesian analysis of testing measurement parameters across groups described here is based on the work by Muthén and Asparouhov (2013). One of the features in the Bayesian approach to measurement invariance is that researchers can incorporate hypotheses or substantive theory into tests of measurement invariance. That is, if researchers have strong knowledge about differences in parameters before conducting a measurement invariance test, the information can be incorporated into tests of measurement invariance. If small differences between model parameters could be tolerable for group comparison in latent construct, for example, then researchers can allow for small differences in the model parameters across groups and the small differences between model parameters could provide sufficient evidence for valid comparison on latent construct across populations. Muthén and Asparouhov introduced this concept of *approximate measurement invariance* using Bayesian SEM that relaxes the constraint that differences in parameters be exactly zero, allowing these parameters to be estimated slightly differently, but approximately the same. This approximate measurement invariance can be accomplished by assigning the prior distribution with mean of zero and small variance to the difference in measurement model parameters. The authors explained that from the Bayesian approach, the frequentist approach to measurement invariance can be seen as test of stringent invariance that differences of all the measurement parameters can be considered to have a very strong prior distribution with mean of zero and variance of zero.

Approximate measurement invariance, proposed by Muthén and Asparouhov (2013), takes a two-step approach to detect measurement noninvariance. First, all measurement model parameters are estimated with Bayesian estimation in each group simultaneously. Under Muthén's and Asparouhov's approach, there is no need to choose a reference indicator to be constrained to be equal across groups in the Bayesian approximate measurement invariance test. By assigning the strong informative prior distributions to differences in parameter estimates, the model has enough information for identification. The second step, then, involves testing differences in terms of model parameters. Given that Bayesian inference yields posterior distributions of the model parameters, measurement invariance tests involve testing hypotheses about differences between posterior means of individual parameter and corresponding parameters' averages across the groups. A z-test is used for testing the statistical significance of the differences. If a posterior mean of a parameter in a group significantly deviates from its average across groups, it is considered to reflect noninvariance. Unlike the frequentist approach, approximate measurement invariance tests are not performed in hierarchical order and instead all levels of measurement invariance are tested simultaneously.

Muthén and Asparouhov (2013) noted that the estimated parameters through approximate measurement invariance test are biased estimates due to the alignment issue. That is, an alignment issue occurs because estimation through approximate measurement invariance tends to pull all of the parameters toward their averages across groups. Thus, it results in biased measurement model parameters (e.g., factor loadings and intercepts) and structural parameters. To resolve this problem, they recommended freeing the determined noninvariant items to get correct estimation of parameters.

Given that the exact equality of parameters across population is unlikely to hold in practice, a Bayesian approach to measurement invariance provides a flexible and practical approach to testing hypotheses about parameter differences. Recently, two studies compared the Bayesian and frequentist approaches to measurement invariance and found that the Bayesian approach is more likely to conclude in favor of measurement invariance than the frequentist approach through real data analyses (Cieciuch, Davidov, Schmidt, Algesheimer, & Schwartz, 2014; van de Schoot, Kluytmans, Tummers, Lugting, Hox, & Muthén, 2013). For example, van de Schoot and colleagues (2013) presented an example of a real data analysis of measurement invariance that the frequentist approach to measurement invariance yielded poor-fitting model, but when they applied Bayesian approximate measurement invariance, data-model fit turned out to be good while providing evidence of measurement invariance. These authors explained that researchers who employ approximate measurement invariance make tradeoff between the degree of measurement invariance and the degree of model fit. Further, they conducted a simulation study and found that approximate measurement invariance performed better than the frequentist approach to measurement invariance in terms of detecting true difference in mean of latent construct under the presence of partial invariance.

In addition, Muthén and Asparouhov (2013) have pointed out that approximate measurement invariance can be a very useful approach because all factor loadings related to the latent construct can be tested. Unlike a frequentist approach to tests of measurement invariance, a Bayesian approach does not require choosing a reference indicator for model identification when assessing measurement invariance. Because no reference indicator is chosen and constrained to be equal across groups for model identification, all factor loadings can be freely estimated and tested for approximate measurement invariance. Given that the reference indicator will not be tested for measurement invariance when estimation methods from the frequentist approach are used, this may be viewed as a distinct advantage over the frequentist approach to measurement invariance test. Further, the Bayesian approach to measurement invariance test becomes greatly efficient when there are many groups to be tested and also identifying noninvariant items across multiple groups easily can be done within approximate measurement invariance test. With these advantages, a Bayesian approach to measurement invariance can be expected to play an increasing role in the future of measurement invariance testing.

Although the Bayesian approach to test measurement invariance proposed by Muthén and Asparouhov (2013) provides potential benefits, particularly compared to the frequentist approach, there are still some issues to be considered. It should be first noted that although a reference indicator does not need to be chosen and constrained to be equal across groups in measurement invariance testing, it is needed when structural parameters are estimated and compared across groups for metric equivalence of structural parameters as well as model identification. In addition, like general Bayesian analysis, the choice of prior distribution is very important role in measurement invariance testing (Muthén & Asparouhov, 2013; Steinmetz, 2013). van de Schoot and colleagues (2013) demonstrated that the posterior mean estimates of model parameters as well as model fit were affected by different specification of prior distributions to difference in parameters. These results imply that one can lead to different conclusions about measurement invariance depending on how prior distributions are specified. Therefore, prior distributions on difference between parameters should be carefully chosen based on substantive knowledge about the parameters when the Bayesian approach is taken to test measurement invariance.

It is worthwhile to note two issues. First, prior distributions on the *difference* between parameters suggested by Muthén and Asparouhov (2013) have some limitations when understanding approximate measurement invariance. In unstandardized solutions, variance of differences between parameters can represent different magnitudes of variability of differences depending on scales of factors' indicators. In other words, although Muthén and Asparouhov recommended zero-mean, small variance (i.e., variance of .01) prior distributions on differences between parameters, variances of .01 may have very small or large magnitude of variability depending on scales of factors' indicators. Given that the choice of the variance for prior distributions can affect the significance of measurement parameter differences and hence power of noninvariance detection (Muthén & Asparouhov, 2013), simply choosing a variance of .01 could theoretically lead to incorrect results in approximate measurement invariance testing and thus researchers should choose the variance of prior distribution with caution based on the scales of factors' indicators. Further, in their examples, Muthén and Asparouhov (2013) assigned strong informative prior distributions (i.e., normal distribution with mean of zero and variance of .01) to differences between parameters for all factor loadings and intercepts. It should be noted, however, that this approach might not be appropriate unless researchers have substantive theories on the differences between parameters. Particularly, when researchers do not have such detailed knowledge about parameter differences, they should not assign strong prior distributions to differences between parameters, but instead use noninformative prior distributions.

Group Comparisons Involving Latent Construct Parameters

Issue of Partial Invariance in Group Comparison

Under partial invariance, noninvariant items are allowed to vary across groups in the model when assessing group difference in latent constructs. If partial measurement invariance holds, however, it is debatable whether comparisons of latent constructs across populations are valid in the previous literature. Some researchers have argued that measurement invariance, particularly factor loading invariance, is a prerequisite for meaningful group comparison on latent constructs (e.g., Cheung & Rensvold, 1999; Meredith, 1993; Meredith & Teresi, 2006). It has been believed that different magnitudes of factor loadings across populations mean different magnitudes of association of observed variables with latent constructs, indicating possible different meaning of latent constructs in different populations. From this perspective, partial invariance could be problematic because it fails to ensure that the construct measured by a set of items is not equivalent or comparable.

However, many more researchers differentiate construct equivalence from measurement invariance (Byrne et al., 1989; Millsap & Kwok, 2004; Steenkamp & Baumgartner, 1998; van de Vijver & Tanzer, 2004; Wu et al., 2007). These researchers believe that although latent constructs are equivalent across populations, it could be possible to have different relations between observed variables and their associated latent construct. When researchers have strong beliefs of construct equivalence across populations based on the theories or previous empirical studies, construct equivalence can be assumed even under measurement noninvariance. Therefore, the concept of construct equivalence should not be tested simply through measurement invariance tests, but be justified based on substantive theories, researcher's strong beliefs, and/or previous empirical studies. In fact, previous literature has found that measurement noninvariance can occur because of translation errors, cultural bias, or unequal familiarity with the item content or format (Taylor & Lee, 2012; van de Vijver & Tanzer, 2004) although the underlying latent construct is assumed to be equivalent across populations. Another good example is items showing differential item functioning (DIF) in the field of educational measurement. Suppose, for example, that a set of test items are developed to measure students' general math ability and a researcher wants to test whether the test items function differently between genders. Although the construct (i.e., general math ability) is equivalent between girls and boys, items function differently due to different degrees of familiarity with the test item format or context, which results in DIF (Bolt, 2000). In DIF assessment, the presence of DIF items does not imply that the underlying construct measured by the test items is different across genders. Instead it is assumed that they simply function differently due to item impact (i.e., presence of DIF is due to true different ability) or item bias (i.e., presence of DIF is due to some technical issue unintentionally favoring a certain group) (Reise, Smith, & Furr, 2001). Thus, although item parameters may differ between populations, it is still assumed that the underlying construct is equivalent across populations. In addition to conceptual equivalence of latent constructs, Wu and colleagues suggested *metric equivalence* of the latent construct for validity of group comparison of latent constructs. That is, the valid group comparisons of latent construct necessitate that the same latent constructs are measured on the same metric across groups. Under the conceptual equivalence of latent constructs, this can be achieved by linking the metric of the latent variable to its reference indicator which is truly invariant in MGCFA framework.

In summary, although noninvariance of factor loadings across populations may reflect that latent constructs measured by a set of items are different under some situations, it is not always true. As described earlier, although equivalent constructs are measured, there are more situations where the same underlying construct may function differently in different populations. In addition, given that measurement invariance can be difficult to achieve, small parameter differences may be tolerable for making inference about measurement invariance. From this perspective, partial invariance might not be problematic in terms of construct validity under the assumption of conceptual equivalence and metric equivalence of latent constructs. Instead, one of the important issues to be addressed is the extent to which the measurement noninvariance influences statistical conclusions and inferences about group differences in underlying constructs.

Group Comparison of Observed Composite Scores

The common procedure to test for a latent construct mean difference is to create observed composite scores and then test mean differences using a traditional *t*-test or analysis of variance (ANOVA), in large part because composites are easy to create and easily understood by applied researchers (Allen, 1999; Borsboom, 2006). Given that latent constructs are measured using measurement instruments with multiple items, group comparisons regarding latent construct using observed composite scores often raise concern about whether differences in observed composite scores reflect true differences in the latent construct. Regarding this issue, there is agreement that valid comparisons of observed composite scores across groups greatly rests on the extent to which measurement invariance is achieved (Little, 1997; Millsap & Kwok, 2004; Steinmetz, 2013; Vandenberg & Lance, 2000; Wu et al., 2007) unless the effect of noninvariance is inferentially benign. That is, under the situation where the effect of noninvariance is benign, conclusions on group difference in latent constructs can remain valid and meaningful; otherwise, validity of group comparison depends on the degree of measurement invariance (Borsboom, 2006).

Recently, Steinmetz (2013) conducted a simulation study to investigate the impact of partial invariance on accuracy of statistical conclusion of latent mean differences across groups when observed composite score is used. In simulation, two-group one-factor models with four or six indicators were used for data generation. Steinmetz manipulated varying degrees of partial invariance with varying factor loading and intercept differences across groups as well as latent mean differences and sample size. Using individual composite scores that were created from several indicators, regression was conducted to investigate latent mean differences across groups. The results showed that the presence of one or two noninvariant factor loadings and intercepts in the model increased Type I error. Also, it was observed that power of detecting true mean difference in the latent construct decreased. Further results showed that the effect of unequal intercepts substantially affected the observed mean difference while the effect of unequal factor loadings was relatively small. As the number of unequal intercepts increased or sample size decreased, the percentage of correct conclusions on true difference in latent construct decreased. Given these results, the author concluded that use of the observed composite scores from the model with partial invariance across populations can lead to inaccurate conclusions on group differences in the latent construct.

Millsap and Kwok (2004) conducted a simulation study to investigate the impact of using observed sum scores from several partial invariant measurement models on the selection accuracy of persons. Through Monte Carlo simulation study, they compared the selection accuracy using the observed sum scores with selection accuracy with the known factor scores in terms of four different indices of selection accuracy. These include (1) proportion of persons selected per group, (2) the success ratio (the proportion of persons selected using their true scores among those selected using the observed sum scores), (3) sensitivity (the proportion of persons selected using their observed sum scores among those selected using the true scores) and (4) specificity (the proportion of persons not selected using their observed sum scores among those not selected using their true scores). In their simulation, lack of invariance always had smaller parameter values of factor loadings and intercepts in the focal group. The results showed that lack of both weak and strong invariance resulted in a lower sum of observed scores in the focal group and hence the proportion of people selected based on the observed sum scores became lower compared to the proportion of people selected based on the factor scores in the focal group, and in turn, this increased the selection proportion in the reference group. The results of these previous studies imply that group comparisons on the latent variable using observed scores may not be the true representation of difference in latent constructs unless measurement invariance holds.

These studies demonstrated that if measurement invariance does not hold, group comparison of observed scores may yield misleading conclusions on group difference in the latent construct. As stated earlier, observed score means are a function of factor loadings, intercepts, and latent construct means, and hence differences in observed score means can unambiguously be true differences in the latent construct means only when the factor loadings and intercepts are invariant across populations. When factor loadings for observed variables are equivalent, but intercepts differ across groups, the difference in means on the observed variables is due to not only the difference in means of latent construct, but also the difference in intercepts for the observed variables (Thompson & Green, 2013). Therefore, group comparisons of observed scores to make inference in latent construct differences can be drawn only when measurement invariance holds. In other words, the presence of measurement invariance is a critical assumption for valid comparisons of observed scores across groups.

Group Comparison Involving Latent Construct Parameters

Given that many studies are presenting noninvariant measurement models across populations (Schmitt & Kuljanin, 2008), group comparisons of observed scores are more likely to lead to incorrect conclusions on group differences in latent constructs. As such, some methodological researchers have suggested analyzing group differences in the latent constructs within the MGCFA model rather than using traditional methods (e.g., ANOVA) to test group differences with the observed scores when studying group differences in the latent constructs (Steinmetz, 2013; Thompson & Green, 2013). Under the assumption of conceptual equivalence of latent constructs and metric equivalence of latent constructs, researchers may assess group difference in latent construct mean within the MGCFA while allowing only a set of parameters to vary across populations and constraining another subset of parameters to be invariant under the partial invariance condition. For example, some researchers have argued that comparison of latent construct means across populations can be meaningful and valid when a few items are not invariant, and that under these conditions, failure to achieve measurement invariance does not affect a significant portion of the results (e.g., Byrne et al., 1989). Other researchers have suggested that group comparisons of latent construct are meaningful as long as two of the factor loadings (including reference indicator) are invariant across populations (e.g., Steenkamp & Baumgartner, 1998). From a statistical viewpoint, the number of noninvariant item may not be important. Under noninvariance situations, researchers may directly perform group comparison involving latent constructs within the MGCFA framework while not imposing any constraints in the first place. Note that from the frequentist approach, freely estimating all of the factor loadings is not feasible because at least one indicator variable must be fixed (e.g., to 1) in each group for model identification.

Although a reference indicator also should be chosen for model identification in the Bayesian MGCFA framework, a Bayesian approach provides more flexibility than a frequentist approach in that researchers can allow for small differences in the reference indicators' parameters across groups. For the Bayesian approach, by assigning a strong prior distribution with small variance to differences between reference indicators' parameters across groups, researchers can have the measurement model to be identified (Muthén & Asparouhov, 2013). It should be noted that the imposed constraint on the reference indicators is necessary for metric equivalence as well as model identification in the MGCFA framework. When the parameters are estimated separately for each group, the estimate of the parameters of the same item might be on different metrics (Zumbo, 2007), making group comparison of latent constructs invalid. Therefore, different item parameter estimate metrics need to be linked in order to place them on a common metric (i.e., metric equivalence). In order to ensure metric equivalence, a constraint is imposed so that reference indicators are the same across groups. When the strong prior distributions were assigned to the reference indicators' parameters in the Bayesian MGCFA framework, it should be noted that the exact metric equivalence may not be achieved, but *approximate metric equivalence* may be achieved. If small differences in the metric of model parameter estimates could be tolerated for group comparisons in latent constructs, this could be sufficient evidence for making valid comparisons on latent constructs across populations. Given that it is not clear what degrees of approximate metric equivalence may be appropriate for comparison of latent constructs, it is recommended to perform a sensitivity analysis of parameter estimates to varying degrees of variance of prior distributions for the differences between reference indicators' parameters across populations.

However, previous studies have demonstrated that measurement noninvariance could yield incorrect conclusions of group differences in latent construct parameters within the MGCFA framework (Beuckelaer & Swinnen, 2011; Kaplan & George, 1995). Within the context of a population analysis, for example, Kaplan and George (1995) examined the power of detecting a latent construct mean difference between two groups under partial factor loading invariance and factor loading noninvariance. Kaplan and George (1995) used a six-item, two-factor model and a 12-item, two-factor model. They also varied factor loading sizes and sample sizes (i.e., equal vs. unequal). In their study, latent construct means were estimated for both groups under the assumption of strict invariance (i.e., the factor loadings, the intercepts, and the error terms are equal across populations), which was not true. They found that increased levels of noninvariance decreased power of detecting latent mean difference although magnitude of latent mean difference had more substantial effect on power than other factors. Beuckelaer and Swinnen (2011) expanded on the work of Kaplan and George by considering both factor loading partial invariance and intercept partial invariance through a simulation study. They investigated the impact of weak or strong partial invariance on statistical conclusions regarding the latent variable mean difference between two groups. The models they investigated were a 3-item, one-factor model and a 4-item, one-factor model. In their simulation, only one indicator variable was noninvariant in terms of factor loadings or both factor loadings and intercepts. Similar to Kaplan and George's study, latent construct means were estimated for both groups under the assumption of strict measurement invariance between two groups. They found that when there is no true difference in the latent construct mean, the Type I error increased to 45% for the 3-item, one factor model, and 37.2% for 4-item, one factor model. When there was true difference in the latent construct mean, the power of detecting latent mean difference greatly varied ranging from 28.7% to 95.5%. In addition, this study revealed that the number of indicators (i.e., 3 indicators vs. 4 indicators) did not have an impact on the percentage of correct conclusions on the latent variable mean difference.

Those findings from Kaplan and George (1995) and Beuckelaer and Swinnen (2011) imply the negative impact of partial invariance on structural parameter estimates (i.e., latent construct mean) and suggested that the validity of group comparisons of the latent construct mean is dependent on the extent to which measurement invariance is met. They do, however, have an important limitation. Specifically, although these two studies explicitly or implicitly mentioned the negative impact of model misspecification on structural parameter estimates, in these two studies latent variable means were estimated for both groups under the incorrect assumption of measurement invariance. For instance, in both Kaplan and George's study and that of Beuckelaer and Swinnen, latent variable means were estimated for both groups under the assumption of strict invariance (i.e., the factor loadings, the intercepts, and the error terms are equal across populations) which was not true. That is, equality constraints were imposed that were not true in the population, making their models fundamentally misspecified. Given this, results of previous studies regarding the impact of partial invariance on the correct conclusion for the difference in latent constructs may be confounded with model misspecification. For example, if measurement parameters are constrained to be equal across groups even though they are actually noninvariant, latent construct parameter estimates can be biased and hence comparison of group means on latent construct estimates could yield incorrect conclusions.

Reviews of Simulation Designs

As discussed above, most previous simulation studies related to measurement invariance have been conducted through MGCFA from a frequentist approach. The measurement invariance literature primarily focuses on detecting noninvariant items, while a relatively small number of studies have examined the impact of partial invariance and noninvariance on subsequent comparisons of structural parameters across groups within the MGCGA framework (Beuckelaer & Swinnen, 2011; Kaplan & George, 1995). Although each of previous simulation studies on measurement invariance included conditions that were expected to be found in real data analysis contexts and thus may allow for researchers to generalize to a wider range of conditions, there are some limitations in their simulation design.

First, research on the impact of noninvariance has been conducted by varying the magnitude of the parameter differences and/or by varying the number of invariant items (Beuckelaer & Swinnen, 2011; Finch & French, 2012; Kaplan & George, 1995; Kim & Yoon, 2011; Kim, Yoon, & Lee, 2012; Meade & Bauer, 2007; Millsap & Kwok, 2004). These previous simulation studies simulated noninvariant items in terms of the measurement model parameters by subtracting or adding some values from the reference group parameters for the focal groups. For instance, Kim and Yoon created noninvariant factor loadings in the focal group by subtracting .2 and .4 from reference group's factor loadings for small and large amounts of DIF, respectively. The same study simulated intercept invariant items by adding values of .3 and .6 to the intercept of the reference group for small and large DIF, respectively. Another study by Finch and French created noninvariant factor loading in focal group by subtracting values from .1 through .4 increments of .1 from reference group factor loading. Similarly, Millsap and Kwok created noninvariant factor loading in by subtracting values from .1 through .3 increments of .1 from reference group factor loading. Although some researchers argued that the differences of .2 and .4 in factor loadings represents moderate and large differences, it is not clear whether these differences represent a meaningful factor loading differences in other studies because these differences are not on a standardized scale in most of measurement invariance simulation studies (Beuckelaer & Swinnen, 2011; Kaplan & George, 1995; Kim et al., 2012; Meade & Bauer, 2007). Factor loading differences such as .2 or .4 may have different magnitudes of noninvariance depending on scales of factors' indicators. Although this is a popular way to create noninvariant items in literature, it should be noted that a fixed parameter difference might represent different magnitudes of noninvariance depending on scales of factors' indicators.

Second, previous literature has demonstrated that measurement quality (e.g., factor loading magnitude or communality) plays a critical role in measurement invariance testing (Meade & Bauer, 2007) as well as in parameter estimation (Gagné & Hancock, 2006). Gagné and Hancock found that factor loading magnitude played a significant role in convergence and accuracy of parameter estimates. Specifically, their results showed as the factor loading size increases, the convergence and accuracy of parameter estimates improve, precision and power of measurement invariance testing increases, and accuracy of selection improves. Further, high factor loading magnitude is a significant factor in increasing precision and power of measurement invariance testing (Meade & Bauer, 2007) and further impact of partial invariance on selection of persons based on observed measures (Millsap & Kwok, 2004). This finding also implies that factor loading magnitude can be an important factor to be considered when studying measurement invariance testing.

Summary of the Current Study

Understanding group differences in latent constructs is the basis for some of the most commonly investigated research questions addressed by social and educational researchers. Historically, the group comparisons regarding latent constructs have been conducted with observed composite scores through traditional group comparison procedures, but this common procedure provides valid conclusions on latent construct differences only when measurement invariance indeed holds. Since measurement invariance is difficult to achieve in reality, it is more likely to misinform conclusions on group differences in latent constructs. Given this, what would seem ideal is to directly examine group difference in a latent construct within the MGCFA model and to make inferences regarding group differences. Whereas the measurement invariance literature has emphasized that measurement invariance is prerequisite for valid comparisons of latent construct scores across groups, it is in the case only when observed scores are used for group comparison in latent constructs. Since we expect some noninvariance in places researchers might not be able to detect, researchers may build it into the model and examine group difference in latent constructs.

The purpose of this study is, thus, to empirically investigate the extent to which measurement noninvariance affects structural parameter comparisons across populations from a Bayesian approach. Particularly, this study aims to investigate the sensitivity of the posterior distribution of two of structural parameters, structural regression coefficient differences and factor mean differences, to varying degrees of noninvariant measurement models across populations when noninvariance exists in model. This study uses three types of evaluation criteria along with four simulation design factors (i.e., sample size, factor loading size, structural parameter difference, and prior distribution). In this study, the degrees of noninvariance will be manipulated with two factors (i.e., percentage of noninvariant item and total magnitude of noninvariance). The inference on structural parameters is “sensitive” if varying degrees of noninvariance between populations causes a significant change of (1) accuracy of statistical conclusion on structural parameters (Type I error and power), (2) precision of structural parameter estimates, and (3) bias of structural parameter estimates. In particular, the research questions under study are:

(1) how will varying degrees of factor loading noninvariance influence Type I error and power, precision of structural parameter estimates, and bias of structural parameter difference estimates?

(2) how will varying degrees of intercept noninvariance influence the three outcomes?

(3) how will varying degrees of both factor loading and intercept noninvariance influence the three outcomes?

(4) how will total sample size, factor loading size, and the prior distribution in three types of noninvariance conditions influence the three outcomes?

The following section describes the methods used to investigate the above research questions including the design of the current study: data generating model, simulation design factors, and data analysis.

Chapter 3: Methods

Data Generating Model

Two-group two-factor models with mean structure were used to generate data from a multivariate normal distribution. Selection of the data generating model in this study was based on previous measurement invariance testing application studies (Anderson, Hughes, Fisher, & Nicklas, 2005; Byrne, Shavelson, & Muthén, 1989; Chen & Tang, 2006; Cheung & Watkins, 2000; Crockett, Randall, Shen, & Driscoll, 2005; Dolan, Colom, Abad, Wicherts, Hessen, & Sluis, 2006; Marsh, 1993; Yoo & Donthu, 2001) as well as simulation studies (Beuckelaer & Swinnen, 2011; Finch & French, 2012). Assuming that two populations have the same factor structure (i.e., configural invariance) across all conditions, the data generating model for only one population (i.e., reference population) is presented in Figure 1. Varying degrees of factor loading and intercept difference were manipulated while the other parameters (e.g., factors' variances, error variances) were held constant across all generating conditions. Note that the differences in factor loadings and intercepts between two populations only occurred on the exogenous factor while all model parameters on the endogenous factor were invariant across all generating conditions. The critical parameters of interest in this study were (1) group difference in the regression coefficients from exogenous to endogenous factors between two populations, which is $\gamma^R - \gamma^F$, where the superscripts '*R*' and '*F*' represent reference population and focal population, respectively, and (2) factor mean differences in the exogenous and endogenous factors between two populations, which are captured by κ^F and α^F respectively, where κ^F and α^F are the latent intercepts in the focal population. It should be noted that the exogenous and endogenous factors' means were set to zero in the reference population for model identification, and hence κ^F represents the relative difference in the exogenous factor mean between two populations and α^F represents the relative difference in the endogenous factor mean after controlling for differences in the exogenous factor between two populations.

Given that factor variances are not likely to be same across populations in reality, factor variances were simulated to be different across populations. The variances of both the exogenous and the endogenous factors in the reference populations were set to 1 while the corresponding variances in the focal group were set to 1.3 across all conditions. Across all conditions, for the reference population the regression coefficient between exogenous and endogenous factors was set to .5 and the disturbance variance of endogenous factor was set to .75, resulting in variance of the endogenous factor being 1 as previously stated. Because the regression coefficient between exogenous and endogenous factors varied in the focal group, the values of disturbance variance of the endogenous factor in the focal group were chosen to achieve variance of the endogenous factor of 1.3 across all conditions. In addition, error variances were generated as invariant between the two populations across all conditions. The error variances for both the reference and focal populations were assumed to be homogeneous and were set to .32, resulting in the construct reliability coefficient *H* (Hancock & Mueller, 2001) approximately ranging from .70 and .90 across all conditions. The parameter values for factors variances and error variances were selected with reference to previous similar simulation studies on measurement invariance test (Beuckelaer & Swinnen, 2011; Kaplan

& George, 1995; Kim et al., 2012). Population generating values for the factor loadings and intercepts in the reference group and focal group are presented in Tables 1 through 3.

