The Impact of Varying the Number of Measurement Invariance Constraints on the Assessment of Between-Group Differences of Latent Means

by

Yuning Xu

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Approved May 2014 by the
Graduate Supervisory Committee:

Samuel Green, Chair
Roy Levy
Keke Lai

ARIZONA STATE UNIVERSITY

August 2014

ABSTRACT

Structural equation modeling is potentially useful for assessing mean differences between groups on latent variables (i.e., factors). However, to evaluate these differences accurately, the parameters of the indicators of these latent variables must be specified correctly. The focus of the current research is on the specification of between-group equality constraints on the loadings and intercepts of indicators. These equality constraints are referred to as invariance constraints. Previous simulation studies in this area focused on fitting a particular model to data that were generated to have various levels and patterns of non-invariance. Results from these studies were interpreted from a viewpoint of assumption violation rather than model misspecification. In contrast, the current study investigated analysis models with varying number of invariance constraints given data that were generated based on a model with indicators that were invariant, partially invariant, or non-invariant. More broadly, the current simulation study was conducted to examine the effect of correctly or incorrectly imposing invariance constraints as well as correctly or incorrectly not imposing invariance constraints on the assessment of factor mean differences. The results indicated that different types of analysis models yield different results in terms of Type I error rates, power, bias in estimation of factor mean difference, and model fit. Benefits and risks are associated with imposing or reducing invariance constraints on models. In addition, model fit or lack of fit can lead to wrong decisions concerning invariance constraints.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

Social scientists frequently are interested in testing group differences in means on multiple dependent variables. Multivariate analysis of variance (MANOVA) has commonly been used for evaluating group mean differences; however, it is limiting in that it focuses on linear combinations of the variables (Hancock, Lawrence, & Nevitt, 2000). When the outcome variables are designed to reflect a latent-variable system, structural equation modeling (SEM) is considered to be a more appropriate approach (Cole, Maxwell, Arvey, & Salas, 1993). One of the advantages of SEM over MANOVA techniques is its flexibility in specifying models that are carefully tailored to match substantive theories (Byrne & Stewart, 2006; Green & Thompson, 2012). By making informed choices about model specification, competing structural equation models can be evaluated in terms of model fit. Also certain distributional assumptions that are required for traditional multivariate tests, such as normality and homogeneous variance-covaraince matrices across groups, can be avoided within SEM.

Common practice in testing differences in latent means involves initially evaluating measurement invariance (Byrne, 1998; Hancock, 1997; Hancock, 2004; Kaplan, 2000). A series of increasingly strict measurement invariance assumptions are tested in a building-block fashion until a certain level of invariance is achieved. Then, one can compare two nested models with and without equality constraints on between-group factor means. One concludes that the factor means are different in the populations if the model with constrained factor means fits significantly worse than the model with factors means allowed to vary across groups.

Valid comparison of latent means in SEM relies on measurement invariance assumptions. Ideally, for any particular latent construct, the factor loadings of each of its indicators as well as the intercepts should be invariant across the population groups (Bollen, 1989; Horn & McArdle, 1992; Meredith, 1993). This assumption is known as *scalar invariance* (Meredith, 1993; Millsap, 1998). If scalar invariance is violated and the non-invariance is ignored for analysis models, estimated differences in factor means across groups could be biased (Chen, 2008). Additionally, Type I and Type II error rates associated with testing factor mean differences could be distorted (Kaplan & George, 1995; Whittaker, 2013).

Unfortunately, rigorous scalar invariance is hard to achieve in practice. As an alternative, researchers advocated a partial invariance assumption for making meaningful interpretation of latent mean differences when a full collection of scalar invariant manifest measures is not available (Bagozzi & Edwards, 1998; Byrne, Shavelson, & Muthén, 1989; Reise, Widaman, & Pugh, 1993; Steenkamp & Baumgartner, 1998). Partial invariance applies when the invariance assumption may hold for some but not all manifest measures across different groups (Vandenberg & Lance, 2000). Invariance constraints on non-invariant measures should be relaxed to have interpretable differences of factor means. With a partial invariance assumption, researchers need to assess the invariance and non-invariance of each specific measure. An omnibus hypothesis test on sets of parameters (i.e., equality constraints on all factor loadings or all intercepts) is conducted first. If rejected, one conducted a specification search using model modification indices to assess which specific parameters within the set are non-invariant.

Although the partial invariance assumption has been more frequently adopted in testing the equivalency of factor covariance structures and mean structures, it is criticized for its exploratory nature (Bollen, 1989; Vandenberg & Lance, 2000). Sample-based modification of models should be made based not only on statistical evidence, but also on substantive knowledge of the indicators. Purely data-driven modifications are not recommended (Bollen, 1989; Steenkamp & Baumgartner, 1998). If specification searches are conducted, cross-validation is needed to support the final model. Also, as a post hoc test, the specification search procedure used to locate non-invariance is subject to capitalization on chance. Sample-based specification searches are found to seldom arrive at the true population model with any consistency unless the sample size is very large and the number of parameters in the search is small (Green & Thompson, 2012; MacCallum, Roznowski, & Necowitz, 1992).

Both advantages and disadvantages on partial invariance have been discussed in the literature (e.g., Bollen, 1989; Byrne et al., 1989; Millsap & Hartog, 1988; Steenkamp & Baumgartner, 1998). From one perspective, researchers should carefully assess the loadings and intercepts of each indicator to evaluate whether there are differences between groups or else they might yield incorrect decisions about their assessment of factor means. From an alternative perspective, the assessment based on a single sample is fraught with problems in practice, and thus researchers should feel very uncomfortable about making any decision in this process. Moreover, to the extent that more indicators are allowed to differ, estimation of factor mean differences is based on a limited number of measures (Green & Thompson, 2012). A loss of power in testing latent mean differences is expected with fewer constraints imposed on parameters.

Studies on latent factor mean modeling have been focusing on the impact of violating invariance assumptions (Chen, 2008; Kaplan & George, 1995; Wang, Whittaker, & Beretvas, 2012; Whittaker, 2013). Specifically, existing studies examined the effect of an analysis model with assumed full scalar invariance when data were generated with varying levels of non-invariance of measurement parameters. However for researchers who take on partial invariance as their assumption and wish to conduct model specification search for non-invariance, a clearer understanding is needed about the effect of making such model modifications on the assessment of differences in factor means.

This study is designed to examine the impact of modifying correct and incorrect invariance constraints on loadings and intercepts on testing factor mean differences across groups. More specifically, it focuses on the impact of specifying analysis models with varying numbers of cross-group parameter invariance constraints on factor mean difference testing given generation models that are measurement invariant, partially invariant, or non-invariant. It differs from previous ones in that it examines different choices among analysis models given particular generation models. In so doing, the results should inform researchers about the implications of decisions making in specification searches of cross-group equality constraints of parameters on tests of differences in factor means. Evaluation of testing factor mean differences should include assessment of Type I and II error rates associated with the tests of differences in factor means as well as bias, efficiency, and effect size of estimates of factor mean differences. In addition, how model fit is affected when modifying the between-group constraints should be assessed, as it is unclear whether fit indices are valid for comparing models before and after they are modified (Hancock, 1997).

The next section is a brief literature review, starting with general approaches for multivariate means modeling. After that, measurement invariance is defined in general and at specific levels, followed by the prototypical steps in conducting latent means testing using SEM. Problems in conducting these steps and the implications of model misspecification are then discussed. A statement of the objectives of the study closes this section.

Multiple Group Comparison

MANOVA and SEM are considered as alternative approaches for testing multivariate mean differences. Both approaches take into account covariances among manifest variables in the test of mean differences. But they differ in terms of their null hypotheses, model interpretations, and the relationships between composites/factors and their indicators (Cole et al., 1993; Hancock et al., 2000).

The fundamental difference between MANOVA and SEM in testing group mean differences is their null hypotheses. MANOVA evaluates group differences by forming a composite of the observed variables so that the groups are maximally differentiated on the composites in the multivariate space. The observed variables combine to create the composite, indicating an emergent variable system. SEM, on the other hand, applies when a set of observed variables are believed to reflect a latent variable system rather than forming a composite. Consequently, results from the two modeling approaches should be interpreted differently. MANOVA involves testing composites of manifest variables. These composites, typically determined using discriminant analyses, can be difficult to interpret. In contrast, SEM allows one to interpret differences of means on latent factors, which have been conceptualized by the researcher to underlie the measures. In terms of factor and indicator relationship, MANOVA implicitly assumed that each variable measures the composite in the same way across different population groups, which is analogous to measurement invariance. However this assumption often is not tested in practice. Using SEM to model latent means, researchers initially must specify and evaluate the measurement model. The indicators need to reach a certain level of measurement invariance so that a valid comparison on latent means can be made.

Two SEM modeling approaches are often used in assessing latent mean differences: Multiple Indicator Multiple Cause models (MIMIC) and structured means models (SMM) (Green & Thompson, 2012; Hancock et al., 2000 Sörbom, 1974; Jöreskog & Goldberger, 1975). As stated in the name, MIMIC introduces dummy coded indicators to the modeling system to denote group membership. For SMM, structural models are specified and evaluated for each of the investigated groups simultaneously. Relationship between the two modeling strategies is similar to that between regression and *t*-test approaches to assess univariate differences between two populations (Hancock, 1997). However the statistical assumptions underlying the two methods are different in terms of parameter constraints across-groups. MIMIC implicitly assumes that the same measurement model holds for multiple groups, whereas SMM does allow for differences in specifications of the measurement model across groups (Hancock, et al., 2000). Because MIMIC models have more restrictive assumptions, they may yield fewer parameters to estimate in testing factor mean difference and, in that sense, result in a smaller sample size requirement compared to the SMM approach. The two approaches will have the same number of degrees of freedom only if the same constraints are imposed. For this study, the SMM method is adopted in that it allows for specifying analysis models with varying number of between-group invariance constraints.

Measurement Invariance

An important assumption before conducting tests of differences in factor means in SEM is measurement invariance. Investigation of measurement invariance is often conducted under the framework of Confirmatory Factor Analysis (CFA). In CFA, a linear relationship between *p* observed variables and *m* latent factors is specified for any subject as in the following equation:

$$\mathbf{X}_g = \boldsymbol{\tau}_g + \boldsymbol{\Lambda}_g \boldsymbol{\xi}_g + \boldsymbol{\delta}_g, \tag{1}$$

where $\mathbf{X}$ is a $p \times 1$ vector of observed scores, $\boldsymbol{\tau}$ is a $p \times 1$ vector of measurement intercepts, $\boldsymbol{\Lambda}$ is a $p \times m$ matrix of factor loadings, $\boldsymbol{\xi}$ is a $m \times 1$ vector of latent factor scores, $\boldsymbol{\delta}$ is a $p \times 1$ vector of unique factor scores, and g denotes the group membership. Accordingly, the mean and variance-covariance matrices of the observed variables can be defined as follows:

$$\mathbf{E}\,(\mathbf{X}_g) = \boldsymbol{\tau}_g + \boldsymbol{\Lambda}_g \boldsymbol{\kappa}_g, \tag{2}$$

$$\boldsymbol{\Sigma}_g = \boldsymbol{\Lambda}_g \boldsymbol{\Phi}_g \boldsymbol{\Lambda}_g' + \boldsymbol{\Theta}_g, \tag{3}$$