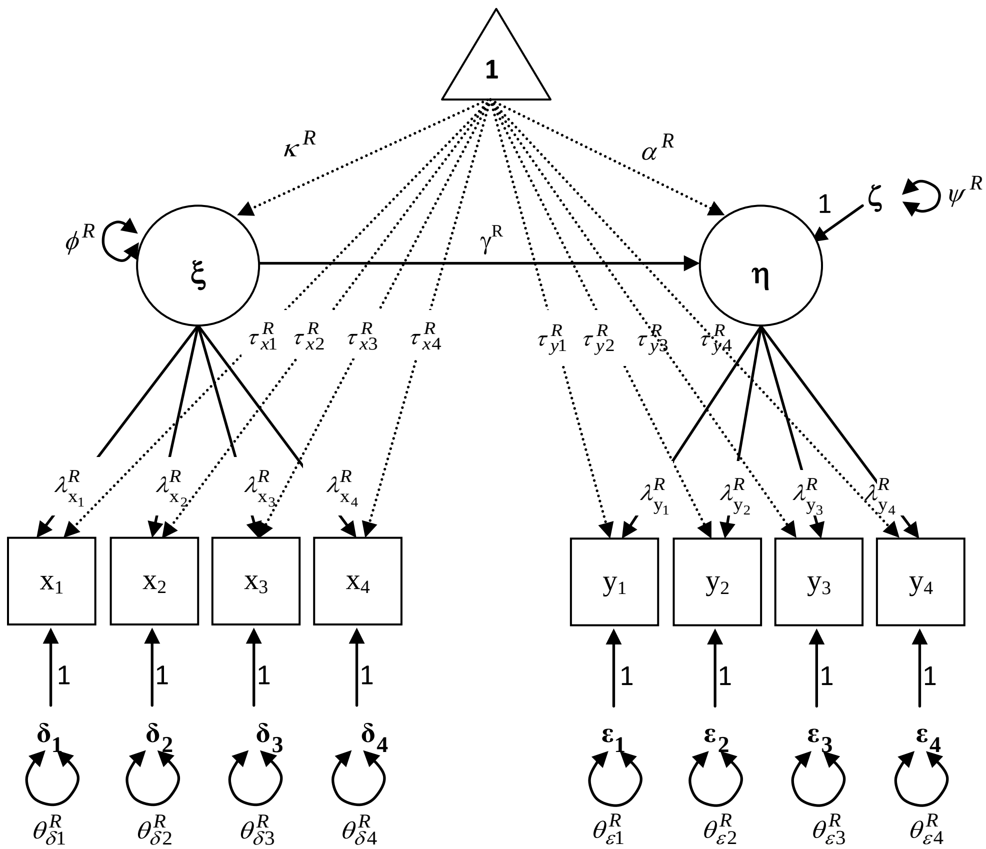

Figure 1. Data Generating Model

Table 1

Population Generating Values for Factor Loading Noninvariance Only Conditions

High factor loading size		
Reference Group	$\boldsymbol{\Lambda}' = \begin{bmatrix} .9 & .9 & .9 & .9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$, $\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$, $\boldsymbol{\theta}' = [.32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32]$	
Focal Group	$\boldsymbol{\Lambda}' = \begin{bmatrix} — & — & — & — & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$, $\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$, $\boldsymbol{\theta}' = [.32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32]$	
Total magnitude of noninvariance	Percentage of noninvariant items	
	25%	75%
20%	$\boldsymbol{\Lambda}' = \begin{bmatrix} .9 & .9 & .9 & .72 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$	$\boldsymbol{\Lambda}' = \begin{bmatrix} .9 & .84 & .84 & .84 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$
50%	$\boldsymbol{\Lambda}' = \begin{bmatrix} .9 & .9 & .9 & .45 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$	$\boldsymbol{\Lambda}' = \begin{bmatrix} .9 & .75 & .75 & .75 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$
80%	$\boldsymbol{\Lambda}' = \begin{bmatrix} .9 & .9 & .9 & .18 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$	$\boldsymbol{\Lambda}' = \begin{bmatrix} .9 & .66 & .66 & .66 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$
Moderate factor loading size		
Reference Group	$\boldsymbol{\Lambda}' = \begin{bmatrix} .5 & .5 & .5 & .5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$, $\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$, $\boldsymbol{\theta}' = [.32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32]$	
Focal Group	$\boldsymbol{\Lambda}' = \begin{bmatrix} — & — & — & — & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$, $\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$, $\boldsymbol{\theta}' = [.32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32]$	
Total magnitude of noninvariance	Percentage of noninvariant items	
	25%	75%
20%	$\boldsymbol{\Lambda}' = \begin{bmatrix} .5 & .5 & .5 & .40 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$	$\boldsymbol{\Lambda}' = \begin{bmatrix} .5 & .47 & .47 & .47 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$
50%	$\boldsymbol{\Lambda}' = \begin{bmatrix} .5 & .5 & .5 & .25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$	$\boldsymbol{\Lambda}' = \begin{bmatrix} .5 & .42 & .42 & .42 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$
80%	$\boldsymbol{\Lambda}' = \begin{bmatrix} .5 & .5 & .5 & .10 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$	$\boldsymbol{\Lambda}' = \begin{bmatrix} .5 & .37 & .37 & .37 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$

Table 2

Population Generating Values for Intercept Noninvariance Only Conditions

High factor loading size		
Reference Group	$\boldsymbol{\Lambda}' = \begin{bmatrix} .9 & .9 & .9 & .9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$, $\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$, $\boldsymbol{\theta}' = [.32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32]$	
Focal Group	$\boldsymbol{\Lambda}' = \begin{bmatrix} .9 & .9 & .9 & .9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$, $\boldsymbol{\tau}' = [\_\_ \;\; \_\_ \;\; \_\_ \;\; \_\_ \;\; .8 \;\; .8 \;\; .8 \;\; .8]$, $\boldsymbol{\theta}' = [.32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32]$	
Total magnitude of noninvariance	Percentage of noninvariant items	
	25%	75%
20%	$\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .64 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$	$\boldsymbol{\tau}' = [.8 \;\; .75 \;\; .75 \;\; .75 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$
50%	$\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .40 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$	$\boldsymbol{\tau}' = [.8 \;\; .67 \;\; .67 \;\; .67 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$
80%	$\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .16 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$	$\boldsymbol{\tau}' = [.8 \;\; .59 \;\; .59 \;\; .59 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$
Moderate factor loading size		
Reference Group	$\boldsymbol{\Lambda}' = \begin{bmatrix} .5 & .5 & .5 & .5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$, $\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$, $\boldsymbol{\theta}' = [.32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32]$	
Focal Group	$\boldsymbol{\Lambda}' = \begin{bmatrix} .5 & .5 & .5 & .5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$, $\boldsymbol{\tau}' = [\_\_ \;\; \_\_ \;\; \_\_ \;\; \_\_ \;\; .8 \;\; .8 \;\; .8 \;\; .8]$, $\boldsymbol{\theta}' = [.32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32]$	
Total magnitude of noninvariance	Percentage of noninvariant items	
	25%	75%
20%	$\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .64 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$	$\boldsymbol{\tau}' = [.8 \;\; .75 \;\; .75 \;\; .75 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$
50%	$\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .40 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$	$\boldsymbol{\tau}' = [.8 \;\; .67 \;\; .67 \;\; .67 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$
80%	$\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .16 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$	$\boldsymbol{\tau}' = [.8 \;\; .59 \;\; .59 \;\; .59 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$

Table 3

Population Generating Values for Both Factor Loading and Intercept Noninvariance Conditions

High factor loading size		
Reference Group	$\boldsymbol{\Lambda}' = \begin{bmatrix} .9 & .9 & .9 & .9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$, $\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$, $\boldsymbol{\theta}' = [.32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32]$	
Focal Group	$\boldsymbol{\Lambda}' = \begin{bmatrix} — & — & — & — & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$, $\boldsymbol{\tau}' = [— \;\; — \;\; — \;\; — \;\; .8 \;\; .8 \;\; .8 \;\; .8]$, $\boldsymbol{\theta}' = [.32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32]$	
Total magnitude of noninvariance	Percentage of noninvariant items	
	25%	75%
20%	$\boldsymbol{\Lambda}' = \begin{bmatrix} .9 & .9 & .9 & .72 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$ $\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .64 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$	$\boldsymbol{\Lambda}' = \begin{bmatrix} .9 & .84 & .84 & .84 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$ $\boldsymbol{\tau}' = [.8 \;\; .75 \;\; .75 \;\; .75 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$
50%	$\boldsymbol{\Lambda}' = \begin{bmatrix} .9 & .9 & .9 & .45 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$ $\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .40 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$	$\boldsymbol{\Lambda}' = \begin{bmatrix} .9 & .75 & .75 & .75 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$ $\boldsymbol{\tau}' = [.8 \;\; .67 \;\; .67 \;\; .67 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$
80%	$\boldsymbol{\Lambda}' = \begin{bmatrix} .9 & .9 & .9 & .18 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$ $\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .16 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$	$\boldsymbol{\Lambda}' = \begin{bmatrix} .9 & .66 & .66 & .66 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .9 & .9 & .9 & .9 \end{bmatrix}$ $\boldsymbol{\tau}' = [.8 \;\; .59 \;\; .59 \;\; .59 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$
Moderate factor loading size		
Reference Group	$\boldsymbol{\Lambda}' = \begin{bmatrix} .5 & .5 & .5 & .5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$, $\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$, $\boldsymbol{\theta}' = [.32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32]$	
Focal Group	$\boldsymbol{\Lambda}' = \begin{bmatrix} — & — & — & — & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$, $\boldsymbol{\tau}' = [— \;\; — \;\; — \;\; — \;\; .8 \;\; .8 \;\; .8 \;\; .8]$, $\boldsymbol{\theta}' = [.32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32 \;\; .32]$	
Total magnitude of noninvariance	Percentage of noninvariant items	
	25%	75%
20%	$\boldsymbol{\Lambda}' = \begin{bmatrix} .5 & .5 & .5 & .40 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$ $\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .64 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$	$\boldsymbol{\Lambda}' = \begin{bmatrix} .5 & .47 & .47 & .47 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$ $\boldsymbol{\tau}' = [.8 \;\; .75 \;\; .75 \;\; .75 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$
50%	$\boldsymbol{\Lambda}' = \begin{bmatrix} .5 & .5 & .5 & .25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$ $\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .40 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$	$\boldsymbol{\Lambda}' = \begin{bmatrix} .5 & .42 & .42 & .42 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$ $\boldsymbol{\tau}' = [.8 \;\; .67 \;\; .67 \;\; .67 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$
80%	$\boldsymbol{\Lambda}' = \begin{bmatrix} .5 & .5 & .5 & .10 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$ $\boldsymbol{\tau}' = [.8 \;\; .8 \;\; .8 \;\; .16 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$	$\boldsymbol{\Lambda}' = \begin{bmatrix} .5 & .37 & .37 & .37 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 & .5 \end{bmatrix}$ $\boldsymbol{\tau}' = [.8 \;\; .59 \;\; .59 \;\; .59 \;\; .8 \;\; .8 \;\; .8 \;\; .8]$

Manipulated Factors

The simulation design used in this study was 3 (noninvariance conditions) × 2 (percentage of noninvariant items) × 3 (total magnitudes of noninvariance) × 2 (factor loading magnitudes) × 3 (total sample sizes) × 3 (structural parameter differences) × 3 (prior distribution) for 972 conditions for this study. In addition, measurement invariance conditions (i.e., 0% of noninvariant items) for 54 conditions were included to serve as a baseline conditions against noninvariance conditions. Previous simulation studies on SEM from a Bayesian approach used various numbers of replications, ranging from 100 (Lee, Song, & Cai, 2010) to 1,000 replications (Sass & Smith, 2006). In order to decide the replication size for this study, a pilot study was conducted with three different replication sizes (100, 500, and 1,000) under the various degrees of invariance conditions. Although there were small differences in precision between two replication sizes (500 and 1,000) in terms of outcomes of this study, 1,000 replications were set in simulation to ensure the stability of outcome measures of the study.

Type of noninvariance. Three different types of noninvariance were examined: (1) factor loading noninvariance only, (2) intercept noninvariance only, and (3) both factor loading and intercept noninvariance. Under the factor loading noninvariance only conditions, various degrees of factor loading difference were manipulated while intercept parameters were generated as invariant between two populations, which is referred to as weak measurement noninvariance in literature. Under the intercept noninvariance only conditions, varying degrees of intercept difference were manipulated while factor loading parameters were generated as invariant between two populations across all conditions. Under the both factor loading and intercept noninvariance conditions, varying degrees of both factor loading and intercept differences were manipulated. Both intercept noninvariance only conditions and both factor loading and intercept noninvariance conditions are often referred to as strong measurement noninvariance in literature. Note that error variances were generated as invariant between the two populations across all conditions.

Magnitude of noninvariance. Because it is not clear whether the impacts of degrees of noninvariance for individual items are confounded with overall magnitudes of noninvariance, the degrees of noninvariance for individual items varied in several ways while the overall magnitudes of noninvariance were controlled. Specifically, this study systematically varied the degrees of noninvariance using two design conditions: percentage of noninvariant items and total magnitude of noninvariance. Although previous studies directly manipulate magnitude of noninvariance for each item (e.g., .2 for small factor loading difference, .4 for large factor loading difference), this study indirectly manipulated magnitude of noninvariance for each item using the percentage of noninvariant items and the total magnitude of noninvariance.

Percentage of noninvariant items. Two levels of percentage of noninvariant items were manipulated (25% and 75%). The 25% and 75% conditions represent 25% of noninvariant items (1 out of 4 items) and 75% of noninvariant items (3 out of 4 items), respectively. For all simulation conditions, all model parameters that differ between two populations only occurred on the exogenous factor (ξ) while all model parameters on the endogenous factor (η) were the same between two populations. For example, in the 25% factor loading difference conditions, one factor loading was different on ξ while all factor loadings on η were the same between two populations. The fourth item (x_4) of ξ was generated as noninvariant for 25% conditions and the second, third, and fourth items (i.e., x_2, x_3, x_4) were generated as noninvariant for 75%. The first item of ξ (x_1) was generated as invariant across all conditions for scaling purposes.

Given the limitations surrounding the generation of noninvariant items in previous studies, this study used a different approach to create noninvariant items for the simulation. That is, the degree of noninvariance for a model parameter was defined in a relative manner which was calculated as the relative difference of the model parameter from the corresponding model parameter value of the reference group. Specifically, the degree of noninvariance was expressed as percentages that represent the relative magnitude in difference in model parameters of the focal group from the corresponding parameters of the reference group. For example, the 10% noninvariance for a model parameter represents that the noninvariant item in the focal group differs by 10% lower than the corresponding factor loading in the reference group. This can give more meaningful parameter difference even in the unstandardized solution. To do so, noninvariant item parameters were simulated by multiplying a multiplicative factor, say k, to the population parameter values of the reference group. For the focal group, population measurement parameters of the noninvariant factor indicator were specified to be always smaller than the corresponding factor loading in the reference group. To illustrate, one of the population values of a factor loading in the reference group was set to .8. The population factor loadings in the focal group that is 10% (i.e., k =.9) smaller than the corresponding loading of .8 would be .72 (.8*.9=.72) and thus the value of .72 was used as a population factor loading for the focal group representing 10% difference in factor loading. For the factor loading of .5, the value of .45 (.5*.9=.45) was used as the population factor loading for the focal group that represents 10% difference.

Total magnitude of noninvariance. As the degrees of noninvariance for each item parameter and percents of noninvariant item vary in simulation conditions, the overall magnitude of noninvariance also varies. Thus, this study also controlled the overall magnitude of noninvariance. In this study, the total magnitude of noninvariance was defined as the total amount of noninvariance calculated by summing individual percentages of the parameters that showed measurement noninvariance. The overall magnitude of noninvariance was manipulated at 3 levels: 20%, 50%, and 80%. Given that there is no previous study to manipulate the overall magnitude in measurement invariance testing within the MGCFA, these levels are selected considering the percentage of noninvariant items and magnitude of noninvariance in each item. Specifically, the total magnitude of 20% noninvariance condition may be considered as small magnitude of total noninvariance in that magnitude of noninvariance in each item has small factor loading difference which has been defined in the previous literature (i.e., less than factor loading difference of .2). Similarly, the total magnitude of 80% noninvariance may be considered as large magnitude of total noninvariance in that at least one noninvariant item has larger factor loading difference which has been defined in the previous literature (i.e., greater than factor loading difference of .4) in majority of conditions. The total magnitude of 20% noninvariance indicates that the total amount of noninvariance for the focal group model is smaller than that for the reference group by 20%, and so on. To illustrate, suppose that there are two noninvariant items and each factor loading differs by being 10% lower than the corresponding factor loading in the reference group. Then the total amount of noninvariance can be calculated by summing individual magnitude of noninvariance, which is 20% (i.e., 10%+10%=20%). Across all conditions, the magnitude of noninvariance for each item was assumed to be equal. To illustrate, when the total amount of invariance is 50% and there are 75% of noninvariant items (i.e., three items noninvariant), then each item contributes approximately 16.17% noninvariance each, leading to total amount of 50% noninvariance.

Factor loading magnitude. Factor loading magnitude for the reference group was manipulated with two levels: moderate and high. In selection of population values for moderate and high factor loading, this study refers to a previous simulation study (Finch & French, 2012; Kim & Yoon, 2011). Homogeneous factor loadings of .5 and .9 were used as moderate and high factor loading that represent moderate measurement quality and high measurement quality, respectively (Finch & French).

Total sample size. This study manipulated total sample size. Based on the previous simulation studies in measurement invariance (Beuckelaer & Swinnen, 2011; Kaplan & George, 1995; Kim & Yoon, 2011; French & Finch, 2011), the total sample size of 200, 800, and 2,000 was used in this study, which represents small, moderate, and large sample size. The total sample size of 200 has been used as a small size but seems to occur in applied measurement invariance research (Anderson, Hughes, Fisher, & Nicklas, 2005), while total sample size of 2,000 reflects a cross-national research situation where the sample size per country is relatively large (Steenkamp & Baumgartner, 1998). Across all conditions, equal sample size per group was used, reflecting a research situation in which two groups have similar sample sizes and these approximately equal sample sizes commonly occur in measurement invariance research in a MGCFA framework.

Structural parameter differences. The structural parameters of interest in this study were differences in regression coefficients from exogenous to endogenous factors between reference and focal groups (i.e., $\square^{R}$- $\square^{F}$) and latent intercepts of exogenous and endogenous factors in the focal group (i.e., κ^{F} and α^{F}). For the regression coefficient from exogenous to endogenous factors in the focal group, it varied with values of .2, .5, and .8 while that in the reference group was set to .5 across all conditions, resulting in the difference in the regression coefficients between two populations of -.3, 0, and .3. The intercepts of κ^{F} and α^{F} used in this study were set to -.57, 0, and .57 yielding moderate, zero, and moderate effect sizes (i.e., d =-.5, 0, .5). The standardized effect sizes for mean differences in exogenous and endogenous factors are calculated such that the factor mean in the focal group is divided by the square root of the variance of that factor (Hancock, 2001). The population values of factor mean difference are very similar to previous simulation studies (Beuckelaer & Swinnen, 2011; Kaplan & George, 1995) and application studies (Byrne et al., 1989) and thus are considered reasonable.

Prior distribution. Across all conditions, the first item of each factor (i.e., x_1, y_1) was selected as reference indicators for model identification. This study employs three different prior distributions that were assigned to *ratios* of reference indicators' factor loadings and intercepts between groups (i.e., $\lambda^{F}_{x_1}/\lambda^{R}_{x_1}$, $\lambda^{F}_{y_1}/\lambda^{R}_{y_1}$, $\tau^{F}_{x1}/\tau^{R}_{x1}$, and $\tau^{F}_{y1}/\tau^{R}_{y1}$,). Although Muthén and Asparouhov (2013) used prior distributions for *differences* between parameters through use of Bayesian measurement invariance, this specification has a limitation in that values of variance of prior distributions may represent different magnitudes of variability of noninvariance depending on scales of factors' indicators in unstandardized solution. Instead, specification of prior distributions for ratio of two factor loadings can provide meaningful variability of parameter difference even in the unstandardized solution. In this study, normal distributions with mean of 1 were used as prior distributions for ratios of reference indicators' factor loadings and intercepts between groups with varying degrees of variances across all conditions. Three levels of prior distribution's variance were manipulated such that noninvariance of the reference indicators' factor loadings and intercepts between two populations vary within 0%, 10%, and 20%. As seen in Figure 2(a), the variance of the first prior distribution was set at zero, which means that the prior distribution does not allow noninvariance of the reference indicators' parameters between two populations. A prior distribution with zero variance can be regarded as a traditional scaling method that constrains reference indicators' factor loadings and intercepts to be equal across groups. This prior distribution is hereafter referred to as *prior distribution with zero variation*. The second level of prior distribution is presented in Figure 2 (b) and was designed to represent situations in which prior distributions allow magnitudes of noninvariance of the reference indicators' factor loadings and intercepts approximately within 10% between two populations, 95% of the time. The second level of the prior distribution is hereafter referred to as *prior distribution with 10% variation*. As shown in Figure 2 (c), the third level of variance for prior distribution was designed to represent situations in which prior distributions allow magnitudes of noninvariance of the reference indicators' factor loadings and intercepts approximately within 20% between two populations, 95% of the time. The third level of the prior distribution is hereinafter referred to as *prior distribution with 20% variation*.

At the time of writing, the software program used in this study, Mplus, does not allow using prior distributions for the *ratio* between parameters and only has an option to use prior distributions for the *difference* between parameters. Given this, this study used an alternative specification, which is to fix reference indicators' factor loadings and intercepts to their population values in the reference group and let those in the focal group have means of the population values and varying small-variance prior distributions. As a result, magnitudes of noninvariance of the reference indicators' factor loadings and intercepts vary approximately within 0%, 10%, and 20% as stated previously. For example, for a measurement model with a factor loading of .9 and intercept of .8 for the reference indicators, $N(.9, .002)$ and $N(.8, .002)$ were used as prior distributions for the factor loadings and intercepts, respectively, in the focal group. With a normal prior distribution with mean of .9 and variance of .002, a reference indicator's factor loading in the focal group is allowed to be freely estimated approximately within 10% (i.e., factor loading of between .81 and .99) from factor loading value of .9, 95% of the time. Similarly, with a normal prior distribution with mean of .8 and variance of .002, a reference indicator's intercept in the focal group is allowed to be freely estimated approximately within 10% (i.e., intercept of between .71 and .88) from intercept value of .8, 95% of the time.

For the other factor loadings and intercepts, this study specifies a noninformative prior that reflected no prior knowledge across all conditions. A normal distribution with a mean of zero and variance of 10^{10} was used as a noninformative prior which is a default prior for factor loadings in Mplus. Across all conditions, noninformative priors based on inverse-gamma distributions (IG(0, -1)) were used as prior distributions for factor variances and error variances, which is also a default prior for those parameters in Mplus.

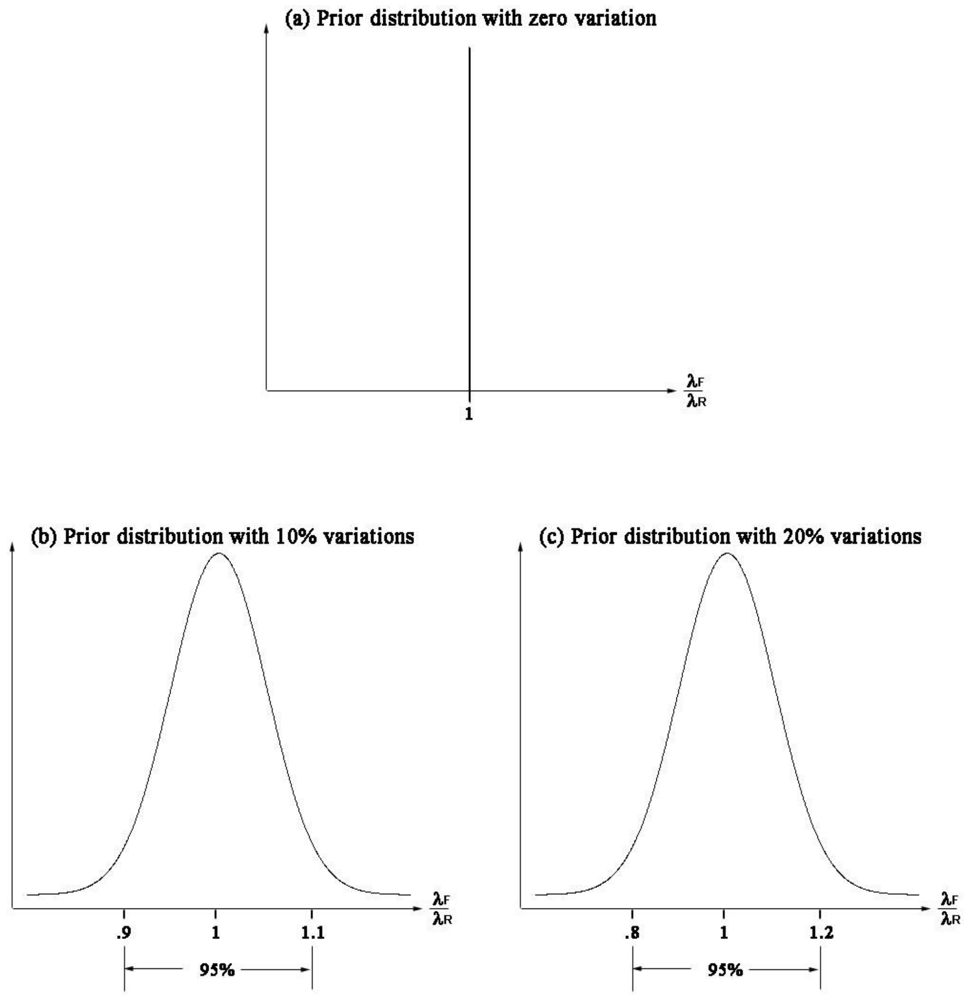

Figure 2. Three Types of Prior Distributions

Outcome Variables

To assess the sensitivity of noninvariance conditions, the three outcome variables from the varying noninvariance conditions were compared with those from the baseline conditions (i.e., invariance conditions). The three primary outcome variables are: (1) accuracy of statistical conclusion on structural parameter comparisons, (2) precision of the estimated structural parameter difference, and (3) bias in the posterior mean of structural parameter difference. The accuracy of statistical conclusions on the structural parameter comparisons was evaluated using the Type I error and power that is determined based on a 95% credibility interval of the posterior distributions of parameters of interest. If the interval from the 2.5th and 97.5th percentiles of the posterior distribution include zero, it is concluded such that there is no structural parameter difference across populations; meanwhile, if the interval from the 2.5th and 97.5th percentiles of the posterior distribution do not include zero, it is concluded such that there is structural parameter difference across populations. In addition, a 95% credibility interval was used to evaluate the precision of the structural parameter difference estimates by examining the width of 95% credibility interval of individual structural parameter difference estimates of interest. Finally, bias is defined as “a systematic difference between a sample estimate and the corresponding population value” (Bandalos & Leite, 2013, p. 642). For bias in the posterior mean of structural parameter difference, average relative bias (ARB) and average bias (AB) were calculated. The ARB of the parameter estimate is defined as “the average deviation of a sample estimate from its population value, relative to the population value” (Bandalos & Leite, 2013) and it has been frequently used to assess bias because it provides a common scale for researchers to compare magnitude of bias across different population parameter values. When the true population difference is zero, the ARB cannot be calculated. Thus, this study also calculated average bias (AB), which is defined as a simple average deviation of a sample estimate from its population value. If the absolute values of ARB are less than .15, they are considered not to be serious, and thus acceptable in most SEM analyses (Muthén, Kaplan, & Hollis, 1987). The ARB and AB are calculated as (Bandalos & Leite, 2013),

$$Average\ relative\ bias\left(\hat{\theta}_i\right)=\sum_{r=1}^{n_r}\left(\frac{\hat{\theta}_{ij}-\theta_i}{\theta_i}\right)/\,n_r\ ,$$

$$Average\ bias\left(\hat{\theta}_i\right)=\sum_{r=1}^{n_r}\left(\hat{\theta}_{ij}-\theta_i\right)/\,n_r$$

where $\hat{\theta}_{ij}$ is the sample estimate, θ_i is the population value, and n_r is the number of replications within the cell. Although there are substantial biases under some conditions, the measures of AB and ARB cannot capture them particularly in cases where both positive and negative biases exist, averaging to zero. In order to obtain a measure of amount of unsigned bias, average absolute relative bias (AARB) and average absolute bias (AAB) were also calculated by taking the absolute values of deviation of a sample estimate from its population value. The AAB and AARB were calculated as (Bandalos & Leite, 2013),

$$Average\ absolute\ bias\left(\hat{\theta}_i\right)=\sum_{r=1}^{n_r}|\,\hat{\theta}_{ij}-\theta_i\,|\,/\,n_r$$

$$Average\ absolute\ relative\ bias\left(\hat{\theta}_i\right)=\sum_{r=1}^{n_r}\left(\frac{|\,\hat{\theta}_{ij}-\theta_i\,|}{|\,\theta_i\,|}\right)/\,n_r$$

It should be noted that when the true population difference is zero, the AARB were not calculated.

Analysis

Mplus 7.11 was used to simulate and analyze sample data from a Bayesian approach. Given that the Gibbs sampler method is a popular method in Bayesian SEM (Fong & Ho, 2013; Golay, Reverte, Rossier, Favez, & Lecerf, 2012; Kaplan & Depaoli, 2012; Muthén & Asparouhov, 2012), the posterior distribution was estimated through the MCMC algorithm with the Gibbs sampler method for the analysis. As recommended by Muthén and Asparouhov (2012) and other application studies (Fong & Ho, 2013; Golay et al., 2012), the Gelman and Rubin convergence diagnostic, PSRF, was used to assess convergence. When PSRF values are between 1 and 1.1, convergence was considered to be achieved. Two MCMC chains were used in this study because use of two chains provides sufficient PSRF information compared to more chains and hence have been often used in previous studies (Muthén & Asparouhov, 2013). In order to determine the number of iterations, a pilot study was conducted. Under both large and small magnitudes of noninvariance conditions with moderate sample size, PSRF values reached to 1.1, on average, at 4,000 iterations and the PSRF values less than 1.1 were maintained until 80,000 iterations. Also, there were negligible differences in and parameter estimates among three iteration sizes (10,000, 50,000, and 80,000) across conditions. Therefore, in each chain, 10,000 iterations were used in the main simulation. The first half of the iteration per chain (i.e., 5,000 iterations) was burn-in iterations and thus was discarded. The remaining 5,000 iterations were used to calculate the posterior mean and 95% credibility intervals in this study. Model fit was assessed with PPP value recommended by previous studies (Lee & Song, 2004; Muthén & Asparouhov, 2012). A model with PPP value less than .05 was considered as a poor model fit; otherwise, it was considered as an adequate model fit. The properly converged replications with adequate model fit were used to evaluate the three outcome measures using SAS 9.2.

Chapter 4: Results

Chapter 4 provides the results of the current simulation study. Convergence and results of model fit assessment are presented in the first section. The remaining three sections present simulation results regarding the three outcome measures (i.e., accuracy of statistical conclusion on structural parameter comparisons, precision of structural parameter estimates, bias of structural parameter estimates) under three different types of noninvariance conditions. The results of each condition represent the average values of the outcome measures over properly converged replications with adequate model fit.

Convergence and Model Fit Assessment

In this study, a replication where the PSRF values ranged between 1 and 1.1 was considered as a properly converged replication. Convergence rates are calculated as the percentage of times a model properly converged over 1,000 replications for each condition. Table 4 presents the convergence rates obtained from the measurement invariance conditions by sample size, factor loading size, and prior distribution. As seen in Table 4, overall convergence rates across all conditions were found to be very good, yielding convergence rates greater than 99.0% in most conditions. All conditions where there were large sample sizes, high factor loadings, and prior distributions with 20% variation provided about 84% in both convergence rates and convergence rates with adequate model fit.

Similar to convergence rates under measurement invariance conditions, varying types and magnitudes of noninvariance conditions yielded good convergence rates. As shown in Table 5, overall convergence rates ranged from 83.3% to 100% in all conditions.

In most conditions, the convergence rates reached 100%. The low convergence rates were observed in the conditions where there were large sample sizes, high factor loadings, and prior distributions with 20% variation. Under those conditions, the convergence rates ranged from 83.3% to 93.0%. Furthermore, the percentage of time a model converged with adequate model fit over 1,000 replications was also kept track for each condition. In this study, a model with PPP value greater than .05 was considered as an adequate model fit. As shown in Tables 5 and 6, converged replications showed adequate model fits in terms of PPP value in most conditions. The convergence rates with adequate model fit were also very similar to those in both measurement invariance and measurement noninvariance conditions.

Given that both measurement invariance and noninvariance conditions provided similar convergence rates and convergence rates with adequate model fit in this simulation, it seemed that the type and magnitude of noninvariance did not impact the convergence and model fit in this study. The properly converged replications with adequate model fit presented in Table 6 were used to evaluate the three main outcome measures.

Table 4
Percentage of Model Convergence and Model Convergence with Adequate Model Fit under Measurement Invariance

		Convergence Rates			Convergence Rates with Adequate Model Fit		
N	λ	P0	P1	P2	P0	P1	P2
100	.5	98.4	98.7	98.6	97.8	97.9	97.9
100	.9	100	100	100	99.8	99.6	99.6
400	.5	100	100	100	99.6	99.5	98.4
400	.9	100	100	99.8	99.2	99.5	99.5
1,000	.5	100	100	98.9	99.7	99.6	99.4
1,000	.9	100	100	84.6	99.6	99.5	84.2

Note.
1. N: sample size per group, λ: factor loading.
2. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 5

Percentage of Model Convergence under Measurement Noninvariance

Factor Loading Noninvariance Only

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	98.4	98.6	98.6	98.1	98.4	98.4	97.0	97.6	97.3	97.0	97.2	97.0	96.1	96.9	96.6	93.6	94.1	93.6
100	.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
400	.5	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
400	.9	100	100	99.9	100	100	99.9	100	100	100	100	100	100	100	100	100	100	100	100
1,000	.5	100	100	99.4	100	100	99.1	100	100	99.0	100	100	99.1	100	100	99.1	100	100	99.2
1,000	.9	100	99.9	84.2	100	100	86.2	100	100	85.3	100	100	88.2	100	100	85.7	100	100	90.3

Intercept Noninvariance Only

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	98.9	99.0	98.9	98.5	99.0	99.0	98.2	98.4	98.3	98.6	98.9	98.8	99.2	99.2	99.0	98.9	99.2	99.3
100	.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
400	.5	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
400	.9	100	100	99.8	100	100	99.7	100	100	99.8	100	100	99.8	100	100	99.7	100	100	99.7
1,000	.5	100	100	99.4	100	100	99.1	100	100	99.0	100	100	98.9	100	100	99.1	100	100	99.3
1,000	.9	100	100	83.3	100	100	85.1	100	100	83.7	100	100	83.9	100	100	83.9	100	100	84.2

Both Factor Loading and Intercept Noninvariance

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	98.1	98.6	98.5	97.9	98.3	98.2	97.4	97.4	97.2	97.2	96.9	96.6	96.1	96.9	96.7	93.7	93.9	93.7
100	.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
400	.5	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
400	.9	100	100	99.8	100	100	99.9	100	100	100	100	100	99.9	100	100	100	100	100	100
1,000	.5	100	100	99.2	100	100	99.0	100	100	99.2	100	100	99.5	100	100	99.1	100	100	99.3
1,000	.9	100	99.9	83.3	100	100	84.5	100	100	84.7	100	100	87.9	100	100	84.0	100	100	90.4

Note.

1. N: sample size per group, λ: factor loading.
2. NF: factor loading noninvariance only, NI: intercept noninvariance only, NFI: both factor loading and intercept noninvariance, NF_2025: factor loading noninvariance only with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 6

Percentage of Model Convergence with Adequate Model Fit under Measurement Noninvariance

Factor Loading Noninvariance Only

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	98.1	98.4	98.3	97.5	97.6	97.6	96.3	96.8	96.5	96.3	96.5	96.3	95.5	96.1	95.8	93.0	93.3	92.8
100	.9	99.2	99.0	99.0	99.8	99.6	99.6	99.8	99.7	99.7	99.9	99.6	99.6	99.9	99.8	99.8	99.8	99.6	99.6
400	.5	99.6	99.4	98.8	99.6	99.5	98.6	99.6	99.4	98.4	99.5	99.5	98.6	99.7	99.5	98.5	99.6	99.5	98.7
400	.9	99.6	99.8	99.8	99.2	99.5	99.6	99.3	99.5	99.5	99.2	99.5	99.6	99.3	99.5	99.4	99.2	99.5	99.5
1,000	.5	99.0	98.8	98.8	99.7	99.6	99.5	99.8	99.6	99.6	99.7	99.6	99.6	99.7	99.6	99.7	99.7	99.6	99.6
1,000	.9	99.6	99.3	83.7	99.6	99.5	85.7	99.5	99.5	84.9	99.5	99.5	87.7	99.4	99.3	85.1	99.5	99.4	89.8

Intercept Noninvariance Only

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	98.7	98.8	98.7	97.8	98.2	98.3	97.3	97.5	97.3	97.9	98.0	97.9	98.5	98.4	98.2	98.8	99.0	99.0
100	.9	99.2	99.0	99.0	99.4	99.1	99.1	99.6	99.7	99.6	99.4	99.3	99.3	99.5	99.2	99.2	99.5	99.5	99.5
400	.5	99.6	99.4	98.8	99.6	99.2	98.3	99.7	99.7	98.6	99.3	99.1	98.0	99.5	99.4	98.6	99.7	99.5	98.7
400	.9	99.6	99.8	99.8	99.6	99.3	99.4	99.0	99.2	99.2	99.1	99.1	99.1	99.2	99.2	99.2	99.9	99.9	99.9
1,000	.5	99.1	98.8	98.6	99.4	99.2	99.0	99.5	99.5	99.3	99.4	99.3	99.1	99.5	99.1	98.9	99.1	99.4	99.1
1,000	.9	99.6	99.3	82.8	99.5	99.1	84.4	99.7	99.6	83.4	99.3	99.1	83.2	99.4	99.2	83.1	99.7	99.4	83.8

Both Factor Loading and Intercept Noninvariance

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	97.6	97.9	97.8	97.3	97.7	97.6	96.7	96.5	96.3	96.7	96.4	96.1	95.5	96.1	96.1	93.2	93.3	93.1
100	.9	99.2	99.3	99.3	99.8	99.6	99.5	99.7	99.1	99.2	99.6	99.6	99.5	99.9	99.8	99.6	99.6	99.6	99.6
400	.5	99.7	99.8	99	99.6	99.5	98.5	99.5	99.2	98.4	99.4	99.1	98.6	99.7	99.5	98.4	99.4	99.2	98.6
400	.9	99.6	99.3	99.3	99.4	99.5	99.4	99.3	99.4	99.4	99.4	99.5	99.5	99.3	99.5	99.6	99.3	99.5	99.5
1,000	.5	99.2	99.2	99.1	99.8	99.5	99.4	99.5	99.2	99.1	99.5	99.3	99.3	99.7	99.6	99.5	99.7	99.7	99.6
1,000	.9	99.7	99.4	82.9	99.7	99.6	84.1	99.2	99.1	84.1	99.5	99.3	87.2	99.4	99.3	83.3	99.5	99.2	89.5

Note.

1. N: sample size per group, λ: factor loading.
2. NF: factor loading noninvariance only, NI: intercept noninvariance only, NFI: both factor loading and intercept noninvariance, NF_2025: factor loading noninvariance only with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Accuracy of Statistical Conclusion for Structural Parameter Comparisons

The accuracy of statistical conclusions for structural parameter comparisons between two populations was evaluated using Type I error and empirical power. In this study, the Type I error rate is the percentage of replications in which models erroneously detect the structural parameter differences when population differences in structural parameters are truly zero. The empirical power is the percentage of replications in which models properly detect the structural parameter differences when population differences in the structural parameters truly exist. Type I error and empirical power were determined based on 95% credibility intervals of the posterior distributions of the three structural parameter differences: structural regression coefficient difference, exogenous factor mean difference, and endogenous factor mean difference.

Type I Error

Table 7 presents the Type I error rates obtained from the measurement invariance models by sample size, factor loading size, and prior distribution. The Type I error rates ranged from < .1% and to 5.3% across all sample sizes, factor loading sizes, and prior distributions. Interestingly, most Type I error rates of the three structural parameter differences decreased as prior distributions allowed more degrees of noninvariance in reference indicators' parameters between populations. For example, all conditions with prior distributions with zero variation (i.e., prior distributions did not allow noninvariance of reference indicators' parameters between two populations) provided the Type I error rates that were close to the nominal 5% error rate. However, all conditions with prior distributions with 10% or 20% variation (i.e., prior distributions allowed noninvariance of the reference indicators' parameters within 10% or 20% between two populations, respectively) provided Type I error rates that were much lower than the nominal 5% error rate. Particularly, the Type I error rates for both exogenous and endogenous factor mean differences were substantially lower, with Type I error rates being close to zero with moderate or large sample sizes and prior distributions with 20% variation.

The results of the Type I error rates in three types of measurement noninvariance conditions are summarized by magnitude of noninvariance, sample size, factor loading size, and prior distribution. As seen in Tables 8 through 10, the three types of noninvariance conditions (i.e., factor loading noninvariance only, intercept noninvariance only, both factor loading and intercept noninvariance) provided very similar patterns of Type I error rates. For example, the three types of measurement noninvariance conditions yielded very similar ranges of Type I error rates which are from < .1% to 6.0% for factor loading noninvariance only conditions, from < .1% to 6.2% for intercept noninvariance only conditions, and from < .1% to 6.1% for both factor loading and intercept noninvariance conditions. The six different magnitudes of noninvariance conditions also provided similar patterns of Type I error rates which are less than 6% in most conditions. In the current study, the Type I error rates obtained from varying degrees of measurement invariance models were not worse than those obtained from the measurement invariance models. These results indicate that the different types and magnitudes of noninvariance do not have an impact on the Type I error rates.

As observed in measurement invariance conditions, it seemed that the sample size and prior distribution impacted the Type I error rates. As sample sizes and variance of prior distributions increased, the Type I error rates decreased in all different types and magnitudes of noninvariance conditions. Under prior distributions with zero variation conditions, the Type I error rates were close to the nominal 5% error rate, yielding excellent Type I error controls across different levels of measurement noninvariance, sample size, and factor loading size. Interestingly, under prior distributions with 10% or 20% variation, the Type I error rates generally tended to fall below the nominal 5% error rate. Under prior distributions with 20% variation, the Type I error rates of structural regression coefficient difference, exogenous factor mean difference, and endogenous factor mean difference were very close to zero particularly when sample size was moderate or large.

Table 7

Type I Error Rates (%): Measurement Invariance

		Structural Regression Coefficient Difference (γ^R - γ^F)			Exogenous Factor Mean Difference (κ^F)			Endogenous Factor Mean Difference (α^F)		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	4.0	4.2	2.9	4.3	2.5	.6	4.0	2.2	.4
100	.9	3.9	3.5	2.3	4.2	3.8	1.4	3.8	3.3	1.0
400	.5	3.5	2.3	1.5	4.3	.1	< .1	5.3	.1	< .1
400	.9	3.9	2.4	.9	4.2	1.1	< .1	5.2	.6	< .1
1,000	.5	3.7	1.5	.4	4.0	< .1	< .1	4.5	< .1	< .1
1,000	.9	4.0	.8	.1	4.0	.1	< .1	4.2	.1	< .1

Note.

1. N: sample size per group, λ: factor loading.
2. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 8

Type I Error Rates (%): Factor Loading Noninvariance Only

Structural Regression Coefficient Difference (γ^R - γ^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	4.1	4.0	3.1	4.0	3.9	2.8	3.8	3.8	2.8	3.8	3.5	2.9	4.1	3.7	2.7	4.0	3.3	2.7
100	.9	4.3	4.1	3.2	3.9	3.5	2.3	3.9	3.4	2.4	3.9	3.5	2.2	3.6	3.5	2.4	3.7	3.4	2.0
400	.5	5.4	3.2	1.1	3.7	2.4	1.5	3.7	2.5	1.8	3.4	2.4	1.6	3.6	3.2	1.5	3.8	2.5	1.6
400	.9	5.2	3.4	.3	3.8	2.3	.9	4.1	2.4	.9	3.7	2.3	.9	3.8	2.5	.9	3.7	2.3	1.0
1,000	.5	4.7	1.9	.2	3.7	1.5	.4	3.4	1.5	.4	3.8	1.5	.4	3.7	1.5	.4	3.8	1.7	.4
1,000	.9	4.0	1.4	.1	4.0	.8	.1	4.2	.9	.1	4.0	.9	.1	3.9	.9	< .1	4.0	.9	.1

Exogenous Factor Mean Difference (κ^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	6.1	3.1	.7	4.3	2.5	.6	4.5	3.2	.6	4.4	2.7	.6	4.5	3.4	.6	4.3	2.6	.7
100	.9	5.9	3.9	2.1	4.2	3.8	1.5	4.0	3.8	1.5	4.1	3.8	1.5	4.0	4.0	1.5	4.3	3.8	1.5
400	.5	4.2	.6	< .1	4.3	.1	< .1	4.5	.1	< .1	4.4	.1	< .1	4.5	.1	< .1	4.6	.1	< .1
400	.9	4.0	1.6	.1	4.2	1.2	< .1	4.2	1.2	< .1	4.2	1.2	< .1	4.2	1.2	< .1	4.2	1.2	< .1
1,000	.5	4.5	< .1	< .1	4	< .1	< .1	4.2	< .1	< .1	4.0	< .1	< .1	4.2	< .1	< .1	4.0	< .1	< .1
1,000	.9	4.5	.2	< .1	4	.1	< .1	4.0	.1	< .1	4.0	.1	< .1	4.1	.1	< .1	4.1	.1	< .1

Endogenous Factor Mean Difference (α^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	3.3	1.0	.1	3.8	2.3	.4	3.8	2.2	.4	3.7	2.4	.4	3.6	2.3	.5	3.8	2.3	.4
100	.9	3.2	2.2	.5	3.8	3.2	1.0	3.9	3	1.0	3.8	3.2	1.0	3.7	3.3	1.2	3.9	3.2	1.0
400	.5	4.5	< .1	< .1	5.4	.1	< .1	5.6	.2	< .1	5.3	.1	< .1	5.5	.2	< .1	5.4	.1	< .1
400	.9	3.9	.7	< .1	5.3	.6	< .1	5.0	.4	< .1	5.3	.6	< .1	5.2	.4	< .1	5.2	.6	< .1
1,000	.5	4.7	< .1	< .1	4.5	< .1	< .1	4.5	< .1	< .1	4.5	< .1	< .1	4.3	< .1	< .1	4.6	< .1	< .1
1,000	.9	4.3	< .1	< .1	4.2	.1	< .1	4.2	.1	< .1	4.4	.1	< .1	4.1	.1	< .1	4.4	.1	< .1

Note.

1. N: sample size per group, λ: factor loading.
2. NF: factor loading noninvariance only, NF_2025: factor loading noninvariance only with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 9

Type I Error Rates (%): Intercept Noninvariance Only

Structural Regression Coefficient Difference ($\gamma^R - \gamma^F$)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	4.0	4.4	3.4	4.6	3.5	3.1	5.4	5.1	3.9	5.1	3.9	3.3	5.4	5.5	4.4	4.0	3.6	3.0
100	.9	4.3	4.0	3.2	4.7	4.1	3.3	4.4	4.8	2.9	3.8	3.2	2.3	4.9	5.4	3.5	4.0	4.0	2.4
400	.5	5.4	3.3	1.2	5.2	3.0	1.7	5.2	3.5	1.2	4.7	2.7	.9	3.8	2.3	.7	5.6	2.9	1.2
400	.9	5.2	3.3	.3	4.7	3.1	.8	4.0	2.3	.4	3.7	2.1	.3	3.9	2.0	.6	4.8	2.7	1.1
1,000	.5	4.7	1.8	.3	3.3	1.2	.3	3.9	1.4	.5	3.6	1.1	.2	4.1	.9	.2	4.3	1.7	.4
1,000	.9	3.9	1.4	.1	4.1	1.3	.1	4.1	1.4	.1	4.2	1.3	.1	4.1	1.0	.2	5.0	1.3	.3

Exogenous Factor Mean Difference (κ^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	6.2	3.0	.7	5.1	2.3	.4	5.2	3.3	.7	4.5	1.7	.5	5.7	2.5	.5	4.8	2.6	.5
100	.9	6.0	3.9	2.2	5.0	3.6	1.2	5.2	4.7	2.1	4.3	3.2	1.3	5.7	4.3	1.5	4.9	3.7	1.8
400	.5	3.9	.6	< .1	4.2	.1	< .1	5.2	.4	< .1	4.1	.5	< .1	4.0	.3	< .1	4.7	.3	< .1
400	.9	4.0	1.6	.2	4.1	1.0	< .1	5.3	2.2	.1	4.0	1.5	.2	4.0	1.5	< .1	4.7	1.3	< .1
1,000	.5	4.4	< .1	< .1	4.6	< .1	< .1	4.6	< .1	< .1	3.7	< .1	< .1	4.2	< .1	< .1	5.2	.1	< .1
1,000	.9	4.5	.2	< .1	4.8	.3	< .1	4.8	.2	< .1	4.2	.1	< .1	4.2	.1	< .1	5.1	.4	< .1

Endogenous Factor Mean Difference (α^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	3.1	1.0	.1	3.9	1.6	.1	4.0	1.6	.2	4.5	1.7	.1	4.3	1.6	< .1	3.5	1.2	< .1
100	.9	3.2	2.3	.5	4.0	2.6	.7	3.9	2.9	.8	4.6	2.7	.9	3.9	2.8	.5	4.0	2.3	.7
400	.5	4.3	< .1	< .1	5.0	.1	< .1	4.5	< .1	< .1	4.1	< .1	< .1	4.8	.1	< .1	5.8	.2	< .1
400	.9	3.8	.7	< .1	5.2	.8	.1	4.4	.3	< .1	3.9	.6	< .1	4.7	.6	< .1	5.3	1.3	< .1
1,000	.5	4.6	< .1	< .1	3.8	< .1	< .1	3.6	< .1	< .1	5.0	< .1	< .1	4.3	< .1	< .1	4.3	< .1	< .1
1,000	.9	4.4	< .1	< .1	4.2	.1	< .1	3.7	< .1	< .1	5.4	.3	< .1	4.0	.1	< .1	4.7	< .1	< .1

Note.

1. N: sample size per group, λ: factor loading.
2. NI: intercept noninvariance only, NI_2025: intercept noninvariance only with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 10

Type I Error Rates (%): Both Factor Loading and Intercept Noninvariance

Structural Regression Coefficient Difference (γ^R - γ^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	4.9	4.5	3.7	3.0	2.8	2.1	4.4	4.3	3.5	3.6	2.9	2.4	4.1	3.7	2.8	5.0	4.4	3.9
100	.9	4.2	3.5	2.9	3.0	3.0	1.9	3.9	3.8	2.3	2.9	3.1	2.0	3.6	3.5	3.0	4.4	4.2	2.3
400	.5	5.1	2.8	.9	4.0	2.2	1.1	3.4	2.0	.9	4.3	2.5	1.4	3.6	3.2	1.7	3.4	2.4	1.4
400	.9	4.0	2.6	.4	3.9	2.4	.6	3.2	2.0	.6	3.6	2.4	.8	3.8	2.5	.9	3.9	2.0	.6
1,000	.5	4.4	2.1	.3	4.7	1.8	.5	3.9	1.4	.6	3.9	1.6	.5	3.7	1.5	.3	5.7	2.3	.8
1,000	.9	4.1	1.9	.2	4.9	1.3	< .1	4.8	1.4	.1	4.7	1.6	.1	3.9	.9	< .1	5.6	2.0	.3

Exogenous Factor Mean Difference (κ^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	6.1	4.0	.9	4.9	2.7	.4	5.0	2.7	.2	5.2	2.9	.3	4.5	3.4	.4	4.9	3.0	.5
100	.9	6.0	4.6	2.7	5.0	4.1	1.4	5.0	3.8	1.2	5.0	4.2	1.4	4.0	4.0	1.6	4.7	4.3	1.7
400	.5	4.1	.4	< .1	4.9	.3	< .1	5.1	.1	< .1	4.1	< .1	< .1	4.5	.1	< .1	4.0	.1	< .1
400	.9	4.0	1.2	.1	5.0	1.6	.1	4.8	1.4	< .1	3.9	1.1	< .1	4.2	1.2	< .1	3.6	1.1	< .1
1,000	.5	4.0	< .1	< .1	3.3	< .1	< .1	4.1	< .1	< .1	3.9	< .1	< .1	4.2	< .1	< .1	4.4	< .1	< .1
1,000	.9	3.9	.3	< .1	3.3	.1	< .1	3.9	.1	< .1	4.0	.2	< .1	4.1	.1	< .1	4.9	.1	< .1

Endogenous Factor Mean Difference (α^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	3.1	1.4	.1	4.2	1.7	.1	3.8	1.6	.1	4.4	2.0	.2	3.6	2.3	.3	3.9	1.8	.3
100	.9	2.9	3.0	.8	4.3	3.1	.7	3.9	2.4	.7	4.4	3.1	.9	3.7	3.3	1.1	3.4	2.7	.8
400	.5	5.0	.2	< .1	5.2	< .1	< .1	5.5	.1	< .1	4.7	< .1	< .1	5.5	.2	< .1	4.5	< .1	< .1
400	.9	4.2	1.1	< .1	4.8	.6	< .1	5.2	.8	< .1	5.0	.8	< .1	5.2	.4	< .1	4.1	.8	< .1
1,000	.5	5.0	< .1	< .1	3.8	< .1	< .1	4.3	< .1	< .1	4.3	< .1	< .1	4.3	< .1	< .1	3.1	< .1	< .1
1,000	.9	4.8	< .1	< .1	3.6	.1	< .1	4.0	.1	< .1	4.6	.1	< .1	4.1	.1	< .1	3.3	.1	< .1

Note.

1. N: sample size per group, λ: factor loading.
2. NFI: both factor loading and intercept noninvariance, NFI_2025: both factor loading and intercept noninvariance with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Power

Table 11 summarizes the empirical power rates obtained from the measurement invariance models by sample size, factor loading size, and prior distribution. In general, the power rates for detecting exogenous factor mean differences were highest, followed by power rates of endogenous factor mean differences and power rates of structural regression coefficient differences. The power rates of structural regression coefficient differences were much lower than those of exogenous factor mean differences when sample size was small or moderate. As expected, the sample size and factor loading size impacted the empirical power rates. As sample size and factor loading size increased, the power rates for detecting the three structural parameter differences increased. It seemed that prior distribution and interaction between prior distribution and sample size also impacted the empirical power rates. Most conditions with prior distributions with 10% variation provided the highest power rates, followed by conditions with prior distributions with zero variation and conditions with prior distributions with 20% variation when sample size was small. When sample size was moderate or large, the empirical power rates for detecting exogenous and endogenous factor mean differences were equal to or close to 100%. For power rates for detecting structural regression coefficient differences, both prior distribution with zero variation and with 10% variation provided similar power rates while prior distribution with 20% variation provided lowest power rates when sample size was moderate or large.

Tables 12 through 14 summarize the empirical power rates obtained from the three types of measurement noninvariance conditions by magnitude of noninvariance, sample size, factor loading size, and prior distribution. It should be noted that when sample size was small or moderate, overall empirical power rates for detecting all three structural parameter differences were slightly lower when the focal group had higher values of the three structural parameters rather than when focal group had lower values of the three structural parameters in population models. However, the difference between the two conditions was small and the pattern of power results was similar across most of other conditions and hence the results were averaged over these conditions.

In general, the three types of noninvariance conditions provided very similar patterns of power rates across all conditions. The differences in empirical power rates across the three types of noninvariance conditions were within 3% for structural regression coefficient differences, within 3.4% for exogenous factor mean differences, and within 4.2% for endogenous factor mean differences. It seemed that the different magnitudes of measurement noninvariance also did not impact empirical power rates for the three structural parameter differences in this simulation. The empirical power rates for detecting the three structural parameter differences were very similar across different magnitudes of measurement noninvariance when holding the other factors constant. As observed in the measurement invariance conditions, the empirical power rates for detecting exogenous factor mean differences were highest while the empirical power rates for detecting structural regression coefficient differences were lowest. In the current study, empirical power rates obtained on the basis of varying degrees of measurement noninvariance models were not worse than those obtained from the measurement invariance models. These results indicate that the different types and magnitudes of noninvariance have no effect on the power rate.

As observed in measurement invariance conditions, most power rates of the three structural parameter differences were influenced by sample sizes, factor loadings, and prior distributions across all measurement noninvariance conditions. As sample size and factor loading increased, the power rates also increased in all different types and magnitudes of noninvariance conditions. Regarding the prior distributions, most conditions with prior distributions with 10% variation provided highest power rates of all three structural parameter differences when sample size was small. When sample size was moderate or large, both prior distributions with zero variation and prior distributions with 10% variation provided similar power rates while prior distributions with 20% variation provided lowest power rates. Irrespective of prior distributions, all conditions with moderate or large sample size showed almost perfect power for detecting the exogenous and endogenous factor mean differences.

Table 11

Power Rates (%): Measurement Invariance

		Structural Regression Coefficient Difference (γ^R - γ^F)			Exogenous Factor Mean Difference (κ^F)			Endogenous Factor Mean Difference (α^F)		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	28.9	31.7	28.9	73.3	88.5	73.3	54.9	67.5	48.2
100	.9	49.5	51.0	46.1	89.3	98.6	96.4	84.0	95.2	90.6
400	.5	82.6	81.2	70.4	100	100	99.0	99.0	99.7	93.7
400	.9	97.8	96.5	87.5	100	100	100	100	100	100
1,000	.5	99.2	99.0	96.0	100	100	100	100	100	99.0
1,000	.9	100	100	99.8	100	100	100	100	100	100

Note.

1. N: sample size per group, λ: factor loading.

2. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 12

Power Rates (%): Factor Loading Noninvariance Only

Structural Regression Coefficient Difference (γ^R - γ^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	27.9	31.8	29.0	28.6	31.3	28.4	27.7	30.7	28.4	28.1	30.4	27.8	27.5	30.6	28.5	27.3	29.3	26.9
100	.9	49.9	51.0	46.0	49.4	50.9	46.0	49.4	50.3	46.2	48.6	50.5	45.6	49.3	50.2	45.9	48.3	49.7	45.2
400	.5	81.2	80.5	70.2	82.2	80.8	70.1	81.5	80.4	69.8	81.2	79.4	69.2	80.9	80.0	69.2	79.5	78.2	68.3
400	.9	97.6	96.3	87.7	97.8	96.5	87.2	97.5	96.6	87.2	97.6	96.5	86.9	97.4	96.7	87.3	97.6	96.4	86.6
1,000	.5	99.3	99.4	96.3	99.2	99.0	95.8	99.0	99.0	95.4	99.0	98.8	95.3	99.2	98.8	95.3	98.8	98.6	94.7
1,000	.9	100	100	100	100	100	99.8	100	100	99.8	100	100	99.8	100	100	99.8	100	100	99.8

Exogenous Factor Mean Difference (κ^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	73.0	90.1	75.0	73.3	88.8	73.3	72.8	88.9	74.4	73.2	88.5	73.4	72.6	89.0	74.8	72.8	88.5	73.0
100	.9	89.1	98.7	97.3	89.2	98.6	96.4	89.1	98.6	96.4	89.1	98.5	96.4	88.9	98.6	96.6	89.1	98.6	96.4
400	.5	99.9	100	99.2	100	100	99.2	100	100	99.3	100	100	99.2	100	100	99.4	100	100	99.2
400	.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
1,000	.5	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
1,000	.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Endogenous Factor Mean Difference (α^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	55.4	68.5	50.5	54.4	67.1	47.8	53.7	66.8	47.0	53.3	66.1	46.9	53.0	66.4	47.2	52.3	63.9	45.4
100	.9	85.2	96.5	92.2	84.0	95.1	90.5	84.0	95.2	90.2	83.8	94.9	90.2	83.9	95.1	90.2	83.5	94.5	90.0
400	.5	99.0	99.6	93.8	99.0	99.7	93.7	99.0	99.4	93.0	99.0	99.6	93.5	99.0	99.4	92.7	99.0	99.5	93.0
400	.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
1,000	.5	100	100	99.1	100	100	99.0	100	100	99.1	100	100	98.8	100	100	99.1	100	100	98.6
1,000	.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Note.

1. N: sample size per group, λ: factor loading.
2. NF: factor loading noninvariance only, NF_2025: factor loading noninvariance only with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 13

Power Rates (%): Intercept Noninvariance Only

Structural Regression Coefficient Difference ($\gamma^R - \gamma^F$)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	28.6	32.6	29.4	30.0	32.0	29.8	30.7	32.5	30.6	30.6	32.2	30.2	30.0	33.2	29.9	30.1	32.1	29.3
100	.9	50.5	51.2	46.2	50.1	50.7	45.3	49.2	50.8	46.4	50.2	50.8	46.0	49.4	50.8	46.2	50.1	51.5	45.8
400	.5	81.5	80.8	70.8	82.2	81.3	71.0	82.2	82.5	71.5	82.0	81.0	70.6	82.3	81.0	69.5	81.5	80.5	69.8
400	.9	97.6	96.4	88.0	97.6	96.2	87.9	97.3	95.8	88.2	98.0	96.8	88.3	98.0	97.0	88.0	97.6	97.0	88.1
1,000	.5	99.4	99.4	96.4	99.7	99.2	96.0	99.6	99.0	96.0	99.6	99.2	96.8	99.2	99.2	96.5	99.0	99.0	96.6
1,000	.9	100	100	100	100	100	99.9	100	100	99.8	100	100	99.8	100	100	99.7	100	100	99.7

Exogenous Factor Mean Difference (κ^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	73.2	90.1	74.4	74.6	89.7	76.0	73.1	88.5	75.1	74.5	89.1	76.8	72.4	88.9	74.2	73.3	90.3	75.4
100	.9	89.2	98.7	97.2	90.5	98.8	97.1	88.8	98.4	96.6	89.6	99.0	97.0	89.4	98.5	96.8	90.4	98.6	97.0
400	.5	99.9	100	99.2	100	100	99.4	99.9	100	99.0	99.8	100	99.2	100	100	99.3	99.8	100	99.1
400	.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
1,000	.5	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
1,000	.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Endogenous Factor Mean Difference (α^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	56.0	69.5	50.8	56.5	67.7	48.4	55.5	68.0	48.7	53.9	67.7	47.8	53.9	66.3	47.5	55.1	68.0	49.6
100	.9	85.1	96.8	92.2	85.9	95.8	91.8	84.0	95.2	90.9	83.7	95.6	91.4	83.0	95.8	90.9	84.5	96.2	90.8
400	.5	99.0	99.6	94.1	98.9	99.4	93.2	99.0	99.9	94.2	98.9	99.6	93.2	98.8	99.8	92.8	99.0	99.4	94.0
400	.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
1,000	.5	100	100	99.0	100	100	99.0	100	100	99.4	100	100	99.0	100	100	99.4	100	100	99.2
1,000	.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Note.

1. N: sample size per group, λ: factor loading.
2. NI: intercept noninvariance only, NI_2025: intercept noninvariance only with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 14

Power Rates (%): Both Factor Loading and Intercept Noninvariance

Structural Regression Coefficient Difference (γ^R - γ^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	29.3	32.0	29.3	29.4	32.1	29.3	27.5	30.7	28.4	27.4	29.3	26.7	27.5	30.6	28.5	27.4	30.3	27.5
100	.9	50.0	51.4	45.6	48.9	50.7	45.9	48.4	50.1	46.5	48.2	49.9	44.9	49.3	50.2	46.2	47.7	49.2	44.7
400	.5	80.7	80.3	69.2	81.5	79.8	70.0	81.0	80.0	70.0	80.8	79.7	69.8	80.9	80.0	69.4	78.3	77.6	67.7
400	.9	97.8	96.6	87.6	97.6	96.4	87.0	97.8	97.2	87.7	97.7	96.7	87.1	97.4	96.7	87.7	97.5	96.4	86.3
1,000	.5	99.4	99.4	96.1	99.0	99.0	95.9	99.2	98.7	95.2	99.1	98.8	95.1	99.2	98.8	95.3	98.4	98.2	95.0
1,000	.9	100	100	100	100	100	99.7	100	100	99.6	100	100	99.7	100	100	99.8	100	100	99.8

Exogenous Factor Mean Difference (κ^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	73.2	90.3	76.2	72.9	88.7	74.0	72.3	89.9	75.7	73.2	89.4	74.2	72.6	89.0	75.0	73.1	88.5	74.6
100	.9	89.9	98.1	96.6	88.3	98.5	96.2	89.7	99.0	96.8	89.1	98.4	96.5	88.9	98.6	96.6	89.1	98.5	96.2
400	.5	99.9	100	99.4	100	100	98.8	100	100	99.2	100	100	99.4	100	100	99.1	100	100	99.2
400	.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
1,000	.5	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
1,000	.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Endogenous Factor Mean Difference (α^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	54.2	66.8	49.0	53.6	66.0	46.1	52.7	66.3	46.0	52.2	65.7	45.5	53.0	66.4	46.9	49.6	63.4	44.1
100	.9	84.2	96.0	91.5	84.0	95.1	90.8	84.6	95.0	90.2	85.0	95.9	91.0	83.9	95.1	90.5	83.4	94.8	89.5
400	.5	98.8	99.4	93.0	99.0	99.6	93.7	98.8	99.5	92.3	99.0	99.8	92.8	99.0	99.4	92.0	98.8	99.4	92.9
400	.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
1,000	.5	100	100	98.7	100	100	99.1	100	100	98.6	100	100	99.2	100	100	99.0	100	100	99.0
1,000	.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Note.

1. N: sample size per group, λ: factor loading.
2. NFI: both factor loading and intercept noninvariance, NFI_2025: both factor loading and intercept noninvariance with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Precision of the Estimated Structural Parameter Difference

The precision of the estimated structural parameter difference between two populations was evaluated using the width of 95% credibility intervals of individual structural parameter difference estimates. Tables 15 through 18 present the precision of the estimated three structural parameter differences by magnitude of invariance, sample size, factor loading size, and prior distribution in measurement invariance and noninvariance conditions. It should be noted that the pattern of results was similar to different magnitudes of structural parameter differences and hence the results were averaged over these conditions.

Table 15 summarizes the precision of the estimated three structural parameter differences obtained from measurement invariance conditions by sample size, factor loading size, and prior distribution. It was clear that the width of 95% credibility intervals of all structural parameter difference parameter estimates became smaller when sample size or factor loading size increased. It seemed that prior distribution also had an impact on the width of 95% credibility intervals of all structural parameter difference parameter estimates. Generally, the width of 95% credibility intervals were narrowest under the conditions with the prior distribution with zero variation while the width of 95% credibility intervals were widest under the conditions with prior distributions with 20% variation.

Tables 16 through 18 summarize the width of 95% credibility intervals of the estimated three structural parameter differences obtained from the three types of measurement noninvariance conditions by magnitude of noninvariance, sample size, factor loading size, and prior distribution. In general, the three types of noninvariance conditions provided very similar patterns of precision. The averaged differences in the width of 95% credibility intervals between factor loading noninvariance only conditions and intercept noninvariance only conditions were trivial, yielding less than .001 for all three structural parameter differences. The largest differences in the width of 95% credibility intervals between factor loading noninvariance only conditions and intercept noninvariance only conditions were .009 for structural regression coefficient differences, .022 for exogenous factor mean differences, and .038 for endogenous factor mean differences when holding the other factors constant. Further, it was observed that the width of 95% credibility interval of individual structural parameter difference estimates was very similar across different magnitudes of measurement noninvariance. These results indicate that the magnitude of measurement noninvariance has no impact on the precision of the three structural parameter difference estimates.