where $\mathbf{E}\,(\mathbf{X}_g)$ is a $p \times 1$ vector of observed means, $\boldsymbol{\Sigma}_g$ is a $p \times p$ matrix of observed variances and covariances, $\boldsymbol{\kappa}_g$ is a $m \times 1$ vector of factor means, $\boldsymbol{\Phi}_g$ is a $m \times m$ matrix of factor variances and covariances, and $\boldsymbol{\Theta}_g$ is a $p \times p$ diagonal matrix of unique variances. Measurement invariance is usually defined from liberal to strict in a hierarchical manner. Common taxonomy of measurement invariance defines four levels of invariance: *configural invariance, metric invariance, scalar invariance*, and *strict invariance* (e.g., Horn & McArdle, 1992; Meredith, 1993; Millsap, 1997; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000). Configural invariance requires the same patterns of zero and non-zero factor loadings underlying each factor across population groups. A model with no invariance constraint on cross-group parameters is fit to data. If the model fits data well, configural invariance is considered to hold and the model is taken as a baseline model. The next level is metric invariance, under which the factor loadings of each indicator should be invariant across all population groups. In other words, the matrices of factor loadings should be identical across groups (i.e., $\mathbf{\Lambda}_g = \mathbf{\Lambda}$). Next, more strictly, scalar invariance requires that all the corresponding intercepts should be invariant across population groups, in addition to invariant loadings, (i.e., $\mathbf{\Lambda}_g = \mathbf{\Lambda}$ and $\boldsymbol{\tau}_g = \boldsymbol{\tau}$). Finally, as the most restrictive level, strict invariance, requires that the unique variances of each indicator are invariant across groups as well as invariant loadings and intercepts (i.e., $\mathbf{\Lambda}_g = \mathbf{\Lambda}$, $\boldsymbol{\tau}_g = \boldsymbol{\tau}$, and $\mathbf{\Theta}_g = \mathbf{\Theta}$).

Prototypical Procedures for Assessment of Latent Mean Differences

Assumptions. Traditional opinion states that full metric invariance should hold before testing scalar invariance; and only when a full scalar invariance holds, can one proceed to analyzing factor mean differences (e.g., Bollen, 1989; Horn & McArdle, 1992). Different factor loadings across groups would indicate that a unit change in factor scores will not result in the same change in the observed indicators for the different groups. Similarly, different intercepts would indicate that subjects with same changes in factor scores will have different changes in observed scores.

However, it is found that full metric invariance or full scalar invariance is hard to achieve in practice. A literature review of cross-cultural and cross-ethnic studies showed that at least 60% of the studies had factor loadings that were not equivalent between different cultural groups (Chen, 2008). Factor loadings in the focal group were usually higher than those in the reference group. Also, it is found that survey measurement instruments typically do not exhibit scalar invariance across populations, especially in many large-scale international studies (De Beuckelaer & Swinnen, 2011). As a result, partial invariance has been considered an attractive alternative for latent mean modeling, given that stringent invariance assumption on loadings and intercepts is not tenable in most substantive studies (e.g., Byrne et al., 1989; Byrne & Stewart, 2006; Carle, Millsap, & Cole, 2008).

With a partial invariance assumption, it is assumed that a subset of indicators with invariant factor loadings and intercepts is sufficient for assessing factor mean differences. As stated by Marsh and Hocevar (1985), comparison of factor means is still feasible when most of the indicators are invariant, and under these conditions, failure to achieve full factorial invariance is trivial from a practical point of view. More liberally, Byrne et al. (1989) argued that, other than the one indicator loading fixed to 1.00 for identification purpose, and one indicator intercept constrained to be equal between groups, further constraints are unwarranted for testing factor means differences.

Procedures. A typical procedure for assessment of latent mean differences starts with the test of measurement invariance. A specification search procedure could be conducted to explore the invariance of specific loading and intercept pairs under a partial invariance assumption.

- Configural invariance is tested first as a baseline model to see whether the same indicators represent the same latent factors across groups.

- If configural invariance is supported, metric invariance is examined where all between-group loading pairs are constrained to be equal. Overall fit of this model is evaluated, and the model is then compared to the configural invariance model to see whether the model fit decreases significantly using a chi-square difference test.
    - If overall fit of the metric invariance model is tenable and there is no significant decrease in fit compared to the configural invariance model, metric invariance is considered to hold.
    - If there is a significant decrease in fit, metric invariance of models fails. One suggestion under this situation is to refer to modification information such as the Lagrangian Multiplier (LM) tests, which indicate constraints on parameters that can be relaxed to improve model fit in the sample. By synthesizing this information as well as substantive knowledge, researchers might allow some loadings to differ between groups. The updated model is assessed for fit and compared with the metric invariance model using a chi-square difference test. If this model is tenable, it is believed that partial metric invariance is met.
- In this step, the between-group intercept constraints are evaluated. Initially, intercepts are constrained to be equal between groups for an indicator if the factor loadings for that intercept were tested to be equal between groups based on the previous step.
    - If the model that imposes between-group intercept constraints is tenable and there is no significant decrease in fit compared to the previous model,

a full scalar invariance is met and the differences in factor means can be assessed.

- If the model that imposes between-group intercept constraints does not fit well, constraints on potential non-invariant intercept pairs can be assessed using modification indices and substantive knowledge about the indicators. Based on this information, one or more of the constrained intercept pairs might be relaxed. At this point, differences in factor means can be assessed.

After determining (partial) scalar invariance, two sets of models are specified to test latent mean differences: one with the latent means constrained to be equivalent across groups (restricted model) and the other with the means freely estimated (full model). The rest of the model is specified as determined through the previous steps of assessing measurement invariance. Both models are fitted to the data, and the model fit is compared using a chi-square difference test. If the increment of fit is significant from the restricted model to the full model, the latent means are considered to be different across groups.

Issues with Specification Search Methods

Specification searches are commonly conducted in practice to assess partial invariance before testing between-group factor mean differences. Several issues should be considered in conducting searches when making decisions about modifying invariance constraints on parameters, as modifications based simply on model fit indices might lead to misspecified models.

Preference for Conceptual Choices. Empirical search procedures for assessing partial measurement invariance can lead to freeing invariance constraints on parameter pairs that have no clear interpretation. In theory, researchers should decide when to constrain or relax a specific pair of parameters across groups based on their substantive knowledge, and then assess these decisions based on model fit. However theory might not be available or can be inaccurate, leading to a dependence in practice on empirical methods, such as modification indices (i.e., LM tests) and Expected Parameter Change (EPC). On the other hand, confidence in making decisions about measurement invariance of parameters should come from both understanding investigated constructs and their indicators as well as statistical support from empirical data (Hancock, Stapleton, & Arnold-Berkovits, 2009). Without theoretical support, relaxing constrained parameters based purely on empirical findings should be done very cautiously, perhaps only when the modification indices are significant and with a substantial change in parameter estimates in a cross validation process (Kaplan, 1989; SteenKamp & Baumgartner, 1998).

Choice of Referent Variable. Selection of referent indicators (RI) is critical in assessing measurement invariance of multi-group models. A referent indicator is the variable that is chosen for assigning the metric for a factor, typically by fixing its loading to one for all groups. For modeling means, typically the intercepts for that indicator are also constrained across groups. A non-invariant RI could lead to biased estimates of model parameters and inadequate model fit initially (Vandenberg & Lance, 2008). When a selected referent indicator is non-invariant, the discrepancy between the loadings or/and intercepts on that indicator is transferred to the parameters of other indicators as well as to the factors. For example, Johnson, Meade, & DuVernet (2009) showed that inappropriate selection of a referent indicator produced biased results for indicator-level tests under partial metric invariance. Several strategies have been proposed to address on how to identify RIs that are truly invariant (Cheung & Rensvold, 1999; Reise et al., 1993; Yoon & Millsap, 2007). However the methods has not been widely adopted in practice because they either are too labor intensive or have requirements about the data that cannot be met in practice (e.g., require large sample sizes) (French & Finch, 2008; Yoon & Millsap, 2007).

Problems with Significance Testing. To conduct model modifications in SEM, three asymptotically equivalent significance tests are often used by most researchers: Likelihood Ratio (LR) test, Lagrangian Multiplier (LM) test, and Wald (W) test. As exploratory tools, these tests have their merits but also limitations briefly in that (a) the order in which parameters are freed or restricted can affect the significance tests for the remaining parameters; (b) probability levels associated with the W and LM statistics in the stepwise procedures are not likely to be accurate (Bollen, 1989); (c) multiple tests are conducted with little or no attempt to control for familywise error rates in practice (Green & Babyak, 1997); and (d) non-rejection of null hypothesis does not imply that the constraints are appropriate (potentially due to a lack of power). Especially in testing invariant parameter pairs, Cheung and Rensvold (1999) pointed out that the commonly used LM test based on a decomposition of the multivariate LM test is suspect because the index for each fixed parameter is calculated in the fully constrained model, where all other indicators in the model are assumed to be invariant.

Implications of Model Misspecification

Misspecification of measurement models can negatively affect assessment of factor mean differences between groups. Simulation studies have been conducted to show that incorrectly imposing between-group constraints on factor loadings or intercepts can lead to inaccurate results in estimation of factor mean differences and distorted model fit.

Bias of Differences in Factor Means. To investigate the impact of incorrectly constraining non-invariant loading on factor means, Chen (2008) simulated factor loadings in one group that were uniformly larger than their counterparts in the other group, while the latent means were equal across groups. When all cross-group loading pairs were constrained to be invariant in the model, a pseudo group difference in latent means appeared. On the other hand, when the non-invariance of loadings was simulated to form a mixed pattern with the average loadings being equal between groups, the estimated difference in latent means was close to zero (i.e., the true difference). In the same study, Chen also included conditions in which the factor loadings were simulated as invariant, but intercepts were non-invariant and the pattern of non-invariance was manipulated. In one condition, the intercepts in one group were uniformly larger than the intercepts in the second group, whereas in the second condition, the intercepts had a mixed pattern in which the average intercept was equivalent between groups. When the intercepts incorrectly constrained to be the same, the factor mean differences were biased in the first condition and were minimized to zero in the second condition. The study by Wang et al. (2012) reached similar conclusion. Additionally, it is found that bias in latent mean difference estimates increases as the differences in factor loadings increased across groups (in a uniform pattern). Again, when the differences in factor loadings follow a mixed pattern and the differences are balanced across groups, the estimated latent mean differences are unbiased. Although not explicitly stated in Chen (2008), it appeared that incorrectly constraining intercepts might have a greater impact on latent mean estimation than incorrectly constraining loadings (when the non-invariance patterns are uniform).

Type I Error and Power of Testing Factor Mean Differences. The impact of inappropriate invariance constraints on statistical properties of latent mean estimation has been studied as well. Kaplan and George (1995) conducted a study to investigate the impact of degrees of non-invariance of factor loadings on the power associated with the test of factor mean differences. In their study, the proportion and magnitude of uniform non-invariance in cross-group loading pairs were manipulated for data generation, while all loading pairs were constrained to be equal in model analyses. The results illustrated that power is artificially boosted as the proportion and magnitude in factor loading non-invariance increased. Consistently, Wang et al. (2012) found that both Type I error and power associated with latent mean difference testing appear to be slightly inflated when all loading pairs were constrained to be equal under partial metric invariance. Further, unequal sample sizes and factor variance ratios between-groups caused problematic Type I error rate and power. In addition, they found that constraining or not constraining non-invariant intercepts had a greater impact on Type I error rate and power than constraining or not constraining non-invariant loadings. In Whittaker (2013), it was shown that Type I error rates were inflated as the proportion of non-invariant intercept pairs increased and as the magnitude of the non-invariance in intercepts increased. Similarly, she demonstrated that constraining intercepts that were uniformly different could artificially increase the power of testing factor mean differences. For conditions with an equal proportion of non-invariant intercepts and all intercepts constrained to be equal in the analysis model, Type I error rates became more inflated and power artificially increased as the total number of indicators became greater.