Instead, it seemed that sample size, factor loading size, and prior distribution influenced the precision of the three structural parameter difference estimates. As expected, the precision increased as sample size or factor loading increased. As observed in measurement invariance conditions, prior distribution with 10% variation provided highest precision levels when sample size was small while prior distribution with zero variation conditions provided highest precision levels when sample size was moderate or large. Generally, prior distribution with 20% variation conditions provided the lowest precision levels.

Tables 19 through 21 summarize the results of the factorial ANOVA on the precision of the three structural parameter difference estimates with five main manipulated factors (i.e., total magnitude of noninvariance, percentage of noninvariance items, sample size, factor loading size, prior distribution) and their combinations (i.e., two-way interactions) in three types of measurement invariance conditions. The effect size measure, $\eta^2 = SS_{effect} / SS_{total}$, was used to examine significant main and interaction effects on the precision of the three structural parameter difference estimates. η^2 can be interpreted as the proportion of variance associated with each of main or interaction effects in an ANOVA study (Thompson, 2013). η^2 of .01, .06, and .14 were used to represent small, moderate, and large effects, respectively, for factorial ANOVA analysis (Cohen, 1988). As seen in Tables 19 through 21, the results of ANOVA were very similar across three types of noninvariance conditions. The total magnitude of noninvariance and percentage of noninvariance items had no effects on the precision of the three structural parameter difference estimates. The significant factors influencing precision of the three structural parameter difference estimates were sample size, factor loading size, and prior distribution. Interestingly, the sample size, factor loading size, and prior distribution were found to have large, moderate, and small effects, respectively, on the precision of structural regression coefficient differences while the sample size, factor loading size, and prior distribution were found to have large effects on the precision of both exogenous and endogenous factor mean differences. For interaction effects, two interaction effects (i.e., sample size × factor loading size, sample size × prior distribution) were found to have small effects on the precision of structural regression coefficient difference estimates. For the precision of exogenous factor mean difference estimates, two interaction effects (i.e., sample size × prior distribution, factor loading size × prior distribution) were found to have moderate and small effects. Lastly, for the precision of endogenous factor mean difference estimates, three interaction effects (i.e., sample size ×

prior distribution, sample size × prior distribution, factor loading size × prior distribution) were found to have small, moderate, and small effects, respectively.

Table 15

Average Precision: Measurement Invariance

		Structural Regression Coefficient Difference (γ^R - γ^F)			Exogenous Factor Mean Difference (κ^F)			Endogenous Factor Mean Difference (α^F)		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	1.066	.879	.928	.948	.741	.948	1.175	.879	1.127
100	.9	.652	.613	.670	.706	.562	.671	.754	.594	.723
400	.5	.442	.444	.527	.469	.492	.776	.501	.549	.880
400	.9	.311	.331	.426	.361	.338	.491	.360	.356	.543
1,000	.5	.276	.308	.396	.300	.421	.698	.319	.473	.770
1,000	.9	.193	.233	.325	.232	.269	.430	.23	.294	.475

Note.

1. N: sample size per group, λ: factor loading.

2. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 16

Average Precision: Factor Loading Noninvariance Only

Structural Regression Coefficient Difference ($\gamma^R - \gamma^F$)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	1.050	.880	.927	1.079	.891	.941	1.094	.911	.959	1.106	.917	.964	1.107	.930	.976	1.148	.950	.997
100	.9	.650	.617	.670	.655	.615	.672	.657	.618	.676	.660	.619	.676	.660	.621	.678	.666	.625	.682
400	.5	.447	.443	.523	.446	.447	.530	.451	.450	.532	.452	.453	.536	.455	.452	.533	.461	.461	.545
400	.9	.310	.333	.423	.312	.332	.427	.313	.332	.426	.314	.333	.429	.314	.332	.426	.316	.335	.431
1,000	.5	.277	.310	.397	.277	.310	.397	.279	.312	.398	.280	.313	.400	.281	.313	.399	.285	.317	.403
1,000	.9	.193	.233	.323	.194	.234	.326	.195	.234	.326	.195	.234	.327	.195	.234	.327	.196	.235	.328

Exogenous Factor Mean Difference (κ^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.947	.743	.947	.948	.741	.948	.956	.737	.940	.949	.741	.948	.962	.733	.934	.949	.742	.949
100	.9	.707	.567	.670	.707	.562	.671	.711	.560	.668	.707	.561	.670	.714	.559	.666	.708	.561	.670
400	.5	.467	.493	.773	.469	.492	.778	.464	.493	.776	.469	.493	.780	.461	.494	.774	.469	.494	.781
400	.9	.363	.340	.493	.361	.338	.491	.360	.338	.489	.361	.338	.491	.359	.339	.487	.361	.338	.491
1,000	.5	.297	.420	.697	.299	.421	.697	.297	.422	.698	.299	.421	.695	.297	.423	.700	.298	.421	.693
1,000	.9	.230	.270	.427	.232	.269	.429	.231	.270	.430	.231	.269	.428	.229	.270	.430	.230	.270	.427

Endogenous Factor Mean Difference (α^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	1.150	.873	1.117	1.179	.884	1.131	1.186	.886	1.130	1.187	.892	1.138	1.194	.889	1.129	1.198	.904	1.148
100	.9	.750	.593	.720	.753	.594	.724	.754	.594	.721	.754	.595	.724	.753	.593	.719	.755	.596	.725
400	.5	.500	.550	.880	.501	.550	.882	.501	.551	.884	.503	.550	.885	.501	.552	.887	.504	.552	.888
400	.9	.360	.357	.543	.360	.356	.544	.359	.357	.544	.360	.356	.545	.359	.357	.544	.361	.356	.545
1,000	.5	.320	.470	.770	.320	.473	.770	.320	.472	.772	.320	.472	.769	.320	.472	.773	.321	.472	.769
1,000	.9	.230	.297	.477	.230	.293	.475	.230	.293	.474	.230	.293	.473	.230	.293	.474	.230	.292	.472

Note.

1. N: sample size per group, λ: factor loading.
2. NF: factor loading noninvariance only, NF_2025: factor loading noninvariance only with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 17

Average Precision: Intercept Noninvariance Only

Structural Regression Coefficient Difference (γ^R - γ^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	1.035	.872	.917	1.053	.884	.934	1.075	.886	.934	1.055	.886	.933	1.058	.884	.933	1.058	.877	.921
100	.9	.648	.611	.667	.651	.614	.671	.656	.613	.671	.649	.613	.670	.651	.615	.672	.651	.611	.667
400	.5	.441	.442	.525	.443	.444	.528	.442	.443	.526	.442	.443	.527	.444	.446	.531	.441	.442	.525
400	.9	.312	.331	.424	.311	.332	.427	.312	.331	.426	.311	.331	.426	.312	.333	.430	.311	.331	.425
1,000	.5	.275	.308	.395	.274	.308	.395	.275	.307	.394	.275	.308	.396	.275	.308	.395	.275	.308	.395
1,000	.9	.193	.233	.324	.193	.233	.325	.193	.233	.324	.193	.233	.326	.193	.233	.324	.193	.233	.325

Exogenous Factor Mean Difference (κ^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.943	.742	.949	.94	.739	.946	.937	.739	.946	.941	.742	.948	.94	.741	.947	.937	.743	.949
100	.9	.705	.563	.671	.703	.560	.669	.701	.561	.669	.703	.562	.671	.704	.561	.670	.703	.563	.672
400	.5	.467	.492	.777	.469	.492	.777	.467	.492	.776	.469	.492	.776	.470	.492	.776	.468	.492	.776
400	.9	.361	.337	.491	.361	.337	.491	.361	.337	.490	.362	.338	.491	.361	.337	.491	.361	.338	.491
1,000	.5	.299	.421	.698	.300	.421	.698	.299	.421	.698	.300	.421	.698	.300	.421	.698	.299	.421	.697
1,000	.9	.232	.269	.430	.232	.269	.430	.232	.269	.43	.233	.269	.430	.232	.269	.431	.232	.269	.429

Endogenous Factor Mean Difference (α^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	1.146	.872	1.119	1.156	.881	1.129	1.179	.881	1.13	1.153	.881	1.13	1.156	.883	1.131	1.165	.876	1.126
100	.9	.749	.592	.721	.751	.594	.723	.755	.593	.722	.747	.594	.723	.750	.595	.724	.751	.591	.721
400	.5	.500	.548	.879	.501	.549	.880	.500	.548	.879	.501	.549	.88	.502	.55	.881	.500	.549	.881
400	.9	.360	.356	.543	.360	.356	.543	.36	.356	.543	.360	.356	.544	.361	.357	.544	.359	.356	.544
1,000	.5	.319	.473	.769	.319	.473	.769	.319	.473	.769	.320	.473	.770	.319	.473	.769	.319	.473	.769
1,000	.9	.229	.294	.474	.229	.294	.474	.230	.294	.475	.229	.294	.475	.229	.294	.474	.229	.294	.475

Note.

1. N: sample size per group, λ: factor loading.
2. NI: intercept noninvariance only, NI_2025: intercept noninvariance only with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 18

Average Precision: Both Factor Loading and Intercept Noninvariance

Structural Regression Coefficient Difference (γ^R - γ^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	1.079	.893	.942	1.076	.899	.949	1.094	.919	.969	1.091	.917	.963	1.107	.930	.981	1.142	.948	.998
100	.9	.656	.616	.674	.654	.615	.672	.656	.621	.678	.656	.619	.676	.660	.621	.678	.666	.626	.684
400	.5	.446	.446	.529	.446	.447	.529	.450	.449	.532	.451	.453	.537	.455	.452	.533	.462	.463	.547
400	.9	.312	.332	.426	.312	.332	.427	.313	.332	.426	.313	.334	.430	.314	.332	.426	.316	.336	.433
1,000	.5	.277	.309	.397	.276	.309	.397	.279	.310	.397	.279	.312	.399	.281	.313	.398	.284	.317	.404
1,000	.9	.194	.234	.326	.194	.234	.326	.195	.234	.326	.195	.234	.326	.195	.234	.327	.196	.235	.328

Exogenous Factor Mean Difference (κ^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.943	.739	.945	.940	.739	.946	.963	.737	.940	.946	.741	.948	.962	.733	.933	.949	.740	.947
100	.9	.707	.561	.669	.703	.560	.669	.712	.560	.668	.705	.561	.670	.714	.559	.665	.707	.560	.668
400	.5	.465	.492	.777	.469	.492	.778	.464	.493	.776	.468	.493	.780	.461	.494	.773	.469	.493	.782
400	.9	.360	.337	.490	.362	.338	.491	.359	.338	.489	.361	.338	.491	.359	.339	.486	.361	.338	.491
1,000	.5	.298	.422	.698	.299	.421	.697	.297	.422	.698	.298	.421	.695	.297	.423	.700	.298	.421	.693
1,000	.9	.231	.269	.430	.232	.269	.430	.230	.270	.430	.231	.269	.428	.229	.270	.430	.230	.270	.427

Endogenous Factor Mean Difference (α^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	1.173	.88	1.128	1.167	.886	1.134	1.185	.890	1.136	1.166	.892	1.138	1.194	.889	1.131	1.191	.902	1.152
100	.9	.754	.593	.722	.750	.592	.722	.752	.595	.722	.749	.595	.723	.753	.593	.718	.753	.595	.725
400	.5	.501	.550	.882	.503	.550	.881	.50	.551	.885	.501	.551	.885	.501	.552	.887	.504	.552	.888
400	.9	.360	.356	.544	.361	.356	.544	.359	.357	.544	.360	.357	.545	.359	.357	.544	.360	.356	.545
1,000	.5	.320	.473	.771	.319	.473	.769	.320	.472	.772	.319	.472	.769	.320	.472	.773	.320	.472	.768
1,000	.9	.229	.293	.475	.230	.293	.474	.230	.293	.474	.230	.293	.473	.230	.293	.474	.229	.293	.471

Note.

1. N: sample size per group, λ: factor loading.
2. NFI: both factor loading and intercept noninvariance, NFI_2025: both factor loading and intercept noninvariance with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 19

ANOVA Results on Precision: Factor Loading Noninvariance Only

Source	Sum of Squares	DF	Mean Square	*F*	*p*	η^2
Structural regression coefficient difference ($\gamma^R - \gamma^F$)						
Total magnitude of noninvariance (Tot_NI)	11.682	2	5.841	249.340	<.001	<.001
Percentage of noninvariant items (Per_NI)	2.117	1	2.117	9.360	<.001	<.001
Sample size	1615.107	2	8075.053	344691	<.001	.579
Loading size	2573.886	1	2573.886	109869	<.001	.092
Prior distribution	329.413	2	164.706	703.640	<.001	.012
Tot_NI * Per_NI	.509	2	.254	1.860	<.001	<.001
Tot_NI * Sample size	1.215	4	2.554	109.01	<.001	<.001
Tot_NI * Loading size	5.876	2	2.938	125.4	<.001	<.001
Tot_NI * Prior distribution	.091	4	.023	.97	.420	<.001
Per_NI * Sample size	1.406	2	.703	30	<.001	<.001
Per_NI * Factor loading size	.885	1	.885	37.76	<.001	<.001
Per_NI * Prior distribution	.058	2	.029	1.23	.292	<.001
Sample size * Loading size	1049.100	2	524.550	2239.9	<.001	.038
Sample size * Prior distribution	388.750	4	97.187	4148.53	<.001	.014
Loading size * Prior distribution	68.338	2	34.169	1458.54	<.001	.002
Error	7451.693	318082	.023			
Corrected Total	27913.982	318115				
Exogenous factor mean difference (κ^F)						
Total magnitude of noninvariance (Tot_NI)	.014	2	.007	1.96	.141	<.001
Percentage of noninvariant items (Per_NI)	.049	1	.049	13.55	<.001	<.001
Sample size	783.393	2	3915.196	1085957	<.001	.497
Loading size	2936.408	1	2936.408	814471	<.001	.186
Prior distribution	2359.409	2	1179.705	327214	<.001	.150
Tot_NI * Per_NI	.011	2	.006	1.55	.212	<.001
Tot_NI * Sample size	.001	4	<.001	.05	.996	<.001
Tot_NI * Loading size	<.001	2	<.001	< .01	.997	<.001
Tot_NI * Prior distribution	.067	4	.017	4.65	<.001	<.001
Per_NI * Sample size	.117	2	.058	16.19	<.001	<.001
Per_NI * Factor loading size	.027	1	.027	7.36	.007	<.001
Per_NI * Prior distribution	.054	2	.027	7.53	.001	<.001
Sample size * Loading size	67.417	2	33.708	9349.70	<.001	.004
Sample size * Prior distribution	1107.659	4	276.915	76807.8	<.001	.070
Loading size * Prior distribution	288.684	2	144.342	40036.1	<.001	.018
Error	1146.780	318082	.004			
Corrected Total	15769.365	318115				
Endogenous factor mean difference (α^F)						
Total magnitude of noninvariance (Tot_NI)	1.152	2	.576	43.58	<.001	<.001
Percentage of noninvariant items (Per_NI)	.546	1	.546	41.31	<.001	<.001
Sample size	1182.020	2	591.010	447115	<.001	.432
Loading size	555.523	1	555.523	419918	<.001	.203
Prior distribution	3235.421	2	1617.710	122386	<.001	.118
Tot_NI * Per_NI	.096	2	.048	3.64	.026	<.001
Tot_NI * Sample size	1.550	4	.387	29.31	<.001	<.001
Tot_NI * Loading size	.949	2	.474	35.89	<.001	<.001
Tot_NI * Prior distribution	.069	4	.017	1.31	.263	<.001
Per_NI * Sample size	.956	2	.478	36.16	<.001	<.001
Per_NI * Factor loading size	.207	1	.207	15.64	<.001	<.001
Per_NI * Prior distribution	.005	2	.002	.18	.836	<.001
Sample size * Loading size	536.913	2	268.457	20309.8	<.001	.020
Sample size * Prior distribution	1746.797	4	436.699	33038	<.001	.064
Loading size * Prior distribution	287.541	2	143.770	10876.8	<.001	.011
Error	4204.442	318082	.013			
Corrected Total	27368.056	318115				

Table 20

ANOVA Results on Precision: Intercept Noninvariance Only

Source	Sum of Squares	DF	Mean Square	*F*	*p*	η^2
Structural regression coefficient difference ($\gamma^R - \gamma^F$)						
Total magnitude of noninvariance (Tot_NI)	.292	2	.146	7.38	.001	<.001
Percentage of noninvariant items (Per_NI)	.001	1	.001	.06	.800	<.001
Sample size	1507.729	2	7535.365	380390	<.001	.597
Loading size	2208.446	1	2208.446	111484	<.001	.087
Prior distribution	337.225	2	168.612	8511.66	<.001	.013
Tot_NI * Per_NI	.594	2	.297	15	<.001	<.001
Tot_NI * Sample size	.557	4	.139	7.03	<.001	<.001
Tot_NI * Loading size	.173	2	.086	4.36	.013	<.001
Tot_NI * Prior distribution	.085	4	.021	1.07	.369	<.001
Per_NI * Sample size	.009	2	.004	.21	.807	<.001
Per_NI * Factor loading size	.002	1	.002	.11	.739	<.001
Per_NI * Prior distribution	.012	2	.006	.3	.742	<.001
Sample size * Loading size	847.240	2	423.620	21384.6	<.001	.034
Sample size * Prior distribution	371.268	4	92.817	4685.46	<.001	.015
Loading size * Prior distribution	61.313	2	3.656	1547.55	<.001	.002
Error	6303.998	318230	.020			
Corrected Total	25246.848	318263				
Exogenous factor mean difference (κ^F)						
Total magnitude of noninvariance (Tot_NI)	.009	2	.004	1.29	.276	<.001
Percentage of noninvariant items (Per_NI)	.002	1	.002	.53	.467	<.001
Sample size	7757.234	2	3878.617	1119964	<.001	.495
Loading size	2924.905	1	2924.905	844576	<.001	.187
Prior distribution	2402.201	2	1201.101	346822	<.001	.153
Tot_NI * Per_NI	.036	2	.018	5.22	.005	<.001
Tot_NI * Sample size	.017	4	.004	1.23	.297	<.001
Tot_NI * Loading size	.001	2	.001	.22	.806	<.001
Tot_NI * Prior distribution	.009	4	.002	.65	.625	<.001
Per_NI * Sample size	.002	2	.001	.23	.797	<.001
Per_NI * Factor loading size	<.001	1	<.001	.01	.917	<.001
Per_NI * Prior distribution	<.001	2	<.001	.02	.984	<.001
Sample size * Loading size	65.841	2	32.920	9505.88	<.001	.004
Sample size * Prior distribution	1024.581	4	256.145	73962.8	<.001	.065
Loading size * Prior distribution	291.050	2	145.525	4202.8	<.001	.019
Error	1102.082	318230	.003			
Corrected Total						
Endogenous factor mean difference (α^F)						
Total magnitude of noninvariance (Tot_NI)	.162	2	.081	7.11	.001	<.001
Percentage of noninvariant items (Per_NI)	.001	1	.001	.06	.808	<.001
Sample size	11487.066	2	5743.533	503133	<.001	.437
Loading size	534.425	1	534.425	467821	<.001	.203
Prior distribution	3259.973	2	1629.987	142787	<.001	.124
Tot_NI * Per_NI	.203	2	.102	8.9	.000	<.001
Tot_NI * Sample size	.255	4	.064	5.59	.000	<.001
Tot_NI * Loading size	.113	2	.057	4.95	.007	<.001
Tot_NI * Prior distribution	.044	4	.011	.96	.426	<.001
Per_NI * Sample size	.002	2	.001	.1	.906	<.001
Per_NI * Factor loading size	.002	1	.002	.18	.674	<.001
Per_NI * Prior distribution	.022	2	.011	.95	.386	<.001
Sample size * Loading size	476.017	2	238.008	20849.5	<.001	.018
Sample size * Prior distribution	1671.773	4	417.943	36611.8	<.001	.064
Loading size * Prior distribution	29.711	2	145.356	12733.1	<.001	.011
Error	3632.766	318230	.011			
Corrected Total	26308.495	318263				

Table 21

ANOVA Results on Precision: Both Factor Loading and Intercept Noninvariance

Source	Sum of Squares	DF	Mean Square	F	p	η^2
Structural regression coefficient difference ($\gamma^R - \gamma^F$)						
Total magnitude of noninvariance (Tot_NI)	11.605	2	5.802	247.7	< .0001	.000
Percentage of noninvariant items (Per_NI)	2.089	1	2.089	89.19	< .0001	.000
Sample size	16151.817	2	8075.909	344755	< .0001	.579
Loading size	2574.256	1	2574.256	109893	< .0001	.092
Prior distribution	328.898	2	164.449	702.22	< .0001	.012
Tot_NI * Per_NI	.491	2	.246	1.49	< .0001	.000
Tot_NI * Sample size	1.274	4	2.569	109.65	< .0001	.000
Tot_NI * Loading size	5.901	2	2.951	125.96	< .0001	.000
Tot_NI * Prior distribution	.097	4	.024	1.04	.387	.000
Per_NI * Sample size	1.426	2	.713	3.44	< .0001	.000
Per_NI * Factor loading size	.893	1	.893	38.12	< .0001	.000
Per_NI * Prior distribution	.060	2	.030	1.28	.278	.000
Sample size * Loading size	1048.782	2	524.391	22385.9	< .0001	.038
Sample size * Prior distribution	388.438	4	97.110	4145.54	< .0001	.014
Loading size * Prior distribution	68.268	2	34.134	1457.15	< .0001	.002
Error	745.784	318069	.023			
Corrected Total	27914.241	318102				
Exogenous factor mean difference (κ^F)						
Total magnitude of noninvariance (Tot_NI)	.015	2	.007	2.07	.126	.000
Percentage of noninvariant items (Per_NI)	.048	1	.048	13.25	.000	.000
Sample size	783.464	2	3915.232	1085919	< .0001	.497
Loading size	2936.227	1	2936.227	814385	< .0001	.186
Prior distribution	2358.861	2	1179.430	327124	< .0001	.150
Tot_NI * Per_NI	.011	2	.005	1.46	.233	.000
Tot_NI * Sample size	.001	4	.000	.05	.995	.000
Tot_NI * Loading size	.000	2	.000	0	.998	.000
Tot_NI * Prior distribution	.069	4	.017	4.78	.001	.000
Per_NI * Sample size	.119	2	.059	16.47	< .0001	.000
Per_NI * Factor loading size	.026	1	.026	7.33	.007	.000
Per_NI * Prior distribution	.053	2	.026	7.33	.001	.000
Sample size * Loading size	67.419	2	33.709	9349.56	< .0001	.004
Sample size * Prior distribution	1107.486	4	276.872	76792.4	< .0001	.070
Loading size * Prior distribution	288.637	2	144.319	40027.9	< .0001	.018
Error	1146.783	318069	.004			
Corrected Total	15769.130	318102				
Endogenous factor mean difference (α^F)						
Total magnitude of noninvariance (Tot_NI)	1.134	2	.567	42.88	< .0001	.000
Percentage of noninvariant items (Per_NI)	.534	1	.534	4.42	< .0001	.000
Sample size	11821.206	2	591.603	447155	< .0001	.432
Loading size	555.376	1	555.376	419903	< .0001	.203
Prior distribution	3233.818	2	1616.909	122324	< .0001	.118
Tot_NI * Per_NI	.094	2	.047	3.56	.029	.000
Tot_NI * Sample size	1.569	4	.392	29.68	< .0001	.000
Tot_NI * Loading size	.950	2	.475	35.94	< .0001	.000
Tot_NI * Prior distribution	.072	4	.018	1.37	.241	.000
Per_NI * Sample size	.969	2	.485	36.65	< .0001	.000
Per_NI * Factor loading size	.207	1	.207	15.68	< .0001	.000
Per_NI * Prior distribution	.005	2	.003	.19	.823	.000
Sample size * Loading size	536.854	2	268.427	20307.3	< .0001	.020
Sample size * Prior distribution	1746.280	4	436.570	33027.8	< .0001	.064
Loading size * Prior distribution	287.552	2	143.776	10877.1	< .0001	.011
Error	4204.313	318069	.013			
Corrected Total	27367.660	318102				

Bias in Posterior Mean of Structural Parameter Difference

In this study, the bias of the three structural parameter difference estimates was assessed using the four measures (i.e., ARB, AB, AAB, and AARB). Tables 22 through 33 present the four measures of bias in posterior means of the three structural parameter differences by varying magnitude of invariance, sample size, factor loading size, and prior distribution under the three types of measurement invariance conditions. It should be noted that the patterns of the four measures of bias were similar across three structural parameter differences and hence the results were averaged over these conditions.

As seen in Tables 22 through 24, all three types of noninvariance conditions generally provided acceptable ARB values of the three structural parameter difference estimates across varying magnitudes of noninvariance conditions. The ARB values of the exogenous and endogenous factor mean difference estimates were all less than .050, ranging from -.014 to .033 for exogenous factor mean differences and from -.042 to .032 for endogenous factor mean differences across different types and magnitudes of invariance conditions. For the structural regression coefficient difference estimates, the ARB values yielded less than .110 across types and magnitudes of noninvariance with one exception. The exception was observed when sample size was small, factor loading was low, and prior distribution had no variance. Under that cell, the ARB values exceeded .150 and reached up to .204. Similarly, the AB values of the three structural parameter difference estimates were acceptable across different types and magnitudes of noninvariance conditions. As seen in Tables 25 through 27, the AB values of the exogenous and endogenous factor mean difference estimates were all less than .025 while the AB values yielded less than .076 for the structural regression coefficient difference estimates. These results suggest that there was no systematic difference between sample estimates of the three structural parameter differences and corresponding population values. Further it indicates that measurement noninvariance conditions provide acceptable parameter estimates of the three structural parameter differences.

As seen in Tables 28 through 33, further examination using AARB and AAB revealed that there were substantial unsigned amounts of bias particularly when sample size was small. For example, the AARB values of structural regression coefficient difference estimates yielded less than .150 only when sample size was large and factor loading was high regardless of prior distribution. When sample size was small and factor loading was low, the AARB values of structural regression coefficient difference estimates reached up to .714. For both exogenous and endogenous factor mean differences, the AARB values yielded less than .150 when sample size was moderate and factor loading was high or sample size was large. The AAB values of structural regression coefficient difference estimates were much smaller in magnitude compared to AARB values. When sample size was moderate or large, the AAB values of all three structural parameter difference estimates were less than .15. When sample size was small and factor loading was high, the AAB values of all three structural parameter difference estimates were also less than .15.

As seen in Tables 34 through 37, a comparison of the four measures of biases obtained from the measurement invariance conditions and varying degrees of noninvariance conditions indicate that the different types and magnitudes of noninvariance do not have impact on the biases in the estimates of structural parameter differences in these simulations. Using the four measures of bias for the three structural parameter differences as the dependent variables, several factorial ANOVAs were conducted to examine the effects of the five main manipulated factors on each of the four bias measures. Five main effects of the manipulated factors and all two-way interaction effects were included in the ANOVA model. Effect size, η^2, was also calculated to examine significant main and interaction effects on the four measures of bias. Tables 38 through 49 summarize the results of the factorial ANOVA on the ARB, AB, AARB, and AAB. The results of ANOVA showed that the total magnitude of noninvariance and percentage of noninvariance items had no effects on the four bias measures for the three structural parameter difference estimates. All effect sizes of these two factors were all less than .001 across three types of noninvariance conditions. Sample size, factor loading size, prior distribution, and any two-way interactions had little effects on both ARB and AB for all three structural parameter differences. However, sample size and factor loading size were found to have large and small effects on both AARB and AAB for all three structural parameter differences. The prior distribution appeared to have little effects on AARB and AAB for structural regression coefficient differences and had small effects on AARB and AAB for exogenous and endogenous factor mean differences.

Table 22

Average Relative Bias (ARB): Factor Loading Noninvariance Only

Structural Regression Coefficient Difference ($\gamma^R - \gamma^F$)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.155	.070	.085	.178	.074	.086	.180	.079	.090	.182	.085	.094	.184	.080	.092	.204	.098	.105
100	.9	.045	.020	.030	.048	.018	.030	.048	.019	.032	.050	.018	.032	.048	.020	.034	.051	.020	.034
400	.5	.040	.020	.025	.043	.022	.032	.045	.024	.032	.044	.024	.033	.046	.024	.033	.046	.026	.036
400	.9	.015	.005	.010	.016	.006	.011	.016	.007	.010	.016	.007	.012	.016	.007	.010	.016	.008	.012
1,000	.5	.020	.030	.055	.020	.027	.057	.021	.027	.057	.020	.028	.058	.022	.028	.057	.022	.030	.060
1,000	.9	<.001	.010	.040	.004	.012	.039	.005	.012	.038	.005	.012	.040	.005	.012	.038	.005	.012	.040

Exogenous Factor Mean Difference (κ^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.020	.005	.010	.026	.006	.010	.030	.006	.010	.026	.006	.010	.033	.005	.010	.020	.004	.008
100	.9	-.010	.005	.005	-.007	.002	.007	-.005	.002	.006	-.008	.002	.007	-.004	.002	.006	-.008	.002	.006
400	.5	<.001	<.001	<.001	<.001	.002	.004	<.001	.002	.004	<.001	.002	.003	-.001	.002	.005	<.001	.001	.002
400	.9	-.010	<.001	.005	-.007	.001	.003	-.008	.001	.004	-.007	<.001	.002	-.008	.002	.004	-.007	<.001	.002
1,000	.5	.005	<.001	<.001	.006	<.001	-.002	.006	<.001	-.002	.006	-.001	-.002	.005	<.001	-.002	.005	-.002	-.003
1,000	.9	<.001	-.005	-.005	.002	<.001	-.007	.002	<.001	-.006	.002	-.001	-.007	.001	<.001	-.006	.002	-.002	-.007

Endogenous Factor Mean Difference (α^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.025	-.020	-.020	.028	-.032	-.032	.027	-.034	-.034	.024	-.035	-.034	.022	-.040	-.040	.017	-.042	-.040
100	.9	.005	<.001	.005	-.004	-.006	-.002	-.004	-.005	-.002	-.004	-.006	-.003	-.006	-.006	-.003	-.006	-.007	-.005
400	.5	.020	.010	.030	.013	.006	.026	.012	.004	.024	.012	.005	.024	.012	.002	.022	.011	.003	.023
400	.9	.005	.005	.025	.002	.007	.024	.002	.006	.022	.002	.006	.024	.002	.005	.022	.002	.006	.022
1,000	.5	<.001	-.005	-.010	-.002	-.005	-.009	-.003	-.005	-.008	-.002	-.005	-.009	-.003	-.005	-.008	-.003	-.005	-.010
1,000	.9	-.005	-.005	<.001	-.005	-.002	-.001	-.005	-.002	<.001	-.005	-.001	-.001	-.005	-.002	<.001	-.005	-.001	-.001

Note.