Model Fit. Few works have focused on model fit in the measurement invariance literature for latent mean modeling. Whittaker (2013) investigated changes in model fit when invariance constraints were misspecified. The results suggested that average CFI values tended to decrease as the proportion of misspecified non-invariant intercepts increased as well as the magnitude of the between-group differences in intercepts increased. RMSEA also demonstrated poorer fit under the same set of conditions.

Study Objective

Measurement invariance needs to be tested prior to factor mean modeling. For researchers who assume their factor indicators are partially invariant and wish to test the invariance for specific parameters, the primary method used to make decisions about measurement invariance involves specification searches. These searches could be problematic and require researchers to make decisions involving choice of significance tests, combining results from both modification indices and ECP, controlling for Type I error, and integration of substantive theory and empirical outcome. In making these complex decisions, it would be helpful to understand the impact of their decisions on the evaluation of factor mean differences.

In the literature, studies were typically designed to assess a particular analysis model given data generated using various models. Results from these studies indicated that violation of a full scalar invariance assumption has a profound impact on estimation of factor mean differences. However for those who plan to test invariance based on specification searches, it stays unclear that what is the impact of modifying analysis models in terms of the invariance constraints on factor mean estimation. The current study addressed this issue by conducting analysis using multiple analysis models on data generated using any one model. The approach is similar to the one faced by applied researchers, who must decide what parameters to constrain or freely estimate for their dataset.

More specifically, the focus of this study is on the impact of correct or incorrect decisions about invariance of loadings and intercepts on assessment of factor mean differences. Three levels of measurement invariance were simulated in data generation: full invariance, partial invariance, and non-invariance. For each generation model, a number of analysis models were considered: those with a minimal number of between-group constraints on intercepts and loadings, those with all between-group constraints imposed on intercepts and loadings, and those in-between these two alternatives. In terms of results, bias, efficiency, and effect size of the estimates of factor mean differences were examined. Type I error rates, and power associated with testing factor mean differences were focused on. Also, the magnitude of model fit indices was studied with changes in the number of correctly and incorrectly imposed invariance constraints.

CHAPTER 2

METHODS

A simulation study was conducted to explore the effect of different numbers of invariance constraints on factor loadings and intercepts on tests of differences in latent factor means between groups. Data were simulated using three types of models: models with a) invariant loadings and invariant intercepts for all indicators between groups (full invariance, F-IV); b) non-invariant loadings and non-invariant intercepts for half of the indicators between groups (partial invariance, P-IV); c) non-invariant loadings and non-invariant intercepts for all the indicators between groups (non-invariance, N-IV). Analysis models with different numbers of invariance constraints were applied to datasets simulated using the three types of models. Type I error rates and power of tests of between-group latent mean differences were assessed, as well as bias, efficiency, and effect size in the estimates of factor mean differences. In addition, model fit indices were evaluated. These indices included CFI, RMSEA, and SRMR.

A two-group, single-factor model with 8 measured indicators was investigated in the current study. The choice of 8 indicators was consistent with the design of the study by Chen (2008), who examined bias in factor mean difference estimation under non-invariance of factor loadings and intercepts. A path diagram of the model is presented in Figure 1. The triangle containing 1 in each of the groups represented the unit predictor. Coefficients from the unit predictors represented the indicator intercepts and factor means. A number of simulation and analysis conditions were manipulated, as described in the following section.