1. N: sample size per group, λ: factor loading.
2. NF: factor loading noninvariance only, NF_2025: factor loading noninvariance only with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 23

Average Relative Bias (ARB): Intercept Noninvariance Only

Structural Regression Coefficient Difference (γ^R - γ^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.156	.068	.079	.167	.075	.088	.190	.074	.088	.156	.070	.083	.162	.075	.087	.175	.069	.086
100	.9	.047	.018	.031	.046	.019	.032	.056	.018	.032	.042	.016	.028	.044	.018	.031	.055	.019	.032
400	.5	.041	.020	.028	.037	.018	.028	.040	.018	.027	.038	.018	.028	.043	.024	.032	.038	.018	.027
400	.9	.016	.006	.010	.013	.006	.008	.015	.004	.009	.012	.005	.010	.017	.006	.011	.016	.005	.010
1,000	.5	.017	.026	.054	.014	.024	.053	.020	.024	.053	.016	.023	.053	.019	.025	.053	.018	.025	.055
1,000	.9	.004	.011	.038	.002	.011	.040	.005	.012	.038	.003	.011	.038	.004	.011	.037	.003	.012	.038

Exogenous Factor Mean Difference (κ^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.020	.004	.008	.022	.006	.010	.014	.006	.010	.019	.006	.010	.020	.006	.010	.014	.006	.009
100	.9	-.010	.002	.007	-.008	.002	.007	-.014	.002	.007	-.012	.002	.007	-.008	.002	.006	-.013	.002	.007
400	.5	-.001	<.001	.002	.002	.002	.003	-.003	.002	.003	.001	.001	.002	.003	.001	.002	-.001	.001	.003
400	.9	-.007	<.001	.002	-.005	.001	.003	-.008	<.001	.002	-.006	.001	.002	-.006	.001	.002	-.008	.001	.003
1,000	.5	.005	<.001	-.002	.007	<.001	-.002	.004	<.001	-.002	.006	<.001	-.002	.005	-.001	-.002	.004	<.001	-.002
1,000	.9	.002	-.001	-.007	.002	<.001	-.008	.002	-.001	-.007	.002	<.001	-.006	.002	<.001	-.007	.002	<.001	-.008

Endogenous Factor Mean Difference (α^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.032	-.019	-.018	.018	-.034	-.034	.026	-.034	-.034	.006	-.034	-.034	.012	-.038	-.037	.030	-.027	-.026
100	.9	.002	<.001	.003	-.006	-.005	-.002	-.005	-.008	-.005	-.017	-.007	-.004	-.011	-.008	-.006	-.002	-.003	<.001
400	.5	.017	.009	.030	.010	.003	.024	.012	.007	.027	.012	.006	.027	.010	-.001	.020	.010	.007	.028
400	.9	.006	.008	.025	<.001	.004	.022	.001	.006	.024	<.001	.007	.024	<.001	.001	.018	-.002	.006	.024
1,000	.5	-.003	-.005	-.009	-.005	-.005	-.009	<.001	-.002	-.006	-.003	-.005	-.009	-.002	-.003	-.008	-.004	-.002	-.007
1,000	.9	-.004	-.001	-.001	-.008	-.002	-.001	-.005	<.001	<.001	-.007	-.002	-.001	-.006	-.002	<.001	-.006	<.001	.002

Note.

1. N: sample size per group, λ: factor loading.
2. NI: intercept noninvariance only, NI_2025: intercept noninvariance only with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 24

Average Relative Bias (ARB): Both Factor Loading and Intercept Noninvariance

Structural Regression Coefficient Difference (γ^R - γ^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.184	.077	.090	.180	.084	.096	.170	.083	.096	.164	.080	.091	.184	.080	.098	.188	.093	.105
100	.9	.051	.020	.034	.049	.018	.031	.043	.020	.033	.044	.018	.030	.048	.020	.034	.052	.022	.036
400	.5	.042	.019	.028	.042	.020	.030	.040	.020	.029	.040	.024	.034	.046	.024	.032	.046	.028	.037
400	.9	.015	.005	.009	.015	.004	.010	.014	.005	.008	.014	.006	.012	.016	.007	.010	.017	.008	.012
1,000	.5	.022	.028	.056	.018	.026	.055	.020	.026	.055	.019	.028	.057	.022	.028	.055	.019	.029	.059
1,000	.9	.004	.012	.040	.004	.011	.040	.005	.011	.038	.004	.012	.038	.005	.012	.038	.004	.013	.039

Exogenous Factor Mean Difference (κ^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.022	.006	.010	.020	.006	.010	.041	.006	.011	.023	.006	.011	.033	.005	.010	.026	.005	.011
100	.9	-.008	.002	.006	-.009	.002	.007	-.003	.002	.006	-.012	.002	.006	-.004	.002	.006	-.008	.002	.007
400	.5	-.002	.002	.004	.003	.002	.003	-.001	.002	.004	<.001	.001	.002	-.001	.002	.004	.001	.001	.002
400	.9	-.008	.001	.003	-.006	<.001	.002	-.008	.001	.003	-.007	<.001	.002	-.008	.002	.004	-.007	<.001	.002
1,000	.5	.005	<.001	-.002	.006	<.001	-.002	.003	-.001	-.002	.003	-.002	-.003	.005	<.001	-.002	.005	-.002	-.003
1,000	.9	.002	<.001	-.007	.002	-.001	-.007	<.001	-.001	-.007	.002	-.001	-.007	.001	<.001	-.006	.002	-.002	-.007

Endogenous Factor Mean Difference (α^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.030	-.030	-.030	.014	-.039	-.038	.010	-.04	-.041	.005	-.040	-.040	.022	-.040	-.043	-.008	-.052	-.052
100	.9	<.001	-.006	-.003	-.005	-.008	-.005	-.008	-.008	-.005	-.010	-.007	-.004	-.006	-.006	-.005	-.016	-.012	-.010
400	.5	.014	.006	.026	.014	.006	.026	.009	.002	.021	.006	.002	.022	.012	.002	.020	.006	-.002	.018
400	.9	.002	.006	.022	.003	.007	.024	<.001	.004	.020	-.002	.005	.022	.002	.005	.019	-.002	.002	.019
1,000	.5	-.003	-.005	-.008	-.003	-.004	-.008	-.004	-.004	-.008	-.005	-.004	-.008	-.003	-.005	-.006	-.008	-.006	-.010
1,000	.9	-.005	-.002	<.001	-.005	-.002	-.002	-.006	-.002	<.001	-.006	-.001	<.001	-.005	-.002	<.001	-.008	-.003	-.002

Note.

1. N: sample size per group, λ: factor loading.
2. NFI: both factor loading and intercept noninvariance, NFI_2025: both factor loading and intercept noninvariance with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 25

Average Bias (AB): Factor Loading Noninvariance Only

Structural Regression Coefficient Difference (γ^R - γ^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	-.043	-.017	-.023	-.068	-.032	-.038	-.068	-.035	-.041	-.071	-.035	-.041	-.068	-.038	-.043	-.076	-.04	-.045
100	.9	-.003	.003	<.001	-.014	-.007	-.012	-.014	-.007	-.013	-.015	-.008	-.013	-.014	-.008	-.013	-.016	-.009	-.014
400	.5	-.017	-.007	-.020	-.012	-.007	-.016	-.014	-.007	-.017	-.013	-.007	-.017	-.015	-.008	-.017	-.014	-.009	-.018
400	.9	-.003	<.001	-.010	<.001	.001	-.007	-.001	.001	-.007	<.001	.001	-.008	-.001	<.001	-.007	<.001	<.001	-.008
1,000	.5	-.007	-.013	-.030	-.006	-.013	-.027	-.007	-.013	-.028	-.007	-.013	-.028	-.008	-.014	-.028	-.007	-.014	-.029
1,000	.9	<.001	-.007	-.020	-.001	-.005	-.017	-.001	-.005	-.017	-.001	-.005	-.017	-.001	-.005	-.017	-.001	-.005	-.018

Exogenous Factor Mean Difference (κ^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	-.023	-.007	-.017	-.015	-.006	-.015	-.014	-.005	-.014	-.014	-.005	-.014	-.012	-.005	-.014	-.011	-.004	-.012
100	.9	-.023	<.001	-.007	-.014	-.001	-.007	-.014	-.001	-.006	-.013	-.001	-.006	-.012	-.001	-.006	-.012	-.001	-.005
400	.5	<.001	<.001	<.001	.004	<.001	-.002	.003	.001	<.001	.004	.001	<.001	.002	.002	.002	.004	.001	.001
400	.9	<.001	<.001	<.001	.001	<.001	-.003	<.001	<.001	-.003	.001	<.001	-.003	-.001	.001	-.001	.001	<.001	-.002
1,000	.5	.003	<.001	.007	.003	.002	.006	.002	.001	.004	.002	.002	.006	.003	.001	.002	.002	.002	.005
1,000	.9	<.001	<.001	.003	.001	.001	.002	.001	.001	.001	.001	.001	.002	.001	<.001	.001	.001	.001	.002

Endogenous Factor Mean Difference (α^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.013	.003	.013	.001	-.001	.009	.001	-.001	.007	<.001	-.001	.007	<.001	-.001	.007	-.001	-.003	.006
100	.9	.010	.003	.010	.004	.002	.006	.004	.002	.006	.004	.001	.006	.003	.001	.006	.004	.001	.006
400	.5	-.007	-.003	-.013	-.006	-.004	-.012	-.006	-.005	-.014	-.006	-.004	-.014	-.006	-.005	-.016	-.007	-.005	-.015
400	.9	-.003	-.003	-.003	-.003	<.001	-.005	-.003	-.001	-.006	-.003	-.001	-.006	-.003	-.001	-.007	-.003	-.001	-.006
1,000	.5	<.001	<.001	.007	-.002	.002	.007	-.002	.002	.009	-.002	.002	.007	-.002	.003	.010	-.002	.002	.008
1,000	.9	<.001	<.001	.010	<.001	.003	.010	<.001	.003	.010	<.001	.003	.010	<.001	.003	.011	<.001	.003	.010

Note.

1. N: sample size per group, λ: factor loading.
2. NF: factor loading noninvariance only, NF_2025: factor loading noninvariance only with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 26

Average Bias (AB): Intercept Noninvariance Only

Structural Regression Coefficient Difference (γ^R - γ^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	-.041	-.013	-.019	-.057	-.026	-.032	-.062	-.023	-.029	-.059	-.025	-.031	-.058	-.025	-.031	-.054	-.016	-.024
100	.9	-.003	.003	-.002	-.006	<.001	-.005	-.011	-.002	-.007	-.007	<.001	-.005	-.006	.001	-.004	-.004	.004	-.001
400	.5	-.014	-.008	-.017	-.011	-.006	-.015	-.012	-.006	-.015	-.013	-.008	-.017	-.016	-.011	-.021	-.015	-.009	-.019
400	.9	-.004	-.003	-.010	-.001	-.001	-.009	-.001	.001	-.007	-.001	-.001	-.009	-.005	-.004	-.012	-.005	-.004	-.011
1,000	.5	-.008	-.015	-.029	-.005	-.012	-.026	-.004	-.009	-.024	-.007	-.014	-.029	-.003	-.009	-.023	-.003	-.009	-.024
1,000	.9	-.002	-.007	-.018	-.001	-.006	-.017	.001	-.003	-.015	-.003	-.007	-.019	.002	-.003	-.014	.001	-.003	-.016

Exogenous Factor Mean Difference (κ^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	-.025	-.007	-.017	-.011	-.001	-.011	-.018	-.013	-.022	-.011	-.011	-.021	-.018	-.005	-.015	-.023	-.014	-.025
100	.9	-.022	-.003	-.009	-.013	.001	-.004	-.019	-.007	-.013	-.012	-.006	-.011	-.018	-.001	-.007	-.021	-.009	-.015
400	.5	.002	<.001	-.002	.006	.001	-.002	.008	-.001	-.004	.005	-.002	-.005	.003	<.001	-.003	.006	-.006	-.009
400	.9	-.001	<.001	-.004	.002	<.001	-.004	.004	-.002	-.005	.002	-.002	-.006	<.001	<.001	-.004	.002	-.006	-.009
1,000	.5	.003	.005	.009	.004	.003	.007	.004	.002	.006	.004	.003	.007	.006	.004	.008	.005	.002	.006
1,000	.9	.002	.003	.004	.002	.002	.003	.002	.001	.002	.003	.002	.003	.004	.003	.004	.003	.001	.002

Endogenous Factor Mean Difference (α^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.013	.006	.014	-.005	-.004	.004	.012	.003	.011	.001	.003	.012	-.001	-.002	.006	.001	.001	.010
100	.9	.010	.005	.010	<.001	-.001	.004	.009	.003	.009	.005	.003	.007	<.001	-.001	.005	.002	<.001	.005
400	.5	-.005	-.004	-.012	-.009	-.007	-.015	-.014	-.006	-.015	-.015	-.006	-.015	-.013	-.007	-.015	-.013	-.007	-.016
400	.9	-.003	<.001	-.005	-.005	-.003	-.008	-.009	-.003	-.008	-.010	-.003	-.007	-.008	-.003	-.007	-.008	-.004	-.009
1,000	.5	<.001	.001	.007	-.003	<.001	.005	-.007	-.002	.003	-.006	<.001	.006	-.001	.002	.008	<.001	.001	.006
1,000	.9	.002	.003	.009	-.001	.001	.007	-.004	-.001	.006	-.003	.001	.008	.001	.003	.010	.001	.002	.009

Note.

1. N: sample size per group, λ: factor loading.
2. NI: intercept noninvariance only, NI_2025: intercept noninvariance only with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 27

Average Bias (AB): Both Factor Loading and Intercept Noninvariance

Structural Regression Coefficient Difference (γ^R - γ^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	-.054	-.020	-.027	-.073	-.039	-.046	-.066	-.037	-.043	-.064	-.030	-.036	-.068	-.038	-.037	-.079	-.044	-.050
100	.9	-.007	<.001	-.005	-.015	-.008	-.013	-.010	-.004	-.009	-.011	-.004	-.009	-.014	-.008	-.008	-.016	-.009	-.015
400	.5	-.014	-.007	-.017	-.011	-.006	-.015	-.014	-.008	-.018	-.015	-.010	-.019	-.015	-.008	-.019	-.015	-.010	-.020
400	.9	-.003	-.002	-.009	<.001	.001	-.007	-.001	-.001	-.009	-.002	-.001	-.010	-.001	<.001	-.010	-.001	-.001	-.009
1,000	.5	-.010	-.016	-.031	-.007	-.013	-.028	-.007	-.012	-.027	-.005	-.011	-.026	-.008	-.014	-.027	-.007	-.014	-.030
1,000	.9	-.004	-.008	-.020	-.002	-.006	-.018	-.001	-.005	-.016	<.001	-.004	-.016	-.001	-.005	-.017	-.001	-.006	-.019

Exogenous Factor Mean Difference (κ^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	-.034	-.018	-.028	-.017	<.001	-.010	-.011	-.001	-.010	-.018	-.004	-.013	-.012	-.005	-.014	-.005	-.005	-.014
100	.9	-.029	-.011	-.017	-.016	.003	-.003	-.010	.003	-.002	-.016	<.001	-.005	-.012	-.001	-.005	-.007	-.001	-.006
400	.5	.003	<.001	-.002	.006	<.001	-.002	.004	.003	.002	.003	.001	<.001	.002	.002	<.001	.002	-.001	-.001
400	.9	<.001	-.001	-.004	.003	-.001	-.004	.001	.001	-.001	.001	<.001	-.003	-.001	.001	-.003	<.001	-.001	-.004
1,000	.5	.004	.003	.006	.003	.002	.005	.004	.003	.005	.004	.003	.006	.003	.001	.002	.004	.001	.005
1,000	.9	.003	.002	.003	.002	.001	.002	.003	.002	.003	.003	.002	.002	.001	<.001	.001	.002	<.001	.002

Endogenous Factor Mean Difference (α^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.012	.007	.016	.004	.002	.010	.004	.001	.010	.016	.004	.012	<.001	-.001	.007	-.001	.003	.011
100	.9	.009	.005	.010	.006	.004	.009	.005	.002	.007	.013	.004	.009	.003	.001	.005	.004	.003	.007
400	.5	-.010	-.007	-.016	-.009	-.003	-.012	-.004	-.001	-.011	-.003	-.003	-.012	-.006	-.005	-.017	.001	-.001	-.011
400	.9	-.006	-.003	-.008	-.005	<.001	-.005	-.002	.002	-.004	-.001	.001	-.004	-.003	-.001	-.008	.002	.002	-.004
1,000	.5	<.001	.001	.007	-.004	.002	.008	-.004	.002	.008	-.004	.002	.007	-.002	.003	.008	-.005	<.001	.007
1,000	.9	.002	.003	.010	-.001	.003	.009	-.001	.003	.010	-.001	.003	.009	<.001	.003	.009	-.001	.002	.009

Note.

1. N: sample size per group, λ: factor loading.
2. NFI: both factor loading and intercept noninvariance, NFI_2025: both factor loading and intercept noninvariance with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 28

Average Absolute Relative Bias (AARB): Factor Loading Noninvariance Only

Structural Regression Coefficient Difference ($\gamma^R - \gamma^F$)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.670	.555	.56	.692	.560	.565	.698	.575	.580	.698	.572	.575	.706	.585	.588	.714	.586	.587
100	.9	.425	.395	.395	.431	.398	.400	.433	.400	.402	.434	.401	.403	.436	.402	.404	.436	.404	.406
400	.5	.310	.275	.280	.300	.268	.273	.302	.271	.276	.303	.272	.277	.304	.273	.278	.308	.277	.284
400	.9	.210	.200	.200	.210	.199	.202	.211	.200	.202	.211	.200	.202	.211	.200	.202	.212	.201	.204
1,000	.5	.180	.170	.190	.190	.178	.196	.190	.179	.198	.192	.181	.200	.191	.179	.198	.196	.184	.204
1,000	.9	.125	.125	.140	.132	.127	.142	.132	.128	.141	.132	.128	.142	.132	.127	.140	.134	.129	.144

Exogenous Factor Mean Difference (κ^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.350	.230	.230	.347	.239	.241	.350	.240	.240	.348	.240	.241	.350	.240	.241	.346	.240	.242
100	.9	.260	.180	.190	.260	.194	.196	.260	.194	.196	.260	.194	.196	.260	.194	.195	.259	.194	.195
400	.5	.165	.110	.110	.170	.118	.118	.170	.118	.118	.170	.118	.118	.170	.118	.118	.170	.117	.118
400	.9	.125	.090	.090	.132	.095	.096	.132	.095	.096	.132	.095	.096	.132	.095	.096	.132	.095	.096
1,000	.5	.105	.070	.070	.106	.071	.071	.106	.071	.071	.106	.071	.071	.106	.071	.070	.106	.071	.071
1,000	.9	.080	.060	.060	.082	.058	.058	.082	.058	.058	.082	.058	.058	.082	.058	.058	.082	.058	.058

Endogenous Factor Mean Difference (α^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.385	.255	.255	.407	.276	.278	.405	.278	.281	.407	.279	.280	.407	.282	.283	.409	.284	.283
100	.9	.260	.190	.190	.271	.200	.201	.271	.200	.201	.272	.200	.201	.271	.200	.201	.272	.201	.202
400	.5	.185	.130	.135	.188	.131	.136	.188	.131	.136	.188	.131	.136	.188	.132	.138	.189	.132	.138
400	.9	.130	.095	.100	.134	.097	.101	.134	.097	.101	.134	.097	.101	.134	.097	.101	.134	.098	.101
1,000	.5	.120	.085	.085	.120	.084	.085	.120	.084	.085	.120	.085	.086	.120	.085	.086	.121	.085	.087
1,000	.9	.085	.060	.065	.086	.062	.064	.086	.062	.065	.086	.062	.065	.086	.062	.065	.086	.063	.065

Note.

1. N: sample size per group, λ: factor loading.
2. NF: factor loading noninvariance only, NF_2025: factor loading noninvariance only with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 29

Average Absolute Relative Bias (AARB): Intercept Noninvariance Only

Structural Regression Coefficient Difference (γ^R - γ^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.663	.552	.556	.690	.568	.573	.727	.582	.589	.702	.580	.585	.715	.594	.599	.690	.548	.557
100	.9	.424	.391	.393	.438	.407	.409	.449	.414	.417	.436	.400	.402	.448	.416	.418	.422	.385	.388
400	.5	.308	.271	.277	.301	.263	.269	.298	.261	.268	.302	.267	.274	.298	.262	.269	.306	.271	.277
400	.9	.213	.200	.204	.209	.196	.200	.209	.196	.199	.207	.194	.199	.208	.196	.200	.215	.200	.205
1,000	.5	.180	.167	.186	.177	.168	.186	.187	.174	.189	.181	.168	.187	.182	.166	.182	.189	.171	.187
1,000	.9	.125	.122	.136	.126	.122	.136	.132	.127	.137	.128	.125	.141	.130	.124	.136	.131	.127	.140

Exogenous factor mean difference (κ^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.348	.229	.231	.338	.226	.228	.346	.233	.236	.338	.222	.226	.350	.234	.238	.334	.226	.228
100	.9	.262	.185	.186	.254	.183	.184	.262	.188	.190	.253	.179	.181	.264	.191	.192	.252	.182	.184
400	.5	.162	.112	.112	.165	.116	.116	.170	.116	.116	.167	.110	.110	.164	.112	.112	.167	.115	.116
400	.9	.125	.091	.091	.128	.094	.094	.132	.094	.094	.128	.090	.090	.126	.091	.091	.129	.094	.094
1,000	.5	.104	.069	.069	.105	.074	.074	.104	.070	.070	.102	.070	.071	.105	.070	.070	.108	.075	.075
1,000	.9	.080	.056	.056	.082	.060	.060	.080	.057	.058	.078	.058	.058	.081	.057	.058	.084	.061	.062

Endogenous factor mean difference (α^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.383	.254	.255	.394	.258	.260	.416	.268	.270	.400	.267	.269	.403	.272	.273	.408	.267	.270
100	.9	.261	.186	.186	.262	.186	.187	.273	.192	.194	.268	.192	.193	.269	.194	.196	.270	.194	.195
400	.5	.186	.129	.135	.186	.131	.136	.189	.128	.133	.190	.130	.136	.192	.132	.137	.192	.132	.138
400	.9	.132	.096	.100	.134	.097	.100	.134	.094	.098	.135	.096	.100	.138	.099	.102	.137	.098	.103
1,000	.5	.120	.081	.083	.117	.082	.083	.114	.079	.08	.116	.081	.083	.114	.079	.081	.118	.082	.083
1,000	.9	.087	.062	.063	.083	.060	.062	.082	.058	.060	.084	.061	.064	.082	.059	.061	.084	.060	.064

Note.

1. N: sample size per group, λ: factor loading.
2. NI: intercept noninvariance only, NI_2025: intercept noninvariance only with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 30

Average Absolute Relative Bias (AARB): Factor Loading and Intercept Noninvariance

Structural Regression Coefficient Difference (γ^R - γ^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.704	.567	.572	.695	.564	.57	.702	.584	.590	.676	.552	.557	.706	.585	.590	.719	.603	.609
100	.9	.438	.400	.402	.420	.388	.391	.426	.399	.400	.415	.382	.384	.436	.402	.405	.446	.416	.418
400	.5	.311	.272	.279	.303	.271	.276	.302	.267	.273	.302	.268	.275	.304	.273	.282	.315	.282	.289
400	.9	.213	.198	.203	.211	.200	.203	.210	.198	.201	.211	.198	.202	.211	.200	.205	.217	.206	.210
1,000	.5	.187	.174	.192	.191	.176	.194	.190	.177	.196	.190	.176	.194	.191	.179	.196	.202	.186	.200
1,000	.9	.130	.126	.142	.134	.128	.143	.134	.129	.143	.134	.127	.140	.132	.127	.143	.137	.132	.144

Exogenous factor mean difference (κ^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.348	.229	.232	.349	.241	.242	.349	.230	.230	.346	.234	.236	.350	.240	.240	.348	.236	.238
100	.9	.262	.184	.186	.262	.195	.196	.26	.186	.187	.258	.190	.191	.260	.194	.194	.262	.193	.194
400	.5	.164	.108	.109	.172	.116	.116	.171	.118	.118	.173	.116	.116	.170	.118	.118	.168	.116	.116
400	.9	.127	.088	.088	.132	.094	.094	.132	.096	.096	.134	.094	.094	.132	.095	.095	.130	.094	.095
1,000	.5	.102	.070	.070	.105	.072	.071	.105	.072	.072	.106	.069	.069	.106	.071	.072	.106	.074	.074
1,000	.9	.079	.057	.058	.081	.058	.058	.082	.059	.06	.082	.056	.056	.082	.058	.059	.082	.060	.060

Endogenous factor mean difference (α^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.396	.261	.264	.411	.280	.282	.405	.276	.278	.399	.268	.270	.407	.282	.282	.405	.288	.290
100	.9	.263	.188	.190	.276	.201	.202	.271	.198	.200	.270	.194	.196	.271	.200	.201	.270	.206	.206
400	.5	.189	.129	.135	.190	.129	.134	.188	.130	.134	.186	.131	.136	.188	.132	.136	.185	.130	.134
400	.9	.132	.095	.100	.134	.096	.100	.134	.096	.099	.133	.097	.100	.134	.097	.100	.132	.096	.099
1,000	.5	.119	.081	.083	.115	.081	.082	.118	.086	.088	.115	.084	.085	.120	.085	.087	.116	.084	.085
1,000	.9	.086	.060	.062	.082	.060	.064	.085	.064	.066	.083	.062	.065	.086	.062	.066	.083	.061	.062

Note.

1. N: sample size per group, λ: factor loading.
2. NFI both factor loading and intercept: noninvariance, NFI_2025: both factor loading and intercept noninvariance with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 31

Average Absolute Bias (AAB): Factor Loading Noninvariance Only

Structural Regression Coefficient Difference (γ^R - γ^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.197	.163	.170	.202	.167	.168	.205	.171	.172	.205	.170	.171	.207	.174	.175	.210	.173	.174
100	.9	.127	.117	.117	.127	.120	.120	.128	.120	.120	.128	.120	.120	.128	.120	.121	.129	.121	.122
400	.5	.093	.080	.083	.089	.080	.082	.090	.081	.083	.090	.082	.083	.090	.082	.083	.091	.083	.085
400	.9	.063	.060	.060	.062	.059	.061	.063	.060	.061	.063	.060	.061	.063	.06	.061	.063	.060	.061
1,000	.5	.053	.053	.057	.056	.053	.058	.056	.053	.059	.057	.054	.059	.056	.054	.059	.058	.055	.060
1,000	.9	.040	.037	.043	.039	.038	.042	.039	.038	.042	.039	.038	.042	.039	.038	.042	.039	.038	.042

Exogenous factor mean difference (κ^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.197	.130	.130	.193	.137	.138	.194	.137	.137	.193	.137	.138	.194	.137	.138	.192	.137	.138
100	.9	.147	.110	.110	.146	.111	.111	.146	.111	.111	.146	.111	.111	.146	.111	.111	.145	.111	.111
400	.5	.087	.060	.060	.095	.067	.067	.095	.067	.067	.095	.067	.067	.095	.067	.067	.095	.067	.067
400	.9	.070	.050	.050	.074	.054	.054	.074	.054	.054	.074	.054	.054	.074	.054	.054	.074	.054	.054
1,000	.5	.057	.040	.040	.059	.041	.041	.059	.041	.040	.059	.041	.040	.059	.041	.040	.059	.041	.040
1,000	.9	.043	.030	.030	.046	.033	.033	.046	.033	.033	.046	.033	.033	.046	.033	.033	.046	.033	.033

Endogenous factor mean difference (α^F)

		NF_2025			NF_2075			NF_5025			NF_5075			NF_8025			NF_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.207	.140	.140	.219	.150	.152	.218	.152	.153	.220	.152	.153	.219	.152	.153	.220	.153	.154
100	.9	.143	.100	.100	.149	.110	.111	.148	.110	.111	.149	.110	.111	.148	.110	.111	.149	.111	.111
400	.5	.103	.073	.073	.102	.071	.074	.102	.072	.074	.102	.072	.074	.102	.072	.075	.102	.072	.075
400	.9	.073	.053	.057	.073	.053	.055	.073	.053	.055	.073	.053	.055	.073	.053	.055	.073	.053	.055
1,000	.5	.063	.043	.043	.065	.046	.047	.065	.046	.047	.065	.046	.047	.065	.046	.047	.065	.046	.047
1,000	.9	.047	.033	.037	.047	.034	.036	.047	.035	.036	.047	.035	.036	.047	.035	.036	.047	.035	.036

Note.

1. N: sample size per group, λ: factor loading.
2. NF: factor loading noninvariance only, NF_2025: factor loading noninvariance only with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 32

Average Absolute Bias (AAB): Intercept Noninvariance Only

Structural Regression Coefficient Difference (γ^R - γ^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.196	.165	.166	.204	.168	.171	.215	.174	.176	.207	.173	.174	.212	.177	.179	.204	.165	.167
100	.9	.126	.117	.118	.129	.122	.122	.133	.124	.125	.129	.119	.120	.133	.125	.125	.125	.115	.116
400	.5	.091	.081	.083	.089	.078	.081	.088	.078	.080	.089	.080	.082	.087	.079	.081	.091	.081	.083
400	.9	.063	.060	.061	.062	.058	.06	.062	.059	.060	.061	.058	.059	.062	.058	.060	.064	.060	.061
1,000	.5	.053	.050	.055	.052	.050	.055	.056	.052	.056	.054	.050	.056	.054	.050	.054	.056	.051	.055
1,000	.9	.037	.036	.041	.037	.037	.041	.039	.038	.041	.038	.037	.042	.039	.037	.040	.039	.038	.042

Exogenous factor mean difference (κ^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.194	.131	.132	.188	.129	.130	.193	.133	.135	.187	.127	.129	.195	.134	.135	.185	.129	.130
100	.9	.147	.105	.106	.142	.104	.105	.147	.107	.108	.142	.102	.103	.148	.109	.110	.141	.104	.105
400	.5	.090	.064	.064	.092	.066	.066	.095	.066	.066	.093	.063	.063	.091	.064	.064	.093	.066	.066
400	.9	.070	.052	.052	.071	.054	.054	.074	.053	.054	.072	.051	.052	.070	.052	.052	.072	.053	.054
1,000	.5	.058	.039	.040	.059	.042	.043	.058	.040	.040	.057	.040	.041	.058	.040	.040	.061	.043	.043
1,000	.9	.045	.032	.032	.046	.034	.034	.045	.033	.033	.044	.033	.033	.045	.032	.032	.047	.035	.035

Endogenous factor mean difference (α^F)

		NI_2025			NI_2075			NI_5025			NI_5075			NI_8025			NI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.208	.138	.139	.212	.139	.14	.224	.146	.147	.216	.145	.146	.218	.147	.148	.22	.145	.147
100	.9	.142	.102	.102	.142	.101	.102	.149	.106	.107	.146	.105	.106	.147	.106	.107	.147	.106	.107
400	.5	.101	.071	.073	.100	.071	.074	.102	.069	.072	.103	.072	.074	.104	.072	.075	.104	.073	.075
400	.9	.072	.053	.055	.072	.053	.055	.073	.051	.053	.074	.053	.055	.075	.054	.056	.074	.054	.056
1,000	.5	.065	.045	.046	.063	.044	.045	.061	.043	.044	.064	.045	.046	.062	.043	.045	.064	.045	.046
1,000	.9	.047	.034	.035	.046	.033	.034	.044	.032	.033	.046	.033	.035	.045	.032	.034	.046	.033	.035

Note.