A Two-Group Single-Factor Model

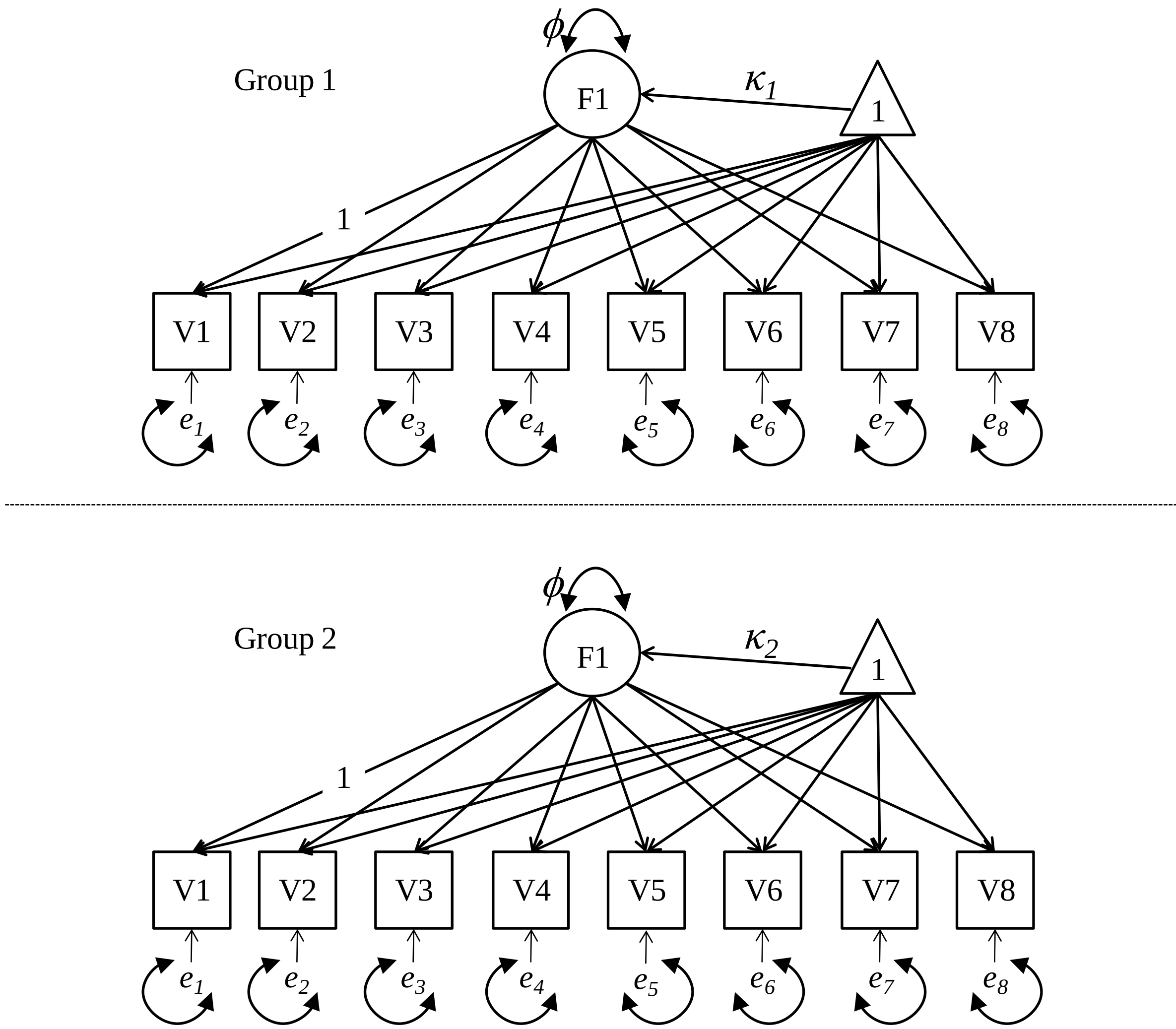


Figure 1. A Two-Group Single-Factor Model with Mean Structure.

Simulation Conditions

Five simulation variables were manipulated: (a) proportion of indicators with non-invariant parameters, (b) pattern of non-invariant parameters, (c) magnitude of between-group differences in factor loadings and intercepts (d) magnitude of between-group difference in latent factor means, and (e) sample size.

Proportion of indicators with non-invariant parameters. As described earlier, three levels of non-invariance were considered regarding factor loadings and intercepts: full invariance (F-IV), partial invariance (P-IV), and non-invariance (N-IV). For an invariant indicator, both its loadings and intercepts were invariant between groups; and for a non-invariant indicator, both the loadings and intercepts were non-invariant. In the F-IV conditions, all the indicators were invariant; in the P-IV conditions, half of the indicators were invariant and the other half of the indicators were non-invariant; and, in the N-IV conditions, all indicators were non-invariant.

Pattern of non-invariance. P-IV and N-IV conditions were combined with uniform and mixed patterns of non-invariance in the design of the study. For a uniform pattern of non-invariance, the differences in parameter values between the two groups were the same across all non-invariant indicators. Specifically, factor loadings in group 1 were always greater than the loadings in group 2, and correspondingly, intercepts in group 1 were always lower than intercepts in group 2. For a mixed pattern of non-invariance, the between-group differences in parameter values were the same, and, in one direction for half of the non-invariant indicators and in the opposite direction for the other half of the non-invariant indicators. For example, in the mixed P-IV condition with 4 invariant indicators and 4 non-invariant indicators, 2 of the 4 non-invariant indicators had higher loadings and lower intercepts in group 1 and the other 2 non-invariant indicators had higher loadings and lower intercepts in group 2.

Between-group differences in factor loadings and intercepts (DIF). For all invariant indicators, the loadings were set to be .50 and the intercepts were 1.0. For all non-invariant indicators, two levels of between-group differences in loadings and intercepts were considered: 10% and 20%. Either 10% or 20% difference in loadings was created by a proportional rescaling method where a higher loading was reduced by either 10% or 20% to yield a lower loading (Yoon & Millsap, 2007). Thus the loading differences were comparable across conditions and did not depend on loading size. The magnitude of any non-invariant loading pair was balanced around .50. For the 10% DIF level, loadings were set to be .56 and .45 (.56 – 10% × .56 = .50 and .50 – 10% × .50 = .45); for the 20% level, loadings were .625 and .40. Intercept differences were created by simply subtracting .05 or .10 (corresponding to 10% or 20% loading differences) from 1.0 in one group and adding .05 or .10 to 1.0 in the other group. It should be noted that any non-invariant indicator is simulated to have a higher loading and a lower intercept in one group, and, a lower loading and a higher intercept in the other group. Combination of a higher loading and a lower intercepts for any indicator reflected the relationship between parameters in substantive research. Analogously, increasing slopes of predictors in regression models leads to decreases in intercepts, holding everything else constant. In a P-IV condition where 4 indicators were invariant and 4 indicators were non-invariant, with 10% DIF, $\boldsymbol{\Lambda}_1$ = [.50, .50, .50, .50, .56, .56, .56, .56], $\boldsymbol{\tau}_1$ = [1.00, 1.00, 1.00, 1.00, .95, .95, .95, .95] for group 1, and $\boldsymbol{\Lambda}_2$ = [.50, .50, .50, .50, .45, .45, .45, .45], $\boldsymbol{\tau}_2$ = [1.00, 1.00, 1.00, 1.00, 1.05, 1.05, 1.05, 1.05] for group 2. In Table 1, population loadings and intercepts were presented for conditions with 10% DIF.