1. N: sample size per group, λ: factor loading.
2. NI: intercept noninvariance only, NI_2025: intercept noninvariance only with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 33

Average Absolute Bias (AAB): Both Factor Loading and Intercept Noninvariance

Structural Regression Coefficient Difference (γ^R - γ^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.208	.169	.171	.205	.168	.169	.206	.173	.175	.200	.164	.165	.207	.174	.175	.214	.179	.181
100	.9	.130	.120	.120	.125	.116	.117	.126	.119	.12	.122	.114	.115	.128	.120	.121	.132	.124	.125
400	.5	.092	.081	.083	.089	.081	.083	.089	.080	.082	.089	.080	.082	.090	.082	.085	.093	.085	.087
400	.9	.063	.059	.061	.062	.060	.061	.062	.059	.061	.062	.060	.061	.063	.06	.062	.064	.061	.063
1,000	.5	.055	.052	.057	.056	.052	.057	.056	.053	.058	.056	.052	.057	.056	.054	.058	.060	.056	.060
1,000	.9	.038	.038	.042	.040	.038	.042	.039	.038	.042	.039	.038	.042	.039	.038	.042	.040	.039	.043

Exogenous factor mean difference (κ^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.194	.131	.132	.195	.138	.138	.194	.131	.132	.192	.134	.135	.194	.137	.137	.193	.135	.136
100	.9	.147	.105	.106	.147	.111	.112	.146	.106	.106	.145	.109	.109	.146	.111	.111	.147	.110	.111
400	.5	.091	.062	.062	.096	.066	.066	.095	.067	.068	.096	.066	.066	.095	.067	.067	.094	.066	.066
400	.9	.071	.050	.051	.074	.053	.054	.074	.055	.055	.075	.053	.054	.074	.054	.054	.073	.054	.054
1,000	.5	.057	.040	.04	.058	.041	.041	.058	.041	.041	.059	.039	.039	.059	.041	.041	.059	.042	.042
1,000	.9	.045	.032	.033	.045	.033	.033	.046	.033	.033	.046	.032	.032	.046	.033	.033	.046	.034	.034

Endogenous factor mean difference (α^F)

		NFI_2025			NFI_2075			NFI_5025			NFI_5075			NFI_8025			NFI_8075		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.214	.143	.144	.223	.152	.153	.218	.150	.152	.216	.145	.147	.219	.152	.152	.220	.156	.157
100	.9	.144	.103	.104	.151	.110	.111	.148	.109	.110	.148	.107	.107	.148	.110	.110	.149	.113	.113
400	.5	.103	.071	.074	.102	.071	.073	.102	.071	.073	.100	.071	.074	.102	.072	.073	.100	.071	.073
400	.9	.073	.052	.054	.073	.053	.054	.073	.053	.054	.072	.053	.055	.073	.053	.054	.071	.053	.054
1,000	.5	.065	.045	.046	.062	.045	.046	.064	.048	.048	.062	.046	.047	.065	.046	.048	.063	.045	.046
1,000	.9	.047	.033	.034	.045	.033	.035	.046	.035	.036	.046	.034	.036	.047	.035	.036	.045	.033	.034

Note.

1. N: sample size per group, λ: factor loading.
2. NFI: both factor loading and intercept noninvariance, NFI_2025: both factor loading and intercept noninvariance with 20% of total magnitudes of noninvariance and 25% of noninvariant items, etc.
3. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 34

Average Relative Bias (ARB): Measurement Invariance

		Structural Regression Coefficient Difference (γ^R - γ^F)			Exogenous Factor Mean Difference (κ^F)			Endogenous Factor Mean Difference (α^F)		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.172	.074	.085	.028	.005	.009	.030	-.030	-.030
100	.9	.047	.018	.030	-.006	.002	.007	-.003	-.005	-.002
400	.5	.042	.022	.032	< .001	.002	.003	.014	.006	.026
400	.9	.016	.006	.010	-.008	< .001	.002	.002	.007	.024
1,000	.5	.020	.027	.056	.006	< .001	-.002	-.002	-.005	-.008
1,000	.9	.004	.012	.040	.002	< .001	-.007	-.005	-.002	-.001

Note.

1. N: sample size per group, λ: factor loading.
2. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 35

Average Bias (AB): Measurement Invariance

		Structural Regression Coefficient Difference (γ^R - γ^F)			Exogenous Factor Mean Difference (κ^F)			Endogenous Factor Mean Difference (α^F)		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	-.065	-.030	-.037	-.016	-.006	-.015	.002	< .001	.009
100	.9	-.014	-.007	-.012	-.015	-.002	-.007	.005	.002	.007
400	.5	-.012	-.006	-.016	.004	< .001	-.003	-.006	-.003	-.012
400	.9	< .001	.001	-.007	.001	-.001	-.004	-.003	< .001	-.005
1,000	.5	-.006	-.012	-.027	.003	.002	.006	-.002	.002	.007
1,000	.9	< .001	-.005	-.017	.002	.001	.003	< .001	.003	.010

Note.

1. N: sample size per group, λ: factor loading.

2. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 36

Average Absolute Relative Bias (AARB): Measurement Invariance

		Structural Regression Coefficient Difference (γ^R - γ^F)			Exogenous Factor Mean Difference (κ^F)			Endogenous Factor Mean Difference (α^F)		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.683	.555	.56	.348	.240	.242	.406	.276	.277
100	.9	.430	.397	.399	.260	.194	.196	.271	.200	.200
400	.5	.298	.266	.271	.170	.118	.118	.188	.130	.136
400	.9	.209	.198	.202	.132	.095	.096	.134	.097	.101
1,000	.5	.188	.176	.194	.106	.071	.071	.120	.084	.085
1,000	.9	.132	.126	.142	.082	.058	.058	.086	.062	.064

Note.
1. N: sample size per group, λ: factor loading.
2. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation
3. All results are rounded to three decimal places.

Table 37

Average Absolute Bias (AAB): Measurement Invariance

		Structural Regression Coefficient Difference (γ^R - γ^F)			Exogenous Factor Mean Difference (κ^F)			Endogenous Factor Mean Difference (α^F)		
N	λ	P0	P1	P2	P0	P1	P2	P0	P1	P2
100	.5	.201	.165	.167	.193	.137	.138	.219	.150	.151
100	.9	.127	.119	.120	.146	.111	.111	.148	.110	.110
400	.5	.088	.079	.081	.095	.067	.067	.101	.071	.074
400	.9	.062	.059	.060	.074	.054	.054	.073	.053	.055
1,000	.5	.055	.053	.058	.059	.041	.041	.065	.046	.047
1,000	.9	.039	.038	.042	.046	.033	.033	.047	.034	.036

Note.

1. N: sample size per group, λ: factor loading.

2. P0: prior distribution with zero variation, P1: prior distribution with 10% variation, P2: prior distribution with 20% variation

Table 38

ANOVA Results on Average Relative Bias: Factor Loading Noninvariance Only

Source	Sum of Squares	DF	Mean Square	*F*	*p*	η^2
Structural regression coefficient difference ($\gamma^R - \gamma^F$)						
Total magnitude of noninvariance (Tot_NI)	1.059	2	.529	2.61	.074	< .001
Percentage of noninvariant items (Per_NI)	.378	1	.378	1.87	.172	< .001
Sample size	128.275	2	64.137	316.21	< .001	.003
Loading size	91.957	1	91.957	453.37	< .001	.002
Prior distribution	21.963	2	1.982	54.14	< .001	.001
Tot_NI * Per_NI	.094	2	.047	.23	.792	< .001
Tot_NI * Sample size	.824	4	.206	1.02	.398	< .001
Tot_NI * Loading size	.649	2	.325	1.6	.202	< .001
Tot_NI * Prior distribution	.029	4	.007	.04	.998	< .001
Per_NI * Sample size	.318	2	.159	.78	.456	< .001
Per_NI * Factor loading size	.242	1	.242	1.19	.275	< .001
Per_NI * Prior distribution	.005	2	.003	.01	.988	< .001
Sample size * Loading size	53.713	2	26.856	132.41	< .001	.001
Sample size * Prior distribution	54.662	4	13.666	67.37	< .001	.001
Loading size * Prior distribution	8.684	2	4.342	21.41	< .001	< .001
Error	4297.172	211854	.203			
Corrected Total	43329.173	211887				
Exogenous factor mean difference (κ^F)						
Total magnitude of noninvariance (Tot_NI)	.007	2	.003	.08	.926	< .001
Percentage of noninvariant items (Per_NI)	.008	1	.008	.18	.672	< .001
Sample size	2.382	2	1.191	27.22	< .001	< .001
Loading size	1.895	1	1.895	43.32	< .001	< .001
Prior distribution	.121	2	.060	1.38	.251	< .001
Tot_NI * Per_NI	.052	2	.026	.59	.555	< .001
Tot_NI * Sample size	.007	4	.002	.04	.997	< .001
Tot_NI * Loading size	.002	2	.001	.03	.975	< .001
Tot_NI * Prior distribution	.008	4	.002	.05	.996	< .001
Per_NI * Sample size	.004	2	.002	.04	.956	< .001
Per_NI * Factor loading size	< .001	1	< .001	.01	.929	< .001
Per_NI * Prior distribution	.001	2	< .001	.01	.990	< .001
Sample size * Loading size	1.248	2	.624	14.26	< .001	< .001
Sample size * Prior distribution	1.590	4	.397	9.09	< .001	< .001
Loading size * Prior distribution	1.883	2	.941	21.52	< .001	< .001
Error	9266.715	211854	.044			
Corrected Total	9275.817	211887				
Endogenous factor mean difference (α^F)						
Total magnitude of noninvariance (Tot_NI)	.534	2	.267	4.82	.008	< .001
Percentage of noninvariant items (Per_NI)	.068	1	.068	1.22	.269	< .001
Sample size	17.396	2	8.698	157.09	< .001	.001
Loading size	.699	1	.699	12.62	< .001	< .001
Prior distribution	3.952	2	1.976	35.69	< .001	< .001
Tot_NI * Per_NI	.084	2	.042	.76	.468	< .001
Tot_NI * Sample size	.397	4	.099	1.79	.127	< .001
Tot_NI * Loading size	.129	2	.065	1.17	.311	< .001
Tot_NI * Prior distribution	.006	4	.002	.03	.998	< .001
Per_NI * Sample size	.092	2	.046	.83	.434	< .001
Per_NI * Factor loading size	.005	1	.005	.09	.761	< .001
Per_NI * Prior distribution	.001	2	.000	.01	.992	< .001
Sample size * Loading size	1.726	2	.863	15.58	< .001	< .001
Sample size * Prior distribution	13.888	4	3.472	62.71	< .001	.001
Loading size * Prior distribution	7.524	2	3.762	67.94	< .001	.001
Error	1173.273	211854	.055			
Corrected Total	11776.414	211887				

Table 39

ANOVA Results on Average Relative Bias: Intercept Noninvariance Only

Source	Sum of Squares	DF	Mean Square	*F*	*p*	η^2
Structural regression coefficient difference ($\gamma^R - \gamma^F$)						
Total magnitude of noninvariance (Tot_NI)	.059	2	.030	.15	.862	<.001
Percentage of noninvariant items (Per_NI)	.059	1	.059	.3	.585	<.001
Sample size	112.132	2	56.066	281.89	<.001	.003
Loading size	7.108	1	7.108	352.49	<.001	.002
Prior distribution	2.932	2	1.466	52.62	<.001	<.001
Tot_NI * Per_NI	.251	2	.125	.63	.532	<.001
Tot_NI * Sample size	.062	4	.016	.08	.989	<.001
Tot_NI * Loading size	.037	2	.019	.09	.911	<.001
Tot_NI * Prior distribution	.092	4	.023	.12	.977	<.001
Per_NI * Sample size	.019	2	.009	.05	.954	<.001
Per_NI * Factor loading size	.014	1	.014	.07	.792	<.001
Per_NI * Prior distribution	.050	2	.025	.13	.882	<.001
Sample size * Loading size	41.978	2	2.989	105.53	<.001	.001
Sample size * Prior distribution	51.468	4	12.867	64.69	<.001	.001
Loading size * Prior distribution	7.392	2	3.696	18.58	<.001	<.001
Error	4214.263	211874	.199			
Corrected Total	42445.575	211907				
Exogenous factor mean difference (κ^F)						
Total magnitude of noninvariance (Tot_NI)	.009	2	.004	.1	.901	<.001
Percentage of noninvariant items (Per_NI)	.003	1	.003	.07	.790	<.001
Sample size	1.397	2	.698	16.31	<.0001	<.001
Loading size	1.636	1	1.636	38.21	<.0001	<.001
Prior distribution	.027	2	.014	.32	.729	<.001
Tot_NI * Per_NI	.043	2	.022	.51	.602	<.001
Tot_NI * Sample size	.009	4	.002	.05	.995	<.001
Tot_NI * Loading size	<.001	2	<.001	<.001	.999	<.001
Tot_NI * Prior distribution	.036	4	.009	.21	.933	<.001
Per_NI * Sample size	<.001	2	<.001	<.001	.998	<.001
Per_NI * Factor loading size	.003	1	.003	.08	.778	<.001
Per_NI * Prior distribution	.001	2	<.001	.01	.989	<.001
Sample size * Loading size	1.006	2	.503	11.76	<.001	<.001
Sample size * Prior distribution	1.443	4	.361	8.43	<.001	<.001
Loading size * Prior distribution	1.502	2	.751	17.55	<.001	<.001
Error	9069.121	211874	.043			
Corrected Total	9076.231	211907				
Endogenous factor mean difference (α^F)						
Total magnitude of noninvariance (Tot_NI)	.234	2	.117	2.17	.114	<.001
Percentage of noninvariant items (Per_NI)	.068	1	.068	1.27	.261	<.001
Sample size	16.959	2	8.480	157.6	<.001	.001
Loading size	.242	1	.242	4.49	.034	<.001
Prior distribution	2.574	2	1.287	23.92	<.001	<.001
Tot_NI * Per_NI	.980	2	.490	9.1	.000	<.001
Tot_NI * Sample size	.556	4	.139	2.58	.035	<.001
Tot_NI * Loading size	<.001	2	<.001	<.001	.997	<.001
Tot_NI * Prior distribution	.023	4	.006	.11	.981	<.001
Per_NI * Sample size	.022	2	.011	.21	.812	<.001
Per_NI * Factor loading size	.004	1	.004	.07	.785	<.001
Per_NI * Prior distribution	.100	2	.050	.93	.394	<.001
Sample size * Loading size	1.790	2	.895	16.63	<.001	<.001
Sample size * Prior distribution	12.231	4	3.058	56.83	<.001	.001
Loading size * Prior distribution	6.585	2	3.293	61.2	<.001	.001
Error	11399.771	211874	.054			
Corrected Total	11442.097	211907				

Table 40

ANOVA Results on Average Relative Bias: Both Factor Loading and Intercept Noninvariance

Source	Sum of Squares	DF	Mean Square	F	p	η^2
Structural regression coefficient difference (γ^R - γ^F)						
Total magnitude of noninvariance (Tot_NI)	1.029	2	.514	2.54	.079	.000
Percentage of noninvariant items (Per_NI)	.363	1	.363	1.79	.181	.000
Sample size	128.445	2	64.223	316.72	<.0001	.003
Loading size	91.858	1	91.858	453.01	<.0001	.002
Prior distribution	21.936	2	1.968	54.09	<.0001	.001
Tot_NI * Per_NI	.085	2	.043	.21	.810	.000
Tot_NI * Sample size	.846	4	.211	1.04	.383	.000
Tot_NI * Loading size	.639	2	.320	1.58	.207	.000
Tot_NI * Prior distribution	.035	4	.009	.04	.997	.000
Per_NI * Sample size	.329	2	.164	.81	.445	.000
Per_NI * Factor loading size	.237	1	.237	1.17	.280	.000
Per_NI * Prior distribution	.007	2	.003	.02	.983	.000
Sample size * Loading size	53.766	2	26.883	132.58	<.0001	.001
Sample size * Prior distribution	54.447	4	13.612	67.13	<.0001	.001
Loading size * Prior distribution	8.704	2	4.352	21.46	<.0001	.000
Error	42956.426	211844	.203			
Corrected Total	43315.308	211877				
Exogenous factor mean difference (κ^F)						
Total magnitude of noninvariance (Tot_NI)	.007	2	.003	.08	.925	.000
Percentage of noninvariant items (Per_NI)	.008	1	.008	.19	.666	.000
Sample size	2.385	2	1.193	27.27	<.0001	.000
Loading size	1.889	1	1.889	43.2	<.0001	.000
Prior distribution	.121	2	.060	1.38	.252	.000
Tot_NI * Per_NI	.052	2	.026	.6	.550	.000
Tot_NI * Sample size	.007	4	.002	.04	.997	.000
Tot_NI * Loading size	.002	2	.001	.03	.975	.000
Tot_NI * Prior distribution	.008	4	.002	.05	.996	.000
Per_NI * Sample size	.004	2	.002	.04	.958	.000
Per_NI * Factor loading size	.000	1	.000	.01	.921	.000
Per_NI * Prior distribution	.001	2	.000	.01	.989	.000
Sample size * Loading size	1.251	2	.626	14.3	<.0001	.000
Sample size * Prior distribution	1.595	4	.399	9.12	<.0001	.000
Loading size * Prior distribution	1.887	2	.943	21.57	<.0001	.000
Error	9265.173	211844	.044			
Corrected Total	9274.287	211877				
Endogenous factor mean difference (α^F)						
Total magnitude of noninvariance (Tot_NI)	.533	2	.267	4.82	.008	.000
Percentage of noninvariant items (Per_NI)	.067	1	.067	1.22	.270	.000
Sample size	17.398	2	8.699	157.13	<.0001	.001
Loading size	.700	1	.700	12.65	.000	.000
Prior distribution	3.953	2	1.977	35.7	<.0001	.000
Tot_NI * Per_NI	.084	2	.042	.76	.467	.000
Tot_NI * Sample size	.397	4	.099	1.79	.127	.000
Tot_NI * Loading size	.130	2	.065	1.18	.309	.000
Tot_NI * Prior distribution	.007	4	.002	.03	.998	.000
Per_NI * Sample size	.093	2	.046	.84	.434	.000
Per_NI * Factor loading size	.005	1	.005	.09	.758	.000
Per_NI * Prior distribution	.001	2	.000	.01	.992	.000
Sample size * Loading size	1.725	2	.862	15.58	<.0001	.000
Sample size * Prior distribution	13.886	4	3.471	62.71	<.0001	.001
Loading size * Prior distribution	7.524	2	3.762	67.96	<.0001	.001
Error	11728.059	211844	.055			
Corrected Total	11774.204	211877				

Table 41

ANOVA Results on Average Bias: Factor Loading Noninvariance Only

Source	Sum of Squares	DF	Mean Square	F	p	η^2
Structural regression coefficient difference ($\gamma^R - \gamma^F$)						
Total magnitude of noninvariance (Tot_NI)	.727	2	.363	2.84	<.001	<.001
Percentage of noninvariant items (Per_NI)	.190	1	.190	1.92	.010	<.001
Sample size	23.010	2	11.505	659.87	<.001	.004
Loading size	25.686	1	25.686	1473.25	<.001	.005
Prior distribution	4.566	2	2.283	13.94	<.001	.001
Tot_NI * Per_NI	.159	2	.079	4.55	.011	<.001
Tot_NI * Sample size	1.736	4	.434	24.89	<.001	<.001
Tot_NI * Loading size	.259	2	.129	7.41	.001	<.001
Tot_NI * Prior distribution	.005	4	.001	.07	.992	<.001
Per_NI * Sample size	.865	2	.432	24.79	<.001	<.001
Per_NI * Factor loading size	.055	1	.055	3.15	.076	<.001
Per_NI * Prior distribution	.012	2	.006	.35	.704	<.001
Sample size * Loading size	12.774	2	6.387	366.32	<.001	.002
Sample size * Prior distribution	1.595	4	2.649	151.92	<.001	.002
Loading size * Prior distribution	1.512	2	.756	43.35	<.001	<.001
Error	5545.792	318082	.017			
Corrected Total	5626.931	318115				
Exogenous factor mean difference (κ^F)						
Total magnitude of noninvariance (Tot_NI)	.059	2	.030	2.14	.117	<.001
Percentage of noninvariant items (Per_NI)	.025	1	.025	1.81	.178	<.001
Sample size	8.281	2	4.140	299.24	<.001	.002
Loading size	.002	1	.002	.16	.689	<.001
Prior distribution	.580	2	.290	2.96	<.001	<.001
Tot_NI * Per_NI	.005	2	.002	.17	.846	<.001
Tot_NI * Sample size	.244	4	.061	4.4	.002	<.001
Tot_NI * Loading size	.003	2	.001	.11	.898	<.001
Tot_NI * Prior distribution	.048	4	.012	.87	.482	<.001
Per_NI * Sample size	.037	2	.018	1.33	.264	<.001
Per_NI * Factor loading size	.001	1	.001	.06	.808	<.001
Per_NI * Prior distribution	.024	2	.012	.88	.417	<.001
Sample size * Loading size	.639	2	.319	23.07	<.001	<.001
Sample size * Prior distribution	1.919	4	.480	34.67	<.001	<.001
Loading size * Prior distribution	.062	2	.031	2.26	.105	<.001
Error	4401.254	318082				
Corrected Total	4413.140	318115				
Endogenous factor mean difference (α^F)						
Total magnitude of noninvariance (Tot_NI)	.162	2	.081	4.970	.007	<.001
Percentage of noninvariant items (Per_NI)	.066	1	.066	4.010	.045	<.001
Sample size	6.235	2	3.117	19.810	<.001	.001
Loading size	.500	1	.500	3.580	<.001	<.001
Prior distribution	.422	2	.211	12.920	<.001	<.001
Tot_NI * Per_NI	.058	2	.029	1.780	.168	<.001
Tot_NI * Sample size	.206	4	.051	3.150	.013	<.001
Tot_NI * Loading size	.023	2	.011	.700	.497	<.001
Tot_NI * Prior distribution	.012	4	.003	.180	.951	<.001
Per_NI * Sample size	.084	2	.042	2.570	.077	<.001
Per_NI * Factor loading size	.010	1	.010	.600	.440	<.001
Per_NI * Prior distribution	.008	2	.004	.250	.782	<.001
Sample size * Loading size	.195	2	.097	5.960	.003	<.001
Sample size * Prior distribution	3.093	4	.773	47.320	<.001	.001
Loading size * Prior distribution	.003	2	.002	.110	.900	<.001
Error	5196.859	318082				
Corrected Total	5207.814	318115				

Table 42

ANOVA Results on Average Bias: Intercept Noninvariance Only

Source	Sum of Squares	DF	Mean Square	*F*	*p*	η^2
Structural regression coefficient difference ($\gamma^R - \gamma^F$)						
Total magnitude of noninvariance (Tot_NI)	.072	2	.036	2.08	.125	<.001
Percentage of noninvariant items (Per_NI)	.006	1	.006	.37	.545	<.001
Sample size	6.063	2	3.031	175.02	<.001	.001
Loading size	19.181	1	19.181	1107.47	<.001	.003
Prior distribution	4.693	2	2.347	135.49	<.001	.001
Tot_NI * Per_NI	.172	2	.086	4.96	.007	<.001
Tot_NI * Sample size	.927	4	.232	13.38	<.001	<.001
Tot_NI * Loading size	.045	2	.022	1.3	.273	<.001
Tot_NI * Prior distribution	.039	4	.010	.57	.685	<.001
Per_NI * Sample size	.029	2	.014	.82	.439	<.001
Per_NI * Factor loading size	.006	1	.006	.36	.548	<.001
Per_NI * Prior distribution	.004	2	.002	.1	.902	<.001
Sample size * Loading size	1.509	2	5.254	303.38	<.001	.002
Sample size * Prior distribution	1.916	4	2.729	157.56	<.001	.002
Loading size * Prior distribution	1.520	2	.760	43.89	<.001	<.001
Error	5511.580	318231	.017			
Corrected Total	5565.864	318264				
Exogenous factor mean difference (κ^F)						
Total magnitude of noninvariance (Tot_NI)	.242	2	.121	8.94	.000	<.001
Percentage of noninvariant items (Per_NI)	.000	1	<.001	.02	.883	<.001
Sample size	14.973	2	7.486	554.16	<.001	.003
Loading size	.004	1	.004	.31	.575	<.001
Prior distribution	.420	2	.210	15.53	<.001	<.001
Tot_NI * Per_NI	.616	2	.308	22.8	<.001	<.001
Tot_NI * Sample size	.208	4	.052	3.85	.004	<.001
Tot_NI * Loading size	.002	2	.001	.06	.940	<.001
Tot_NI * Prior distribution	.309	4	.077	5.71	.000	<.001
Per_NI * Sample size	.080	2	.040	2.97	.051	<.001
Per_NI * Factor loading size	<.001	1	<.001	.01	.929	<.001
Per_NI * Prior distribution	.184	2	.092	6.8	.001	<.001
Sample size * Loading size	.693	2	.347	25.65	<.001	<.001
Sample size * Prior distribution	3.233	4	.808	59.83	<.001	.001
Loading size * Prior distribution	.122	2	.061	4.53	.011	<.001
Error	4299.175		.014			
Corrected Total	432.243					
Endogenous factor mean difference (α^F)						
Total magnitude of noninvariance (Tot_NI)	.115	2	.057	3.62	.027	<.001
Percentage of noninvariant items (Per_NI)	.236	1	.236	14.87	<.001	<.001
Sample size	9.507	2	4.753	299.51	<.001	.002
Loading size	.347	1	.347	21.88	<.001	<.001
Prior distribution	.738	2	.369	23.24	<.001	<.001
Tot_NI * Per_NI	.460	2	.230	14.49	<.001	<.001
Tot_NI * Sample size	.607	4	.152	9.57	<.001	<.001
Tot_NI * Loading size	.001	2	.000	.03	.975	<.001
Tot_NI * Prior distribution	.085	4	.021	1.33	.255	<.001
Per_NI * Sample size	.141	2	.070	4.44	.012	<.001
Per_NI * Factor loading size	.005	1	.005	.3	.583	<.001
Per_NI * Prior distribution	.040	2	.020	1.25	.288	<.001
Sample size * Loading size	.405	2	.203	12.76	<.001	<.001
Sample size * Prior distribution	2.347	4	.587	36.97	<.001	<.001
Loading size * Prior distribution	.004	2	.002	.12	.890	<.001
Error	505.530		.016			
Corrected Total	5065.425					

Table 43

ANOVA Results on Average Bias: Both Factor Loading and Intercept Noninvariance

Source	Sum of Squares	DF	Mean Square	F	p	η^2
Structural regression coefficient difference (γ^R - γ^F)						
Total magnitude of noninvariance (Tot_NI)	.760	2	.380	21.81	<.0001	.000
Percentage of noninvariant items (Per_NI)	.207	1	.207	11.91	.001	.000
Sample size	22.931	2	11.465	657.85	<.0001	.004
Loading size	25.780	1	25.780	1479.19	<.0001	.005
Prior distribution	4.656	2	2.328	133.59	<.0001	.001
Tot_NI * Per_NI	.153	2	.077	4.4	.012	.000
Tot_NI * Sample size	1.707	4	.427	24.48	<.0001	.000
Tot_NI * Loading size	.269	2	.135	7.72	.000	.000
Tot_NI * Prior distribution	.004	4	.001	.06	.993	.000
Per_NI * Sample size	.842	2	.421	24.15	<.0001	.000
Per_NI * Factor loading size	.059	1	.059	3.41	.065	.000
Per_NI * Prior distribution	.011	2	.005	.31	.736	.000
Sample size * Loading size	12.724	2	6.362	365.03	<.0001	.002
Sample size * Prior distribution	1.735	4	2.684	153.99	<.0001	.002
Loading size * Prior distribution	1.500	2	.750	43.03	<.0001	.000
Error	5543.467	318069	.017			
Corrected Total	5624.801	318102				
Exogenous factor mean difference (κ^F)						
Total magnitude of noninvariance (Tot_NI)	.070	2	.035	2.54	.079	.000
Percentage of noninvariant items (Per_NI)	.032	1	.032	2.29	.130	.000
Sample size	8.384	2	4.192	302.98	<.0001	.002
Loading size	.002	1	.002	.17	.677	.000
Prior distribution	.572	2	.286	2.66	<.0001	.000
Tot_NI * Per_NI	.006	2	.003	.23	.796	.000
Tot_NI * Sample size	.221	4	.055	3.99	.003	.000
Tot_NI * Loading size	.003	2	.001	.1	.903	.000
Tot_NI * Prior distribution	.049	4	.012	.88	.477	.000
Per_NI * Sample size	.032	2	.016	1.15	.315	.000
Per_NI * Factor loading size	.001	1	.001	.05	.816	.000
Per_NI * Prior distribution	.021	2	.011	.77	.465	.000
Sample size * Loading size	.637	2	.318	23.01	<.0001	.000
Sample size * Prior distribution	1.956	4	.489	35.35	<.0001	.000
Loading size * Prior distribution	.063	2	.032	2.29	.102	.000
Error	44.500	318069	.014			
Corrected Total	4412.501	318102				
Endogenous factor mean difference (α^F)						
Total magnitude of noninvariance (Tot_NI)	.166	2	.083	5.08	.006	.000
Percentage of noninvariant items (Per_NI)	.068	1	.068	4.16	.041	.000
Sample size	6.217	2	3.109	19.3	<.0001	.001
Loading size	.501	1	.501	3.67	<.0001	.000
Prior distribution	.414	2	.207	12.66	<.0001	.000
Tot_NI * Per_NI	.057	2	.029	1.75	.173	.000
Tot_NI * Sample size	.202	4	.051	3.09	.015	.000
Tot_NI * Loading size	.023	2	.012	.71	.492	.000
Tot_NI * Prior distribution	.011	4	.003	.17	.952	.000
Per_NI * Sample size	.083	2	.042	2.55	.078	.000
Per_NI * Factor loading size	.010	1	.010	.61	.435	.000
Per_NI * Prior distribution	.008	2	.004	.23	.794	.000
Sample size * Loading size	.194	2	.097	5.95	.003	.000
Sample size * Prior distribution	3.072	4	.768	47.02	<.0001	.001
Loading size * Prior distribution	.004	2	.002	.11	.896	.000
Error	5196.001	318069	.016			
Corrected Total	5206.912	318102				

Table 44

ANOVA Results on Average Absolute Relative Bias: Factor Loading Noninvariance Only

Source	Sum of Squares	DF	Mean Square	*F*	*p*	η^2
Structural regression coefficient difference (γ^R - γ^F)						
Total magnitude of noninvariance (Tot_NI)	2.919	2	1.460	17.09	<.001	<.001
Percentage of noninvariant items (Per_NI)	.317	1	.317	3.71	.054	<.001
Sample size	4772.849	2	2386.425	27944.7	<.001	.200
Loading size	681.093	1	681.093	7975.5	<.001	.029
Prior distribution	5.863	2	25.432	297.8	<.001	.002
Tot_NI * Per_NI	.132	2	.066	.77	.462	<.001
Tot_NI * Sample size	1.868	4	.467	5.47	<.001	<.001
Tot_NI * Loading size	.922	2	.461	5.4	.005	<.001
Tot_NI * Prior distribution	.006	4	.001	.02	1.000	<.001
Per_NI * Sample size	.202	2	.101	1.18	.307	<.001
Per_NI * Factor loading size	.022	1	.022	.26	.610	<.001
Per_NI * Prior distribution	.012	2	.006	.07	.930	<.001
Sample size * Loading size	225.441	2	112.720	1319.94	<.001	.009
Sample size * Prior distribution	52.161	4	13.040	152.7	<.001	.002
Loading size * Prior distribution	17.677	2	8.838	103.5	<.001	.001
Error	18091.944	211854	.085			
Corrected Total	23845.491	211887				
Exogenous factor mean difference (κ^F)						
Total magnitude of noninvariance (Tot_NI)	.248	2	.124	7.75	<.001	<.001
Percentage of noninvariant items (Per_NI)	.098	1	.098	6.14	.013	<.001
Sample size	1102.750	2	551.375	34528.5	<.001	.233
Loading size	63.013	1	63.013	3946.01	<.001	.013
Prior distribution	137.191	2	68.595	4295.62	<.001	.029
Tot_NI * Per_NI	.271	2	.135	8.48	<.001	<.001
Tot_NI * Sample size	.045	4	.011	.71	.587	<.001
Tot_NI * Loading size	.004	2	.002	.11	.894	<.001
Tot_NI * Prior distribution	.025	4	.006	.39	.816	<.001
Per_NI * Sample size	.015	2	.007	.46	.633	<.001
Per_NI * Factor loading size	<.001	1	<.001	0	.981	<.001
Per_NI * Prior distribution	.021	2	.010	.65	.522	<.001
Sample size * Loading size	17.597	2	8.799	55.99	<.001	.004
Sample size * Prior distribution	29.284	4	7.321	458.46	<.001	.006
Loading size * Prior distribution	6.298	2	3.149	197.19	<.001	.001
Error	3383.034	211854	.016			
Corrected Total	4728.596	211887				
Endogenous factor mean difference (α^F)						
Total magnitude of noninvariance (Tot_NI)	.798	2	.399	18.78	<.001	<.001
Percentage of noninvariant items (Per_NI)	.339	1	.339	15.98	<.001	<.001
Sample size	1323.222	2	661.611	31144.7	<.001	.213
Loading size	157.353	1	157.353	7407.28	<.001	.025
Prior distribution	155.935	2	77.968	367.26	<.001	.025
Tot_NI * Per_NI	.407	2	.204	9.59	<.001	<.001
Tot_NI * Sample size	.939	4	.235	11.05	<.001	<.001
Tot_NI * Loading size	.111	2	.056	2.62	.073	<.001
Tot_NI * Prior distribution	.015	4	.004	.17	.952	<.001
Per_NI * Sample size	.438	2	.219	1.3	<.001	<.001
Per_NI * Factor loading size	.031	1	.031	1.47	.226	<.001
Per_NI * Prior distribution	.001	2	<.001	.02	.979	<.001
Sample size * Loading size	49.615	2	24.807	1167.79	<.001	.008
Sample size * Prior distribution	43.132	4	1.783	507.6	<.001	.007
Loading size * Prior distribution	1.054	2	5.027	236.63	<.001	.002
Error	45.435	211854	.021			
Corrected Total	6224.340	211887				

Table 45

ANOVA Results on Average Absolute Relative Bias: Intercept Noninvariance Only

Source	Sum of Squares	DF	Mean Square	*F*	*p*	η^2
Structural regression coefficient difference (γ^R - γ^F)						
Total magnitude of noninvariance (Tot_NI)	1.435	2	.717	8.89	.000	<.001
Percentage of noninvariant items (Per_NI)	.378	1	.378	4.69	.030	<.001
Sample size	4971.868	2	2485.934	30789	<.001	.216
Loading size	63.789	1	63.789	7812.52	<.001	.027
Prior distribution	57.087	2	28.543	353.52	<.001	.002
Tot_NI * Per_NI	1.254	2	.627	7.77	<.001	<.001
Tot_NI * Sample size	2.792	4	.698	8.64	<.001	<.001
Tot_NI * Loading size	.314	2	.157	1.94	.143	<.001
Tot_NI * Prior distribution	.146	4	.037	.45	.771	<.001
Per_NI * Sample size	1.192	2	.596	7.38	.001	<.001
Per_NI * Factor loading size	.002	1	.002	.03	.863	<.001
Per_NI * Prior distribution	.005	2	.003	.03	.969	<.001
Sample size * Loading size	231.734	2	115.867	1435.05	<.001	.010
Sample size * Prior distribution	51.649	4	12.912	159.92	<.001	.002
Loading size * Prior distribution	19.467	2	9.733	12.55	<.001	.001
Error	17106.892	211874	.081			
Corrected Total	23064.213	211907				
Exogenous factor mean difference (κ^F)						
Total magnitude of noninvariance (Tot_NI)	.060	2	.030	1.88	.153	<.001
Percentage of noninvariant items (Per_NI)	.135	1	.135	8.4	.004	<.001
Sample size	1018.962	2	509.481	31729.1	<.001	.218
Loading size	6.259	1	6.259	3752.75	<.001	.013
Prior distribution	141.986	2	7.993	4421.26	<.001	.030
Tot_NI * Per_NI	.259	2	.130	8.07	.000	<.001
Tot_NI * Sample size	.062	4	.016	.97	.423	<.001
Tot_NI * Loading size	.001	2	<.001	.02	.984	<.001
Tot_NI * Prior distribution	.015	4	.004	.23	.920	<.001
Per_NI * Sample size	1.115	2	.558	34.72	<.001	<.001
Per_NI * Factor loading size	.001	1	.001	.07	.786	<.001
Per_NI * Prior distribution	.041	2	.021	1.28	.277	<.001
Sample size * Loading size	16.211	2	8.106	504.8	<.001	.003
Sample size * Prior distribution	35.628	4	8.907	554.7	<.001	.008
Loading size * Prior distribution	5.981	2	2.991	186.25	<.001	.001
Error	3402.107	211874	.016			
Corrected Total	4677.647	211907				
Endogenous factor mean difference (α^F)						
Total magnitude of noninvariance (Tot_NI)	.779	2	.389	18.82	<.001	<.001
Percentage of noninvariant items (Per_NI)	.020	1	.020	.97	.325	<.001
Sample size	1251.741	2	625.871	30248.6	<.001	.207
Loading size	149.567	1	149.567	7228.65	<.001	.025
Prior distribution	17.765	2	85.383	4126.58	<.001	.028
Tot_NI * Per_NI	.006	2	.003	.16	.856	<.001
Tot_NI * Sample size	1.429	4	.357	17.26	<.001	<.001
Tot_NI * Loading size	.083	2	.042	2.01	.134	<.001
Tot_NI * Prior distribution	.050	4	.013	.61	.658	<.001
Per_NI * Sample size	.030	2	.015	.73	.484	<.001
Per_NI * Factor loading size	.003	1	.003	.13	.720	<.001
Per_NI * Prior distribution	.005	2	.002	.11	.895	<.001
Sample size * Loading size	46.713	2	23.356	1128.82	<.001	.008
Sample size * Prior distribution	51.791	4	12.948	625.77	<.001	.009
Loading size * Prior distribution	11.245	2	5.623	271.75	<.001	.002
Error	4383.863	211874	.021			
Corrected Total	6061.005	211907				

Table 46

ANOVA Results on Average Absolute Relative Bias: Both Factor Loading and Intercept Noninvariance

Source	Sum of Squares	DF	Mean Square	*F*	*p*	η^2
Structural regression coefficient difference ($\gamma^R - \gamma^F$)						
Total magnitude of noninvariance (Tot_NI)	2.672	2	1.336	15.65	<.0001	.000
Percentage of noninvariant items (Per_NI)	.248	1	.248	2.9	.089	.000
Sample size	4781.408	2	239.704	27998.9	<.0001	.200
Loading size	679.909	1	679.909	7962.82	<.0001	.029
Prior distribution	51.286	2	25.643	3.32	<.0001	.002
Tot_NI * Per_NI	.115	2	.058	.68	.509	.000
Tot_NI * Sample size	1.992	4	.498	5.83	.000	.000
Tot_NI * Loading size	.868	2	.434	5.08	.006	.000
Tot_NI * Prior distribution	.041	4	.010	.12	.976	.000
Per_NI * Sample size	.164	2	.082	.96	.383	.000
Per_NI * Factor loading size	.016	1	.016	.18	.667	.000
Per_NI * Prior distribution	.036	2	.018	.21	.812	.000
Sample size * Loading size	225.994	2	112.997	1323.38	<.0001	.009
Sample size * Prior distribution	51.271	4	12.818	15.11	<.0001	.002
Loading size * Prior distribution	17.788	2	8.894	104.17	<.0001	.001
Error	18088.403	211844	.085			
Corrected Total	23848.974	211877				
Exogenous factor mean difference (κ^F)						
Total magnitude of noninvariance (Tot_NI)	.237	2	.118	7.41	.001	.000
Percentage of noninvariant items (Per_NI)	.088	1	.088	5.51	.019	.000
Sample size	1103.831	2	551.915	34566.1	<.0001	.233
Loading size	62.994	1	62.994	3945.27	<.0001	.013
Prior distribution	137.471	2	68.735	4304.84	<.0001	.029
Tot_NI * Per_NI	.283	2	.142	8.88	.000	.000
Tot_NI * Sample size	.053	4	.013	.83	.507	.000
Tot_NI * Loading size	.003	2	.002	.11	.898	.000
Tot_NI * Prior distribution	.023	4	.006	.36	.836	.000
Per_NI * Sample size	.021	2	.010	.65	.522	.000
Per_NI * Factor loading size	.000	1	.000	0	.971	.000
Per_NI * Prior distribution	.018	2	.009	.56	.572	.000
Sample size * Loading size	17.603	2	8.801	551.23	<.0001	.004
Sample size * Prior distribution	29.162	4	7.291	456.61	<.0001	.006
Loading size * Prior distribution	6.300	2	3.150	197.29	<.0001	.001
Error	3382.507	211844	.016			
Corrected Total	4729.233	211877				
Endogenous factor mean difference (α^F)						
Total magnitude of noninvariance (Tot_NI)	.775	2	.387	18.24	<.0001	.000
Percentage of noninvariant items (Per_NI)	.323	1	.323	15.19	<.0001	.000
Sample size	1324.216	2	662.108	31174.6	<.0001	.213
Loading size	157.262	1	157.262	7404.49	<.0001	.025
Prior distribution	156.177	2	78.089	3676.71	<.0001	.025
Tot_NI * Per_NI	.419	2	.209	9.86	<.0001	.000
Tot_NI * Sample size	.957	4	.239	11.27	<.0001	.000
Tot_NI * Loading size	.109	2	.054	2.56	.077	.000
Tot_NI * Prior distribution	.014	4	.004	.17	.954	.000
Per_NI * Sample size	.451	2	.225	1.61	<.0001	.000
Per_NI * Factor loading size	.030	1	.030	1.41	.236	.000
Per_NI * Prior distribution	.002	2	.001	.05	.954	.000
Sample size * Loading size	49.655	2	24.827	1168.97	<.0001	.008
Sample size * Prior distribution	43.010	4	1.752	506.26	<.0001	.007
Loading size * Prior distribution	1.067	2	5.034	237.01	<.0001	.002
Error	4499.292	211844				
Corrected Total	6224.217	211877				

Table 47

ANOVA Results on Average Absolute Bias: Factor Loading Noninvariance Only

Source	Sum of Squares	DF	Mean Square	*F*	*p*	η^2
Structural regression coefficient difference (γ^R - γ^F)						
Total magnitude of noninvariance (Tot_NI)	.360	2	.180	24.59	<.001	<.001
Percentage of noninvariant items (Per_NI)	.039	1	.039	5.30	.021	<.001
Sample size	628.170	2	314.085	42942.3	<.001	.204
Loading size	88.313	1	88.313	12074.3	<.001	.029
Prior distribution	5.368	2	2.684	366.98	<.001	.002
Tot_NI * Per_NI	.014	2	.007	.99	.372	<.001
Tot_NI * Sample size	.202	4	.051	6.92	<.001	<.001
Tot_NI * Loading size	.108	2	.054	7.39	.001	<.001
Tot_NI * Prior distribution	.001	4	.000	.03	.998	<.001
Per_NI * Sample size	.023	2	.011	1.56	.211	<.001
Per_NI * Factor loading size	.002	1	.002	.30	.585	<.001
Per_NI * Prior distribution	.002	2	.001	.14	.871	<.001
Sample size * Loading size	28.346	2	14.173	1937.73	<.001	.009
Sample size * Prior distribution	5.964	4	1.491	203.85	<.001	.002
Loading size * Prior distribution	2.195	2	1.097	15.02	<.001	.001
Error	2326.489	318082	.007			
Corrected Total	3078.476	318115				
Exogenous factor mean difference (κ^F)						
Total magnitude of noninvariance (Tot_NI)	.119	2	.059	11.81	<.001	<.001
Percentage of noninvariant items (Per_NI)	.048	1	.048	9.5	.002	<.001
Sample size	529.677	2	264.838	52612.3	<.001	.237
Loading size	29.452	1	29.452	585.9	<.001	.013
Prior distribution	57.838	2	28.919	5745.04	<.001	.026
Tot_NI * Per_NI	.129	2	.065	12.81	<.001	<.001
Tot_NI * Sample size	.021	4	.005	1.04	.384	<.001
Tot_NI * Loading size	.001	2	.001	.14	.870	<.001
Tot_NI * Prior distribution	.014	4	.003	.69	.601	<.001
Per_NI * Sample size	.008	2	.004	.75	.474	<.001
Per_NI * Factor loading size	.000	1	.000	.01	.936	<.001
Per_NI * Prior distribution	.011	2	.005	1.05	.349	<.001
Sample size * Loading size	8.243	2	4.121	818.73	<.001	.004
Sample size * Prior distribution	12.074	4	3.019	599.66	<.001	.005
Loading size * Prior distribution	2.390	2	1.195	237.36	<.001	.001
Error	1601.151	318082	.005			
Corrected Total	2235.986	318115				
Endogenous factor mean difference (α^F)						
Total magnitude of noninvariance (Tot_NI)	.304	2	.152	24.38	<.001	<.001
Percentage of noninvariant items (Per_NI)	.139	1	.139	22.25	<.001	<.001
Sample size	585.917	2	292.958	46938.9	<.001	.214
Loading size	65.997	1	65.997	10574.3	<.001	.024
Prior distribution	66.439	2	33.219	5322.55	<.001	.024
Tot_NI * Per_NI	.189	2	.094	15.12	<.001	<.001
Tot_NI * Sample size	.393	4	.098	15.74	<.001	<.001
Tot_NI * Loading size	.030	2	.015	2.39	.092	<.001
Tot_NI * Prior distribution	.006	4	.002	.25	.912	<.001
Per_NI * Sample size	.196	2	.098	15.71	<.001	<.001
Per_NI * Factor loading size	.009	1	.009	1.46	.227	<.001
Per_NI * Prior distribution	<.001	2	<.001	.04	.964	<.001
Sample size * Loading size	2.001	2	1.001	1602.34	<.001	.007
Sample size * Prior distribution	17.971	4	4.493	719.85	<.001	.007
Loading size * Prior distribution	4.009	2	2.005	321.18	<.001	.001
Error	1985.237	318082	.006			
Corrected Total	2738.934	318115				

Table 48

ANOVA Results on Average Absolute Bias: Intercept Noninvariance Only

Source	Sum of Squares	DF	Mean Square	*F*	*p*	η^2
Structural regression coefficient difference (γ^R - γ^F)						
Total magnitude of noninvariance (Tot_NI)	.207	2	.103	14.88	<.001	<.001
Percentage of noninvariant items (Per_NI)	.053	1	.053	7.57	.006	<.001
Sample size	661.244	2	33.622	47528.4	<.001	.220
Loading size	82.933	1	82.933	11921.9	<.001	.028
Prior distribution	6.144	2	3.072	441.62	<.001	.002
Tot_NI * Per_NI	.153	2	.077	11	<.001	<.001
Tot_NI * Sample size	.394	4	.099	14.16	<.001	<.001
Tot_NI * Loading size	.048	2	.024	3.43	.033	<.001
Tot_NI * Prior distribution	.021	4	.005	.76	.549	<.001
Per_NI * Sample size	.172	2	.086	12.37	<.001	<.001
Per_NI * Factor loading size	<.001	1	<.001	.04	.839	<.001
Per_NI * Prior distribution	.001	2	<.001	.05	.954	<.001
Sample size * Loading size	3.315	2	15.158	2178.99	<.001	.010
Sample size * Prior distribution	6.078	4	1.520	218.45	<.001	.002
Loading size * Prior distribution	2.423	2	1.212	174.17	<.001	.001
Error	2213.713	318231	.007			
Corrected Total	3002.077	318264				
Exogenous factor mean difference (κ^F)						
Total magnitude of noninvariance (Tot_NI)	.030	2	.015	2.93	.054	<.001
Percentage of noninvariant items (Per_NI)	.075	1	.075	14.79	<.001	<.001
Sample size	489.097	2	244.548	48150	<.001	.221
Loading size	28.139	1	28.139	554.45	<.001	.013
Prior distribution	59.955	2	29.977	5902.35	<.001	.027
Tot_NI * Per_NI	.140	2	.070	13.75	<.001	<.001
Tot_NI * Sample size	.031	4	.008	1.51	.196	<.001
Tot_NI * Loading size	<.001	2	<.001	.02	.981	<.001
Tot_NI * Prior distribution	.007	4	.002	.35	.843	<.001
Per_NI * Sample size	.600	2	.300	59.1	<.001	<.001
Per_NI * Factor loading size	.001	1	.001	.2	.657	<.001
Per_NI * Prior distribution	.028	2	.014	2.79	.061	<.001
Sample size * Loading size	7.519	2	3.759	74.19	<.001	.003
Sample size * Prior distribution	15.085	4	3.771	742.54	<.001	.007
Loading size * Prior distribution	2.271	2	1.136	223.59	<.001	.001
Error	1616.258	318231	.005			
Corrected Total	2216.899	318264				
Endogenous factor mean difference (α^F)						
Total magnitude of noninvariance (Tot_NI)	.367	2	.183	3.05	<.001	<.001
Percentage of noninvariant items (Per_NI)	.003	1	.003	.49	.486	<.001
Sample size	552.699	2	276.350	45285.1	<.001	.207
Loading size	63.507	1	63.507	10406.8	<.001	.024
Prior distribution	73.246	2	36.623	6001.37	<.001	.027
Tot_NI * Per_NI	.001	2	.001	.1	.904	<.001
Tot_NI * Sample size	.669	4	.167	27.41	<.001	<.001
Tot_NI * Loading size	.031	2	.016	2.56	.077	<.001
Tot_NI * Prior distribution	.017	4	.004	.69	.597	<.001
Per_NI * Sample size	.022	2	.011	1.78	.168	<.001
Per_NI * Factor loading size	<.001	1	<.001	.07	.796	<.001
Per_NI * Prior distribution	.004	2	.002	.32	.730	<.001
Sample size * Loading size	19.262	2	9.631	1578.18	<.001	.007
Sample size * Prior distribution	22.102	4	5.526	905.46	<.001	.008
Loading size * Prior distribution	4.595	2	2.297	376.45	<.001	.002
Error	1941.988	318231	.006			
Corrected Total	2675.521	318264				

Table 49

ANOVA Results on Average Absolute Bias: Both Factor Loading and Intercept Noninvariance

Source	Sum of Squares	DF	Mean Square	*F*	*p*	η^2
Structural regression coefficient difference ($\gamma^R - \gamma^F$)						
Total magnitude of noninvariance (Tot_NI)	.330	2	.165	22.54	<.0001	.000
Percentage of noninvariant items (Per_NI)	.030	1	.030	4.15	.042	.000
Sample size	629.259	2	314.630	43022.9	<.0001	.204
Loading size	88.168	1	88.168	12056.2	<.0001	.029
Prior distribution	5.413	2	2.706	37.09	<.0001	.002
Tot_NI * Per_NI	.012	2	.006	.82	.439	.000
Tot_NI * Sample size	.216	4	.054	7.4	<.0001	.000
Tot_NI * Loading size	.102	2	.051	6.97	.001	.000
Tot_NI * Prior distribution	.005	4	.001	.17	.955	.000
Per_NI * Sample size	.018	2	.009	1.2	.301	.000
Per_NI * Factor loading size	.001	1	.001	.2	.651	.000
Per_NI * Prior distribution	.004	2	.002	.31	.737	.000
Sample size * Loading size	28.413	2	14.207	1942.64	<.0001	.009
Sample size * Prior distribution	5.860	4	1.465	2.31	<.0001	.002
Loading size * Prior distribution	2.208	2	1.104	15.96	<.0001	.001
Error	2326.059	318069	.007			
Corrected Total	3078.945	318102				
Exogenous factor mean difference (κ^F)						
Total magnitude of noninvariance (Tot_NI)	.114	2	.057	11.34	<.0001	.000
Percentage of noninvariant items (Per_NI)	.043	1	.043	8.62	.003	.000
Sample size	53.143	2	265.071	52664.7	<.0001	.237
Loading size	29.444	1	29.444	585.04	<.0001	.013
Prior distribution	57.951	2	28.976	5756.9	<.0001	.026
Tot_NI * Per_NI	.135	2	.067	13.37	<.0001	.000
Tot_NI * Sample size	.024	4	.006	1.2	.308	.000
Tot_NI * Loading size	.001	2	.001	.13	.875	.000
Tot_NI * Prior distribution	.013	4	.003	.64	.631	.000
Per_NI * Sample size	.010	2	.005	1.01	.363	.000
Per_NI * Factor loading size	.000	1	.000	.01	.926	.000
Per_NI * Prior distribution	.009	2	.005	.93	.396	.000
Sample size * Loading size	8.245	2	4.123	819.07	<.0001	.004
Sample size * Prior distribution	12.026	4	3.006	597.31	<.0001	.005
Loading size * Prior distribution	2.391	2	1.195	237.48	<.0001	.001
Error	16.901	318069	.005			
Corrected Total	2236.237	318102				
Endogenous factor mean difference (α^F)						
Total magnitude of noninvariance (Tot_NI)	.297	2	.148	23.79	<.0001	.000
Percentage of noninvariant items (Per_NI)	.133	1	.133	21.29	<.0001	.000
Sample size	586.277	2	293.138	46977.6	<.0001	.214
Loading size	65.932	1	65.932	10566.1	<.0001	.024
Prior distribution	66.526	2	33.263	533.64	<.0001	.024
Tot_NI * Per_NI	.193	2	.097	15.48	<.0001	.000
Tot_NI * Sample size	.399	4	.100	15.98	<.0001	.000
Tot_NI * Loading size	.028	2	.014	2.28	.103	.000
Tot_NI * Prior distribution	.006	4	.001	.24	.918	.000
Per_NI * Sample size	.201	2	.100	16.1	<.0001	.000
Per_NI * Factor loading size	.008	1	.008	1.34	.247	.000
Per_NI * Prior distribution	.001	2	.000	.06	.942	.000
Sample size * Loading size	2.032	2	1.016	1605.11	<.0001	.007
Sample size * Prior distribution	17.927	4	4.482	718.23	<.0001	.007
Loading size * Prior distribution	4.019	2	2.010	322.06	<.0001	.001
Error	1984.738	318069	.006			
Corrected Total	2738.800	318102				

Chapter 5: Discussion

Most previous studies have consistently shown that group comparisons on latent constructs can be valid and meaningfully interpreted when measurement invariance assumption holds. Although some researchers have supported the concept of partial invariance, there is still no clear-cut partial invariance level which is needed to make valid group comparisons. Given this, the current study aimed to examine the extent to which measurement noninvariance affects structural parameter comparisons across populations. Particularly, this study takes a Bayesian approach to investigate the sensitivity of the posterior distribution of structural parameter difference to varying types and magnitudes of noninvariance across two populations. For this purpose, a simulation study was conducted. Data were generated from two-group two-factor models with mean structure with known types and magnitudes of noninvariance in population parameters and varied as a function of sample size, factor loading size and structural parameter difference. The generated data were analyzed using Bayesian estimation with three different prior distributions of reference indicators' parameters. In order to assess the sensitivity of noninvariance conditions, the three outcome variables were evaluated: accuracy of statistical conclusion on structural parameter difference, precision of the estimated structural parameter difference, and bias in the posterior mean of structural parameter difference. This chapter summarizes the main findings of this research, followed by implications for practice, limitations and directions for future study.

Summary of the Main Findings

Overall convergence rates across all conditions were found to be very good even when the sample size was small and factor loading was low. One exception occurred in conditions where there were large sample sizes, high factor loadings, and prior distributions with 20% variation. Further examination showed that the convergence was achieved at between 6,000 and 8,000 iterations under those problematic conditions while convergence was achieved with 4,000 or less under other conditions. One possible explanation for the relatively low convergence rates may be due to the low number of iterations (10,000) used in this study. To examine this issue, further analyses were conducted with three different iteration sizes (10,000, 20,000, and 30,000) under the problematic conditions. The results showed that with 20,000 and 30,000 iterations, the convergence rates under the problematic conditions were found to be very good, yielding convergence rates greater than 99.0%. This result indicates that more than 10,000 iterations (say 20,000 iterations) are necessary for achieving convergence when there are large sample sizes, high factor loadings, and prior distributions with 20% variation.

Overall, the findings of simulation revealed that the three outcome variables examined in this study were not sensitive to varying types and magnitudes of noninvariance across all conditions. Specifically, it seemed clear that the accuracy of statistical conclusion on the three structural parameter differences evaluated by Type I error rates and power rates was not associated with types and magnitudes of invariance. The Type I error rates for all conditions were generally close to or lower than the nominal 5% level across different types and magnitudes of noninvariance. Even in situations where a large magnitude of noninvariance exists in measurement models between two populations, the Type I error rates rarely exceeded 6.5%. Additionally, all conditions provided consistent power rates across different types and magnitudes of noninvariance when the other factors were controlled.

The results of this study particularly on the Type I error and power were not consistent with those of previous studies. For example, previous research has found that the increased level of noninvariance between two populations increased type I error rates (Beuckelaer & Swinnen, 2011). Also, the increased level of noninvariance decreased power of detecting latent mean difference (Beuckelaer & Swinnen, 2011; Kaplan & George, 1995). However, the findings of the current study did not support those of previous studies. A possible explanation for these inconsistent results may be due to different model specifications employed in the current study and previous studies. In previous studies, for instance, structural parameter differences between two groups were estimated and evaluated under model misspecification. Specifically, structural parameters were estimated under the false assumption that measurement invariance holds for all parameters across groups. Given the fact that model misspecification can lead to the increased model non-convergence, increased Type I error rates in measurement invariance testing, and inaccurate measurement and structural parameter estimates (e.g., Anderson & Gerbing, 1988; French & Finch, 2011; Jarvis, MacKenzie, & Podsakoff, 2003), the model misspecification could affect the Type I error rates for structural parameter differences across groups in previous studies. In the current study, however, the structural parameter difference was estimated and evaluated without model misspecification. In other words, for the factor loadings or intercept parameters that are noninvariant across populations, equality constraints were not imposed. It should be also noted that equality constraints were not imposed on the factor loadings and intercepts that were truly invariant in population models except for reference indicators' parameters.

In order to examine the possible effect of the model misspecification on the Type I error and power for structural parameter differences, this study performed further analyses using a dataset exhibiting large magnitudes of noninvariance (e.g., three types of noninvariance conditions with 80% of total magnitudes of noninvariance and 75% of noninvariant items). ML estimation was performed under the false assumption that measurement invariance holds for all factor loadings and intercepts across groups as did in previous studies. Mplus codes for this analysis was given in the Appendix A. Results showed that the Type I error rates were greatly inflated under model misspecification particularly when sample size was large and factor loading was high while the Type I error rates were well controlled under correct model specification (see Appendix B). As seen in Appendix C, the power rates unexpectedly fluctuated for detecting structural regression coefficient differences under model misspecification. Small sample size unexpectedly yielded higher power rates than moderate sample size under some conditions and vice versa. Unlike power rates for detecting structural regression coefficient differences, the power rates for detecting factor mean differences were close to 100% under model misspecification. This might be due to a relatively large factor mean difference in population model of this study. This result supports previous studies that power of detecting factor mean difference was greatly affected by true factor mean difference rather than model misspecification (Beuckelaer & Swinnen, 2011; Kaplan & George, 1995). These results indicate that Type I error rate for the three structural parameter differences is not sensitive to the level of measurement noninvariance and instead might be greatly sensitive to model misspecification.

The results of the simulation have also shown that the precision and bias of the estimated three structural parameter differences were not sensitive to varying types and magnitudes of noninvariance across all conditions, holding the other factors constant. Varying types and magnitudes of noninvariance did not cause any significant changes in the width of 95% credibility intervals of the three structural parameter difference estimates. Similarly, means of posterior distributions for the three structural parameter differences had little systematic biases (i.e., average relative and average biases) in all conditions and the results on the bias were also highly consistent across different types and magnitudes of noninvariance conditions.

As expected, the three outcome variables were sensitive to sample size and factor loading size. The Type I error rates decreased and empirical power rates increased as sample size and factor loading size increased. Similarly, the width of 95% credibility intervals decreased as sample size and factor loading size increased. However, sample size and factor loading size had little effects on average relative bias and average bias in the current study. Consistent with the findings of previous studies in Bayesian analyses (e.g., Lee & Song, 2004), a Bayesian inference on a basis of posterior distributions was sensitive to the choice of prior distribution. Particularly, the results of this study showed that the different choices of variance of prior distributions could lead to different conclusions or inferences on structural parameter difference across groups. That is, the more prior distributions allowed magnitudes of noninvariance of the reference indicators' parameters between two populations, the less Type I error rates were observed.

Interestingly, under large variance of prior distribution conditions, the Type I error rates were close to zero. Although low Type I error rates under large variance of prior distribution conditions seem to sound good, it should be noted that low Type I error is always associated with lower power. As expected, this study found that prior distributions with large variance conditions provided the lowest power rate. In addition, prior distribution with a relatively small variance provided higher precision level than prior distribution with a relatively large variance.

Implications for Practice

There are several implications for practice based on the results of this simulation study. First, although previous literature has strongly argued that establishment of measurement invariance is necessary for accurate and meaningful comparisons on latent constructs across groups, it may not be always true particularly when research questions focus on group comparison in structural parameters. The results of this study showed that even a large magnitude of measurement noninvariance had little impact on the accuracy of statistical conclusion and precision and bias of structural parameter difference estimates. Upon the results, it seemed clear that a lack of measurement invariance did not reduce validity of group comparison on latent constructs. That is, under the situation where measurement noninvariance existed in measurement models across groups, accurate conclusion on structural parameter comparison across groups could be obtained. Previous studies proposed that at least two indicators of the construct exhibited invariance across groups, such group comparison might be appropriate (e.g., Steenkamp & Baumgartner, 1998). However, the findings of this study support previous study indicating that if at least one indicator related to a latent construct displays invariance across groups, drawing correct conclusion for structural parameter difference between populations could be possible (Hancock et al., 2009).

Second, when measurement invariance does not hold, a correct model specification is crucial for accurate statistical conclusion on structural parameter difference across groups. A comparison of simulation results obtained from correct model specification and model misspecification with ML estimation demonstrated that model misspecification could potentially have a strong impact on inference on group difference in latent constructs. Based on the results of this study, correct inference on structural parameter difference across groups largely depended on model specification (or model misspecification) regardless of the level of measurement invariance. Therefore, researchers should carefully examine the nature of measurement models so that the possible model misspecification can be reduced. Researchers could develop models based on existing theory or prior knowledge from experts or analyses of past data. It is also recommended that researchers conduct formal tests on measurement invariance to develop correctly specified models.

Third, researchers may take either a frequentist or Bayesian approach for measurement invariance tests to examine the equality of measurement model parameters across groups. Although measurement invariance tests from the frequentist approach have been commonly used in applied studies, it should be noted that the selection of a reference indicator is of great importance in the frequentist approach to test measurement invariance. Given that researchers never know whether a chosen reference indicator is truly invariant across populations and that invariance of a reference indicator's parameters cannot be tested in the frequentist approach, researchers are more likely to select a reference indicator that is not invariant across populations. If a noninvariant reference indicator is chosen, it will likely to produce misleading measurement invariance results. Using a Bayesian approach has the potential to provide a flexible approach to address this limitation in that a reference indicator does not need to be chosen and constrained to be equal across groups in measurement invariance testing, and all factor loading and intercept parameters can be tested for measurement invariance. This may be regarded as a distinct advantage over the frequentist approach for measurement invariance testing.

As described in Muthén and Asparouhov's (2013) study, this study recommends two-step approach to examine structural parameter differences across two populations within a Bayeisan MGCFA framework particularly when researchers have no prior knowledge on the equality of measurement model parameter. First, a Bayesian approach for measurement invariance testing can be conducted to detect possible noninvariant measurement model parameters across populations. A Bayesian approach for tests of measurement invariance requires researchers to assign prior distributions between parameters representing approximate equality of measurement model parameters across groups. Although Muthén and Asparouhov have recommended using prior distributions for *differences* between parameters within the Bayesian MGCFA framework, this study recommends using prior distributions for ratios between parameters because it provides meaningful variability of parameter differences even in the unstandardized solution. By manipulating variance of the prior dictions, researchers may examine the extent to which measurement invariance is achieved. Based on the results from the first step, the second step then involves specifying models accordingly and testing structural parameter differences in the Bayesian MGCFA framework. In a situation where some parameters appeared to be noninvariant across groups, those measurement parameters should not be constrained to be equal, but should be freely estimated across groups.