Table 1

Population Factor Loadings and Intercepts for conditions with 10% differences in factor loadings

Pattern and proportion of non-invariance		Loadings	Intercepts
F-IV	Group 1	[.50, .50, .50, .50, .50, .50, .50, .50]	[1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00]
	Group 2	[.50, .50, .50, .50, .50, .50, .50, .50]	[1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00]
Uniform P-IV	Group 1	[.50, .50, .50, .50, .56, .56, .56, .56]	[1.00, 1.00, 1.00, 1.00, .95, .95, .95, .95]
	Group 2	[.50, .50, .50, .50, .45, .45, .45, .45]	[1.00, 1.00, 1.00, 1.00, 1.05, 1.05, 1.05, 1.05]
Mixed P-IV	Group 1	[.50, .50, .50, .50, .45, .45, .56, .56]	[1.00, 1.00, 1.00, 1.00, 1.05, 1.05, .95, .95]
	Group 2	[.50, .50, .50, .50, .56, .56, .45, .45]	[1.00, 1.00, 1.00, 1.00, .95, .95, 1.05, 1.05]
Uniform N-IV	Group 1	[.56, .56, .56, .56, .56, .56, .56, .56]	[.95, .95, .95, .95, .95, .95, .95, .95]
	Group 2	[.45, .45, .45, .45, .45, .45, .45, .45]	[1.05, 1.05, 1.05, 1.05, 1.05, 1.05, 1.05, 1.05]
Mixed N-IV	Group 1	[.45, .45, .45, .45, .56, .56, .56, .56]	[1.05, 1.05, 1.05,1 .05, .95, .95, .95, .95]
	Group 2	[.56, .56, .56, .56, .45, .45, .45, .45]	[.95, .95, .95, .95 1.05, 1.05, 1.05, 1.05]

Between-group difference in latent factor means. Factor means for group 1 (κ_1) were set to 0 across all conditions; factor mean for group 2 (κ_2) were set either to 0 or .20. Consequently the differences in latent factor means were equal to the factor means for group 2 (i.e., $\Delta\kappa = \kappa_2$). Type I error rates were examined for conditions with $\Delta\kappa = 0$, whereas power was assessed for conditions with $\Delta\kappa = .20$. Variances of factors were 1.00 for both groups across all conditions. A pilot study was conducted to ensure that power was sensitive to variation across conditions and did not suffer from ceiling effects.

Sample size. Total sample size was set to 300 or 500 in an attempt to mimic small and moderate sample sizes in practice. For simplicity, sample sizes were designed to be same for the two groups for all conditions. Thus each group had either 150 or 250 simulated subjects.

In summary, for F-IV, there were a total of 4 simulation conditions (2 between-group differences in factor means $\times$ 2 sample sizes); for P-IV and N-IV, there were a total of 32 simulation conditions (2 proportions of non-invariant parameters $\times$ 2 patterns of non-invariance $\times$ 2 DIFs $\times$ 2 between-group differences in factor means $\times$ 2 sample sizes). Over all manipulated simulation variables, the design included 36 simulation conditions. For each condition, 1000 replications were generated.

Analysis Conditions

Regardless of the choice of analysis models, the following constraints were imposed on all analysis models for identification purpose: (a) the loadings for one indicator were constrained to be equal across two groups; (b) factor variance was fixed at 1.00 in group 1; (c) the intercepts of the selected indicator in step (a) were constrained to be equal across groups; and (d) the factor mean was fixed at 0 in group 1. Indicator selection in (a) depended on the specification of analysis models.

Four sets of analysis models were conducted on each dataset from the 36 simulation conditions. Within each set of the models, there could be one or multiple analysis models, depending on the condition. Table 2 showed the four different sets of models used to analyze each of the simulated datasets. They differed in terms of the number of invariance constraints imposed on the loading and intercept pairs. An X in the table indicated these constraints were imposed on the parameters of an analysis model. As shown in the table, analysis models had one of four sets of between-group equality constraints on parameters: a) constraints on 1 loading and 1 intercept, b) constraints on 4 loadings and 4 intercepts, c) constraints on 8 loadings and 8 intercepts, and d) constraints on 8 loadings and 1 intercept. For model sets a, b, and c, intercepts of indicators were constrained to be equal whenever the loadings of these indicators were constrained to be equal. For model set d, between-group invariant constraints were imposed on all loading pairs, but only on one of the intercept pairs. Analysis results from model sets a, b, and c reflected the effect of increasing/decreasing the number of invariance constraints on model estimation. Comparison between model sets a and d reflected the effect of changing the number of loading invariance constraints, while keeping the number of intercept constraints constant. The comparison between model sets d and c reflected the effect of the number of intercept invariance constraints, while keeping number of loading constraints constant. Analysis models represented by the top-right cell of Table 2 were excluded in the study because the intercept of an indicator is usually not constrained once the loading of the indicator is considered to be non-invariant in the measurement invariance specification search.

Table 2

Model Sets with Different Numbers of Between-Group Invariance Constraints on Factor Loadings and Intercepts

Number of constraints on loading pairs	Number of constraints on intercept pairs			
	1 intercept	2 intercepts	4 intercepts	8 intercepts
1 loading	X (a)	-	-	-
2 loadings	-	-	-	-
4 loadings	-	-	X (b)	-
8 loadings	X (d)	-	-	X (c)

Models with same numbers of invariant constraints can be specified differently for particular datasets. Each model set in Table 2 included one or multiple analysis models, depending on the generation model. Analysis models across model sets were categorized into five types. Before defining the five model types, a few terms are defined: an *appropriate invariance constraint* is an invariance constraint imposed on parameters for an indicator that were generated to be equal between groups; and an *inappropriate invariance constraint* is an invariance constraint imposed on parameters for an indicator that were generated to be unequal between groups. Furthermore, *inappropriate invariance constraints on unbalanced parameters* across multiple indicators occur if these parameters were generated to have greater values in one of the groups. In contrast, *inappropriate invariance constraints on balanced parameters* across multiple indicators occur if half of the constrained parameters for indicators were generated to have greater values in one group and the other half had greater values in the other group; the absolute magnitude of differences between parameters across indicators were the same.