It is worth noting that in the Bayesian approach, the selection of prior distribution is important for both measurement invariance testing (Lee, 2007; Muthén & Asparouhov, 2013; Steinmetz, 2013) and group comparison on structural parameters between groups. This is particularly important when researchers do not have a large sample size. In order to conduct an approximate measurement invariance tests proposed by Muthén and Asparouhov (2013), researchers need to assign prior distributions on the difference between each of the measurement parameters which allows these parameters to be estimated slightly differently. Depending on the magnitudes of variance in these prior distributions, tests of measurement invariance could lead to different results (Cieciuch et al., 2014; Muthén & Asparouhov, 2013). The results of this study demonstrated that different variances in prior distributions provide different results in terms of the accuracy of statistical conclusion and precision and bias of structural parameter difference estimates. Given the results of the current study, it is not recommended to use large variance in prior distribution for reference indicator parameters, which allows the magnitude of noninvariance of reference indicators' parameters within 20% because it turned out to yield relatively low power, low precision, and high bias for structural parameter difference estimates. Instead, this study recommends using either prior distributions that do not allow the magnitude of noninvariance or do allow the magnitude of noninvariance within 10% because these two prior distributions seem to provide better results on group comparison in structural parameters than prior distributions that allow the magnitude of noninvariance of reference indicators' parameters within 20%.

In addition to benefits of the Bayesian approach to measurement invariance over the frequentist approach, one may question whether the Bayesian approach would give benefits over the frequentist approach in testing structural parameter differences across groups. In order to examine this issue, post hoc analyses were conducted using a dataset from large magnitudes of noninvariance conditions simulated in this study. ML estimation was performed without the incorrect assumption of measurement invariance. In order to make the analyses of two approaches comparable, all measurement parameters except for reference indicators were freely estimated regardless of true equal or unequal population parameters. Appendices D, E, and F contains the Mplus codes for both ML estimation and Bayesian estimation for these analyses. The results of the Type I error rate and power from a traditional approach were presented with those from a Bayesian approach in Appendices G and H. Generally, the frequentist approach provided similar or better results than the Bayesian approach with prior distribution with zero variation. Both approach provided the Type I error rates that were close to the nominal 5% error rate in most conditions. When sample size was small or sample size was moderate with low factor loading, the frequentist approach provided higher power rates for detecting three structural parameter differences. When sample size was moderate with high factor loading or sample size was large, however, both approach provided similar power rates. Bayesian approach provide the results that are similar to those of the frequentist approach when noninformative prior distributions are assigned to model parameters (Kaplan & Depaoli, 2012). That is, because posterior distributions are constructed largely depending on the likelihood distribution of data when noninformative prior distributions are used, it is expected that results are similar in both approaches (Kaplan & Depaoli, 2012). This result implies that although researchers do not have available prior knowledge, the Bayesian approach can provide at least similar results to the frequentist approach.

Limitations and Directions for Future study

Although this was a large simulation study, like other simulation studies, this study has several limitations in its scope. First of all, this study used only one two-group two-factor CFA model to examine the impact of measurement noninvariance in group comparison of structural parameters. Although this model has been frequently used in previous studies, there are a wide variety of models in practice varying or changing the number of latent variables and indicators. Additionally, the CFA model used in this study had a simple structure in that each indicator was loaded on only one latent variable and there were no correlated errors. In practice, some cross-loadings and correlation of error variances are present in measurement models and thus the simple structure CFA can be too restrictive in reality. In order to examine whether the results of the current study can be generalized to a variety of models, various SEM models could be examined for further investigation.

In the current study, structural parameters were estimated and examined under the assumption that the true invariant reference indicators were known. While this is a common assumption in many simulation studies, it ignores possible model misspecification problems that commonly occurred in reality. Given that a noninvariant reference indicator is likely to cause inaccurate measurement and structural parameter estimates, thus, the results of this study may be generalized to situations in which the reference indicator is correctly chosen. If the simulation conditions were expanded such that the simulation included trivial misspecification of reference indicators' factor loadings and intercepts across groups, the impact of model misspecification related to the reference indicator parameters could be examined for further investigation. By doing so, the advantages and disadvantages of using prior distributions on the reference indicators could also be better examined.

Moreover, this study included only equal sample size ratio between two groups. In reality, there are many unequal sample sizes across groups. A good example of unequal sample size situation is a race/ethnicity comparison. In many cases, researchers often compare latent constructs across different race/ethnicity groups where the reference group has generally more sample size than the focal group in some group comparison analyses. Previous research found that unequal sample size yielded low power of detecting true factor mean difference than equal sample size conditions particularly when true factor mean difference was small (Kaplan & George, 1995). Given that estimates of structural parameters in a group with small sample size is more likely to have large standard errors of the estimates than those in a group with large sample size, unequal sample sizes across groups may also have an impact on statistical conclusions and inferences on structural parameter comparisons across groups. Therefore, additional factors for different sample size ratios could be investigated for further study.

There are also important simulation design factors not manipulated in the current study which deserve some attention particularly in a Bayesian analysis. This study used only noninformative prior distributions for all parameters except the reference indicators.

Although it was intended to reflect a situation where prior knowledge on model parameters was not available, the choice of prior distributions can have a great impact on the results on posterior distribution of parameters (Lee, 2007; Lee et al., 2004). Particularly, the impact of prior distributions on the estimations of posterior distributions of parameters is more substantial when sample size is small. If the simulation conditions are expanded with various levels of accurate and inaccurate prior distributions along with varying variance of prior distributions, the Bayesian approach to examine group difference in latent constructs could be thoroughly examined.

Finally, the results of the current study guide the potential study for comparing the empirical performance of the frequentist approach and the Bayesian approach for group comparison in latent constructs. Although the performance of two approaches was compared in post hoc analyses of this study, the comparison between two approaches was conducted under very limited simulation conditions (e.g., noninformative prior, large magnitude of noninvariance). By comparing the results between the Bayesian and frequentist approaches along with expanded simulation conditions, the strengths and weakness of the two approaches could be thoroughly examined and provide useful tips and insights to applied researchers.

Appendices

Appendix A: Mplus Code for ML Estimation under Model Misspecification

```
TITLE: ML estimation under model misspecification
DATA: File=Data_n3f1sd1r1.dat;
VARIABLE:
NAMES=X1-X4 Y1-Y4 G;
USEVARIABLES= X1-X4 Y1-Y4;
Grouping is G (1=reference 2=focal);

MODEL:
F1 BY X1@.5;
F1 BY X2* (efx2);
F1 BY X3* (efx3);
F1 BY X4* (efx4);
F2 BY Y1@.5;
F2 BY Y2* (efy2);
F2 BY Y3* (efy3);
F2 BY Y4* (efy4);
[X1@.8];
[X2*] (eix2);
[X3*] (eix3);
[X4*] (eix4);
[Y1@.8];
[Y2*](eiy2);
[Y3*](eiy3);
[Y4*](eiy4);
F2 ON F1*(CG1FC);");
X1-X4*;
Y1-Y4*;
F1@1;
F2@.75;
[F1@0];
[F2@0];

MODEL focal:
F1 BY X1@.5;
F1 BY X2* (efx2);
F1 BY X3* (efx3);
F1 BY X4* (efx4);
F2 BY Y1@.5;
F2 BY Y2* (efy2);
F2 BY Y3* (efy3);
F2 BY Y4* (efy4);
[X1@.8];
```

```
[X2*] (eix2);
[X3*] (eix3);
[X4*] (eix4);
[Y1@.8];
[Y2*](eiy2);
[Y3*](eiy3);
[Y4*](eiy4);
F2 ON F1*(CG2FC);
X1-X4*;
Y1-Y4*;
F1-F2*;
[F1-F2*];

MODEL CONSTRAINT:
NEW(DIFFCOE);
DIFFCOE=CG1FC-CG2FC;

OUTPUT:
```

Appendix B: Type I Error Rates (%) - ML Estimation under Model Misspecification

		Factor loading noninvariance only			Intercept noninvariance only			Both factor loading and intercept noninvariance		
N	λ	γ^R - γ^F	κ^F	α^F	γ^R - γ^F	κ^F	α^F	γ^R - γ^F	κ^F	α^F
100	.5	8.6	7.3	5.9	4.6	32.4	13.4	9.7	41.9	19.5
100	.9	13.0	8.1	5.9	3.9	14.4	7.6	56.8	61.7	56.0
400	.5	17.3	6.0	4.8	5.1	83.5	41.5	21.9	94.0	58.8
400	.9	31.3	6.4	4.9	5.5	4.4	18.3	66.8	79.0	65.0
1,000	.5	42.6	6.7	7.4	6.3	99.6	79.5	52.3	100	92.9
1,000	.9	66.2	7.6	6.7	5.5	74.2	37.0	85.6	96.4	78.2

Note.
Data from three types of noninvariance conditions with 80% of total magnitudes of noninvariance and 75% of noninvariant items were used for these analyses.

Appendix C: Power Rates (%) - ML Estimation under Model Misspecification

Power Rates (%) under Model Misspecification When Focal Group Had Higher Structural Parameter Values

		Factor loading noninvariance only			Intercept noninvariance only			Both factor loading and intercept noninvariance		
N	λ	$\gamma^R - \gamma^F$	κ^F	α^F	$\gamma^R - \gamma^F$	κ^F	α^F	$\gamma^R - \gamma^F$	κ^F	α^F
100	.5	26.3	96.6	95.0	41.4	100	89.8	23.5	100	82.8
100	.9	79.2	100	100	60.6	100	99.6	60.1	100	98.6
400	.5	99.0	100	100	93.3	100	100	74.3	100	100
400	.9	36.4	99.4	99.3	99.2	100	100	95.1	100	100
1,000	.5	91.4	100	100	99.9	100	100	97.5	100	100
1,000	.9	100	100	100	100	100	100	100	100	100

Note.
Data from three types of noninvariance with 80% of total magnitudes of noninvariance and 75% of noninvariant items were used for these analyses.

Power Rates (%) under Model Misspecification When Focal Group Had Lower Structural Parameter Values

		Factor loading noninvariance only			Intercept noninvariance only			Both factor loading and intercept noninvariance		
N	λ	$\gamma^R - \gamma^F$	κ^F	α^F	$\gamma^R - \gamma^F$	κ^F	α^F	$\gamma^R - \gamma^F$	κ^F	α^F
100	.5	59.2	97.2	84.3	39.7	70.8	99.5	59.5	56.5	99.3
100	.9	99.7	100	100	56.0	96.2	99.9	93.0	96.2	99.9
400	.5	100	100	100	93.7	100	100	99.8	98.5	100
400	.9	86	99.2	97.4	99.1	100	100	100	100	100
1,000	.5	100	100	100	100	100	100	100	100	100
1,000	.9	100	100	100	100	100	100	100	100	100

Note.
Data from three types of noninvariance with 80% of total magnitudes of noninvariance and 75% of noninvariant items were used for these analyses.

Appendix D: Mplus Code for ML Estimation under Correct Model Specification

```
TITLE: ML estimation under correct model specification
DATA: File=Data_n3f1sd1r1.dat;
VARIABLE:
NAMES=X1-X4 Y1-Y4 G;
USEVARIABLES= X1-X4 Y1-Y4;
Grouping is G (1=reference 2=focal);

MODEL:
F1 BY X1@.5;
F1 BY X2-X4*;
F2 BY Y1@.5;
F2 BY Y2-Y4*;
[X1@.8];
[X2-X4*];
[Y1@.8];
[Y2-Y4*];
F2 ON F1*(CG1FC);
X1-X4*;
Y1-Y4*;
F1@1;
F2@.75;
[F1-F2@0];

MODEL focal:
F1 BY X1@.5;
F1 BY X2-X4*;
F2 BY Y1@.5;
F2 BY Y2-Y4*;
[X1@.8];
[X2-X4*];
[Y1@.8];
[Y2-Y4*];
F2 ON F1*(CG2FC);
X1-X4*;
Y1-Y4*;
F1-F2*;
[F1-F2*];

MODEL CONSTRAINT:
NEW(DIFFCOE);
DIFFCOE=CG1FC-CG2FC;

OUTPUT:
```

Appendix E: Mplus Code for Bayesian Estimation with Zero Variation under Correct Model Specification

```
TITLE: Bayesian estimation with zero variation under correct model specification
DATA: File= Data_n3f1sd1r1.dat;
VARIABLE:
NAMES=X1-X4 Y1-Y4 G;
USEVARIABLES= X1-X4 Y1-Y4;
CLASSES=CG(2);
KNOWNCLASS=CG(G=1 G=2);

ANALYSIS:
TYPE=mixture;
ESTIMATOR=bayes;
PROCESSORS=2;
POINT=mean;
CHAIN=2;
FBITERATIONS=10000;

MODEL:
"%OVERALL%
F1 BY X1-X4*;
F2 BY Y1-Y4*;
[X1-X4*];
[Y1-Y4*];
X1-X4*;
Y1-Y4*;
F1-F2*;
F2 ON F1*;
[F1-F2*];

%CG#1%
F1 BY X1* (XL1);
F1 BY X2-X4*;
F2 BY Y1* (YL1);
F2 BY Y2-Y4*;
[X1*](XI1);
[X2-X4*];
[Y1*](YI1);
[Y2-Y4*];
F2 ON F1*(CG1FC);
X1-X4*;
Y1-Y4*;
F1@1;
F2@.75;
[F1-F2@0];
```

```
%CG#2%
F1 BY X1* (XL1);
F1 BY X2-X4*;
F2 BY Y1* (YL1);
F2 BY Y2-Y4*;
[X1*](XI1);
[X2-X4*];
[Y1*](YI1);
[Y2-Y4*];
F2 ON F1*(CG2FC);
X1-X4*;
Y1-Y4*;
F1-F2*;
[F1-F1*];

MODEL CONSTRAINT:
NEW(DIFFCOE);
DIFFCOE=CG1FC-CG2FC;

OUTPUT:
tech8;
```

Appendix F: Mplus Code for Bayesian Estimation with 10% Variation under Correct Model Specification

```
TITLE: Bayesian estimation with 10% variation under correct model specification
DATA: File= Data_n3f1sd1r1.dat;
VARIABLE:
NAMES=X1-X4 Y1-Y4 G;
USEVARIABLES= X1-X4 Y1-Y4;
CLASSES=CG(2);
KNOWNCLASS=CG(G=1 G=2);

ANALYSIS:
TYPE=mixture;
ESTIMATOR=bayes;
PROCESSORS=2;
POINT=mean;
CHAIN=2;
FBITERATIONS=10000;

MODEL:
"%OVERALL%
F1 BY X1-X4*;
F2 BY Y1-Y4*;
[X1-X4*];
[Y1-Y4*];
X1-X4*;
Y1-Y4*;
F1-F2*;
F2 ON F1*;
[F1-F2*];

%CG#1%
F1 BY X1 @.5;
F1 BY X2-X4*;
F2 BY Y1 @.5;
F2 BY Y2-Y4*;
[X1 @.8];
[X2-X4*];
[Y1 @.8];
[Y2-Y4*];
F2 ON F1*(CG1FC);
X1-X4*;
Y1-Y4*;
F1@1;
F2@.75;
[F1-F2@0];
```

```
%CG#2%
F1 BY X1* (G2XL1);
F1 BY X2-X4*;
F2 BY Y1* (G2YL1);
F2 BY Y2-Y4*;
[X1*](G2XI1);
[X2-X4*];
[Y1*](G2YI1);
[Y2-Y4*];
F2 ON F1*(CG2FC);
X1-X4*;
Y1-Y4*;
F1-F2*;
[F1-F1*];

MODEL CONSTRAINT:
NEW(DIFFCOE);
DIFFCOE=CG1FC-CG2FC;

MODEL PRIORS:
G2XL1~N(.5, .001);
G2YL1~N(.5, .001);
G2XI1~N(.8, .002);
G2YI1~N(.8, .002);

OUTPUT:
tech8;
```

Appendix G: Type I Error Rates (%) from ML Estimation and Bayesian Estimation under Correct Model Specification

		Maximum Likelihood Estimation								
		Factor loading noninvariance only			Intercept noninvariance only			Both factor loading and intercept noninvariance		
N	λ	γ^R - γ^F	κ^F	α^F	γ^R - γ^F	κ^F	α^F	γ^R - γ^F	κ^F	α^F
100	.5	4.3	6.1	6.7	4.6	5.4	5.4	4.3	6.1	6.7
100	.9	4.7	6.1	7.5	4.3	5.4	5.6	4.7	6.1	7.5
400	.5	4.7	5.5	4.5	5.3	5.7	6.3	4.7	5.5	4.5
400	.9	5.3	5.5	4.9	5.0	5.6	6.0	5.3	5.5	4.9
1,000	.5	5.6	6.2	6.2	5.5	6.1	4.9	5.6	6.2	6.2
1,000	.9	4.7	6.2	6.4	5.7	6.1	4.6	4.7	6.2	6.4
		Bayesian Estimation with Zero Variation								
		Factor loading noninvariance only			Intercept noninvariance only			Both factor loading and intercept noninvariance		
N	λ	γ^R - γ^F	κ^F	α^F	γ^R - γ^F	κ^F	α^F	γ^R - γ^F	κ^F	α^F
100	.5	4.0	4.3	3.8	4.0	4.8	3.5	5.0	4.9	3.9
100	.9	3.7	4.3	3.9	4.0	4.9	4.0	4.4	4.7	3.4
400	.5	3.8	4.6	5.4	5.6	4.7	5.8	3.4	4.0	4.5
400	.9	3.7	4.2	5.2	4.8	4.7	5.3	3.9	3.6	4.1
1,000	.5	3.8	4.0	4.6	4.3	5.2	4.3	5.7	4.4	3.1
1,000	.9	4.0	4.1	4.4	5.0	5.1	4.7	5.6	4.9	3.3
		Bayesian Estimation with 10% Variation								
		Factor loading noninvariance only			Intercept noninvariance only			Both factor loading and intercept noninvariance		
N	λ	γ^R - γ^F	κ^F	α^F	γ^R - γ^F	κ^F	α^F	γ^R - γ^F	κ^F	α^F
100	.5	3.3	2.6	2.3	3.6	2.6	1.2	4.4	3.0	1.8
100	.9	3.4	3.8	3.2	4.0	3.7	2.3	4.2	4.3	2.7
400	.5	2.5	.1	.1	2.9	.3	.2	2.4	.1	< .1
400	.9	2.3	1.2	.6	2.7	1.3	1.3	2.0	1.1	.8
1,000	.5	1.7	< .1	< .1	1.7	.1	< .1	2.3	< .1	< .1
1,000	.9	.9	.1	.1	1.3	.4	< .1	2.0	.1	.1
		Bayesian Estimation with 20% Variation								
		Factor loading noninvariance only			Intercept noninvariance only			Both factor loading and intercept noninvariance		
N	λ	γ^R - γ^F	κ^F	α^F	γ^R - γ^F	κ^F	α^F	γ^R - γ^F	κ^F	α^F
100	.5	2.7	.7	.4	3.0	.5	< .1	3.9	.5	.3
100	.9	2.0	1.5	1.0	2.4	1.8	.7	2.3	1.7	.8
400	.5	1.6	< .1	< .1	1.2	< .1	< .1	1.4	< .1	< .1
400	.9	1.0	< .1	< .1	1.1	< .1	< .1	.6	< .1	< .1
1,000	.5	.4	< .1	< .1	.4	< .1	< .1	.8	< .1	< .1
1,000	.9	.1	< .1	< .1	.3	< .1	< .1	.3	< .1	< .1

Note.
Data from three types of noninvariance conditions with 80% of total magnitudes of noninvariance and 75% of noninvariant items were used for these analyses.

Appendix H: Power Rates (%) from ML Estimation and Bayesian Estimation under Correct Model Specification

		Maximum Likelihood Estimation								
		Factor loading noninvariance only			Intercept noninvariance only			Both factor loading and intercept noninvariance		
N	λ	$\gamma^R - \gamma^F$	κ^F	α^F	$\gamma^R - \gamma^F$	κ^F	α^F	$\gamma^R - \gamma^F$	κ^F	α^F
100	.5	31.4	93.1	82.4	33.1	94.1	83.8	31.4	93.1	82.4
100	.9	52.5	99.3	97.4	54.4	99.1	98.6	52.5	99.3	97.4
400	.5	84.6	100	100	86.5	100	100	84.6	100	100
400	.9	98.1	100	100	98.5	100	100	98.1	100	100
1,000	.5	99.5	100	100	99.6	100	100	99.5	100	100
1,000	.9	100	100	100	100	100	100	100	100	100
		Bayesian Estimation with zero variation								
		Factor loading noninvariance only			Intercept noninvariance only			Both factor loading and intercept noninvariance		
N	λ	$\gamma^R - \gamma^F$	κ^F	α^F	$\gamma^R - \gamma^F$	κ^F	α^F	$\gamma^R - \gamma^F$	κ^F	α^F
100	.5	27.3	72.8	52.3	30.1	73.3	55.1	27.4	73.1	49.6
100	.9	48.3	89.1	83.5	50.1	90.4	84.5	47.7	89.1	83.4
400	.5	79.5	100	99.0	81.5	99.8	99.0	78.3	100	98.8
400	.9	97.6	100	100	97.6	100	100	97.5	100	100
1,000	.5	98.8	100	100	99.0	100	100	98.4	100	100
1,000	.9	100	100	100	100	100	100	100	100	100
		Bayesian Estimation with 10% variation								
		Factor loading noninvariance only			Intercept noninvariance only			Both factor loading and intercept noninvariance		
N	λ	$\gamma^R - \gamma^F$	κ^F	α^F	$\gamma^R - \gamma^F$	κ^F	α^F	$\gamma^R - \gamma^F$	κ^F	α^F
100	.5	29.3	88.5	63.9	32.1	90.3	68.0	30.3	88.5	63.4
100	.9	49.7	98.6	94.5	51.5	98.6	96.2	49.2	98.5	94.8
400	.5	78.2	100	99.5	80.5	100	99.4	77.6	100	99.4
400	.9	96.4	100	100	97.0	100	100	96.4	100	100
1,000	.5	98.6	100	100	99.0	100	100	98.2	100	100
1,000	.9	100	100	100	100	100	100	100	100	100
		Bayesian Estimation with 20% variation								
		Factor loading noninvariance only			Intercept noninvariance only			Both factor loading and intercept noninvariance		
N	λ	$\gamma^R - \gamma^F$	κ^F	α^F	$\gamma^R - \gamma^F$	κ^F	α^F	$\gamma^R - \gamma^F$	κ^F	α^F
100	.5	26.9	73.0	45.4	29.3	75.4	49.6	27.5	74.6	44.1
100	.9	45.2	96.4	90.0	45.8	97.0	90.8	44.7	96.2	89.5
400	.5	68.3	99.2	93.0	69.8	99.1	94.0	67.7	99.2	92.9
400	.9	86.6	100	100	88.1	100	100	86.3	100	100
1,000	.5	94.7	100	98.6	96.6	100	99.2	95.0	100	99.0
1,000	.9	99.8	100	100	99.7	100	100	99.8	100	100

Note.
Data from three types of noninvariance conditions with 80% of total magnitudes of noninvariance and 75% of noninvariant items were used for these analyses.

References

Allen, D. (1999). Desire to finish college: An empirical link between motivation and persistence. *Research in Higher Education*, *40*, 461–485.

Anderson, C. B., Hughes, S. O., Fisher, J. O., & Nicklas, T. A. (2005). Cross-cultural equivalence of feeding beliefs and practices: The psychometric properties of the child feeding questionnaire among Blacks and Hispanics. *Preventive Medicine*, *41*, 521–531.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological bulletin*, *103*, 411–423.

Bandalos, D. L., & Leite, W. (2013). Use of Monte Carlo studies in structural equation modeling research. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed.) (pp. 625–666). Charlotte, NC: Information Age Publishing, Inc.

Bernardinelli, L., & Montomoli, C. (1992). Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Statistics in medicine, 11*, 983–1007.

Beuckelaer, A. D., & Swinnen, G. (2011). Biased latent variable mean comparisons due to measurement non-invariance: A simulation study. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis: methods and applications* (pp. 117–148). New York, NY: Routledge.

Bolt, D. M. (2000). A SIBTEST approach to testing DIF hypotheses using experimentally designed test items. *Journal of Educational Measurement*, *37*, 307–327.

Borsboom, D. (2006). When does measurement invariance matter? *Medical Care, 44*, S176–S181.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456–466.

Carriquiry, A., & Pawlovich, M. D. (2004). From empirical Bayes to full Bayes: Methods for analyzing traffic safety data. *White Paper, Iowa State University*. http://www.dot.state.ia.us/crashanalysis/pdfs/eb_fb_comparison_whitepaper_october2004.pdf (accessed December 7, 2014).

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*, 464–504.

Chen, Y. J., & Tang, T. L. P. (2006). Attitude toward and propensity to engage in unethical behavior: Measurement invariance across major among university students. *Journal of Business Ethics*, *69*, 77–93.

Cheng, C. H., & Watkins, D. (2000). Age and gender invariance of self-concept factor structure: An investigation of a newly developed Chinese self-concept instrument. *International Journal of Psychology*, *35*, 186–193.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management*, *25*, 1–27.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*, 233–255.

Cieciuch, J., Davidov, E., Schmidt, P., Algesheimer, R., & Schwartz, S. H. (2014). Comparing results of an exact vs. an approximate (Bayesian) measurement invariance test: A cross-country illustration with a scale to measure 19 human values. *Frontiers in Psychology, 5*.

Cohen, J. (1988). *Statistical power analysis for the behavior sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Crockett, L. J., Randall, B. A., Shen, Y. L., Russell, S. T., & Driscoll, A. K. (2005). Measurement equivalence of the center for epidemiological studies depression scale for Latino and Anglo adolescents: A national study. *Journal of Consulting and Clinical Psychology, 73*, 47–58.

Dolan, C. V., Colom, R., Abad, F. J., Wicherts, J. M., Hessen, D. J., & van de Sluis, S. (2006). Multi-group covariance and mean structure modeling of the relationship between the WAIS-III common factors and sex and educational attainment in Spain. *Intelligence, 34*, 193–21.

Finch, W. H., & French, B. F. (2012). The impact of factor noninvariance on observed composite score variances. *International Journal of Research & Reviews in Applied Sciences, 1*,1–13.

French, B. F., & Finch, W. H. (2011). Model misspecification and invariance testing using confirmatory factor analytic procedures. *The Journal of Experimental Education, 79*, 404–428.

Fong, T. C. T., & Ho, R. T. H. (2013). Factor analyses of the hospital anxiety and depression scale: A Bayesian structural equation modeling approach. *Quality of Life Research, 22*, 2857–2863.

Fox, J. P. (2010). *Bayesian item response modeling: Theory and applications*. New York, NY: Springer.

Gagné, P. E., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research*, *41*, 65–83.

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.), Boca Raton: Chapman & Hall/CRC.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statistical Science*, *7*, 473–511.

Golay, P., Reverte, I., Rossier, J., Favez, N., & Lecerf, T. (2012). Further insights on the French WISC–IV factor structure through Bayesian structural equation modeling. *Psychological Assessment*, *25*, 496–508.

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

Hancock, G. R., & Mueller, R. O. (2001), Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sorbom (Eds.), *Structural equation modeling: Present and future—A Festschrift in honor of Karl Jöreskog.* Lincolnwood, IL: Scientific Software International.

Hancock, G. R., Stapleton, L. M., & Arnold-Berkovits, I. (2009). The tenuousness of invariance tests within multisample covariance and mean structure models. In T. Teo & M. S. Khine (Eds.), *Structural equation modeling: Concepts and applications in educational research* (pp. 137–174). Rotterdam, Netherlands: Sense Publishers.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*, 117–144.

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*, 199-218.

Johnson, E. C., Meade, A. W., & DuVernet, A. M. (2009). The role of referent indicators in tests of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 642–657.

Justel, A., & Pena, D. (1996). Gibbs sampling will fail in outlier problems with strong masking. Journal of Computational and Graphical Statistics, 5, 176–189.

Kang, Y., & Hancock, G. R. (2013, April). *Assessing measurement invariance using multigroup confirmatory factor analysis.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 650–673). New York, NY: Guilford Press.

Kaplan, D., & George, R. (1995). A study of the power associated with testing factor mean differences under violations of factorial invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 2,* 101–118.

Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*, 212–228.

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

Kim, S. H., Cohen, A. S., & Park, T. H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement*, *32*, 261–276.

Lee, S. Y. (2007). *Structural equation modeling: A Bayesian approach*. Hoboken, NJ: John Wiley & Sons Inc.

Lee, S. Y., & Song, X. Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research*, *39*, 653-686.

Lee, S. Y., Song, X. Y., & Cai, J. H. (2010). A Bayesian approach for nonlinear structural equation models with dichotomous variables using logit and probit links. *Structural Equation Modeling: A Multidisciplinary Approach*, *17*, 280–302.

Levy, R., & Choi, J. (2013). Bayesian structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed.) (pp. 563–623). Charlotte, NC: Information Age Publishing, Inc.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data : Practical and the theoretical issues. *Multivariate Behavioral Research, 32*, 53–76.

MacCallum, R. C., Edwards, M. C., & Cai, L. (2012). Hopes and cautions in implementing Bayesian structural equation modeling. *Psychological Methods*, *17*, 340–345.

Marsh, H. W. (1993). The multidimensional structure of academic self-concept: Invariance over gender and age. *American Educational Research Journal*, *30*, 841–86.

Meade. A. W., & Bauer, D. J. (2007). Power and precision in confirmatory factor analytic tests of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*, 611–635.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*, 568–592.

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

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care*, *44*, S69–S77.

Millsap, R. (2011). *Statistical approaches to measurement invariance*. New York, NY: Routledge.

Millsap, R., & Kwok, O-M., (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*, *9*, 93–115.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*, 313–335.

Muthén, B., & Asparouhov, T. (2013). BSEM measurement invariance analysis (Technical report). Los Angeles, CA: Muthén & Muthén.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *52*, 431–462.

Muthén, L., & Muthén, B. (2010). *Mplus user's guide*. Sixth edition. Los Angeles, CA: Muthén & Muthén.

Palomo, J., Dunson, D. B., & Bollen, K. (2007). Bayesian structural equation modeling. In S-Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 163–188). Amsterdam, Netherlands: Elsevier.

Raftery, A. E., & Lewis, S. M. (1992a). How many iterations in the Gibbs sampler. *Bayesian statistics*, *4*, 763–773.

Raftery, A. E., & Lewis, S. M. (1992b). Comment: one long run with diagnostics: Implementation strategies for Markov Chain Monte Carlo. *Statistical Science,* 7, 493–497.

Reise, S. P., Smith, L., & Furr, R. M. (2001). Invariance on the NEO PI-R neuroticism scale. *Multivariate Behavioral Research*, *36*, 83–11.

Sass, D. A., & Smith, P. L. (2006). The effects of parceling unidimensional scales on structural parameter estimates in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*, 566–586.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*, 210–222.

Smith, A. F., & Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society*, Series B (Methodolgical), 3-23.

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

Steenkamp, J. B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*, 78–107.

Steinmetz, H. (2013). Analyzing Observed Composite Differences Across Groups. *Methodology, European Journal of Research Methods for the Behavioral and Social Sciences*, *9*, 1–12.

Taylor, C. S., & Lee, Y. (2012). Gender DIF in reading and mathematics tests with mixed item formats. *Applied Measurement in Education*, *25,* 246–28.

Thompson, M. S., & Green, S. B. (2013). Evaluating between-group differences in latent variable means. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed.) (pp. 163–218). Charlotte, NC: Information Age Publishing, Inc.

Thompson, B. (2013). *Foundations of behavioral statistics: An insight-based approach*. New York, NY: Guilford Press.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, *3,* 4–7.

van de Schoot, R., Kluytmans, A., Tummers, L., Lugting, P., Hox, J., & Muthén, B. (2013). Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology*, *4*, 1–15.

van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, *54*, 119–135.

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research and Evaluation*, *12*, 1–26.

Yoo, B., & Donthu, N. (2001). Developing and validating a multidimensional consumer-based brand equity scale. *Journal of Business Research*, *52*, 1–14.

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

Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language assessment quarterly*, *4*, 223–233.