Using these four terms, we next define five types of analysis models. We differentiated among these five types of models in terms of the invariance constraints imposed on parameters, but not in terms of the parameters that were not constrained to be equal between groups.

(1) An *appropriate model* is an analysis model in which all invariance constraints are imposed on parameters that were generated to be equal.

(2) A *100% unbalanced model* is an analysis model in which all invariance constraints are imposed on unbalanced parameters.

(3) A *50% unbalanced model* is a model in which 50% of the invariance constraints are on unbalanced parameters and 50% of the invariance constraints are on parameters that were generated to be equal.

(4) A *100% balanced model* is a model in which all invariance constraints are imposed on balanced parameters.

(5) A *50% balanced model* is a model in which 50% of the invariance constraints are on balanced parameters and 50% of the invariance constraints are on parameters that were generated to be equal.

For F-IV simulation data, all analysis models are *appropriate models* (regardless of the number of invariance constraints) because all indicators were generated to be equal. Likewise, all analysis models are *100% unbalanced models* for uniform N-IV data because all indicators were generated to be uniformly unequal. Different types of analysis models exist when generation models were mixed N-IV, uniform P-IV, and mixed P-IV. Table 3 and Table 4 elucidated the model types for data from these generation models. As shown in Table 3, the model with a total of 8 constraints on loadings and 8 constraints on intercepts when fitting mixed N-IV data is a *100% balanced model* because all the constraints were on balanced parameters. Also, when generation model is mixed P-IV (shown in the right half of Table 4), a model with 8 constraints on loadings and 8 constraints on intercepts is a *50% balanced model* because half of the constraints are appropriate and half are on balanced parameters.

Table 3

Model Specifications When Generation Model Is Mixed N-IV

Generation Model: Mixed N-IV			
Analysis Models			
Total Invariance Constraints	Constraints on non-invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type
$8\ \lambda + 8\ \tau$	$4\ \lambda$ and $4\ \tau$	$4\ \lambda$ and $4\ \tau$	*100% Balanced Model*
$4\ \lambda + 4\ \tau$	$4\ \lambda$ and $4\ \tau$	-	*100% Unbalanced Model*
	$2\ \lambda$ and $2\ \tau$	$2\ \lambda$ and $2\ \tau$	*100% Balanced Model*
$8\ \lambda + 1\ \tau$	$4\ \lambda$ and $1\ \tau$	$4\ \lambda$ and $0\ \tau$	-
$1\ \lambda + 1\ \tau$	$1\ \lambda$ and $1\ \tau$	-	*100% Unbalanced Model*

According to Tables 3 and 4, a total of 22 analysis models were analyzed for these three simulation conditions. There were also 4 analysis models for F-IV conditions, and 4 analysis models for uniform N-IV conditions. Combining with 2 levels of DIF, 2 levels of between-group differences in latent factor means, and 2 levels of sample size, a total of 224 analysis models were conducted in the study.

Table 4

Model Specifications when generation models are Uniform P-IV and Mixed P-IV

Generation Model: Uniform P-IV				Generation Model: Mixed P-IV				
Analysis Models				Analysis Models				
Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type
8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*50% Unbalanced Model*	8 λ + 8 τ	4 λ and 4 τ	2 λ and 2 τ	2 λ and 2 τ	*50% Balanced Model*
4 λ + 4 τ	4 λ and 4 τ	-	*Appropriate Model*	4 λ + 4 τ	4 λ and 4 τ	-	-	*Appropriate Model*
	-	4 λ and 4 τ	*100% Unbalanced Model*		2 λ and 2 τ	2 λ and 2 τ	-	*50% Unbalanced Model*
	2 λ and 2 τ	2 λ and 2 τ	*50% Unbalanced Model*		2 λ and 2 τ	1 λ and 1 τ	1 λ and 1 τ	*50% Balanced Model*
					-	2 λ and 2 τ	2 λ and 2 τ	*100% Balanced Model*
8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	8 λ + 1 τ	4 λ and 1 τ	2 λ and 0 τ	2 λ and 0 τ	-
	4 λ and 0 τ	4 λ and 1 τ	-		4 λ and 0 τ	2 λ and 1 τ	2 λ and 0 τ	-
1 λ + 1 τ	1 λ and 1 τ	-	*Appropriate Model*	1 λ + 1 τ	1 λ and 1 τ	-	-	*Appropriate Model*
	-	1 λ and 1 τ	*100% Unbalanced Model*		-	1 λ and 1 τ	-	*100% Unbalanced Model*

Summarizing Results

*M*plus 6.11 and R 3.0 were used for data generation and model analysis. The results are summarized in terms of empirical Type I error rates and power rates associated with latent factor mean difference testing between groups. Empirical Type I error rates and power were computed from Wald tests of factor mean differences based on critical *z* values at .05 level. A pilot study indicated that Type I error and power rates obtained by the Wald test and chi-square difference test of factor mean differences were consistent to the third decimal places. Bias, efficiency, and effect size of factor mean difference estimates also were examined. Bias was defined as the mean of the differences between the estimated factor mean differences and the population factor mean differences. For conditions with factor mean differences simulated to be .20, relative bias was also examined. Relative bias was defined as bias in mean difference estimates dividing by the difference in population means. A standardized effect size statistic was computed by dividing the difference between the factor means by the square root of the pooled variance of the latent variables from each group (Hancock, 2001). Efficiency is defined as the standard deviation of estimated factor mean differences. In addition, model fit was assessed based on three indices: CFI, RMSEA, and SRMR.

CHAPTER 3

RESULTS

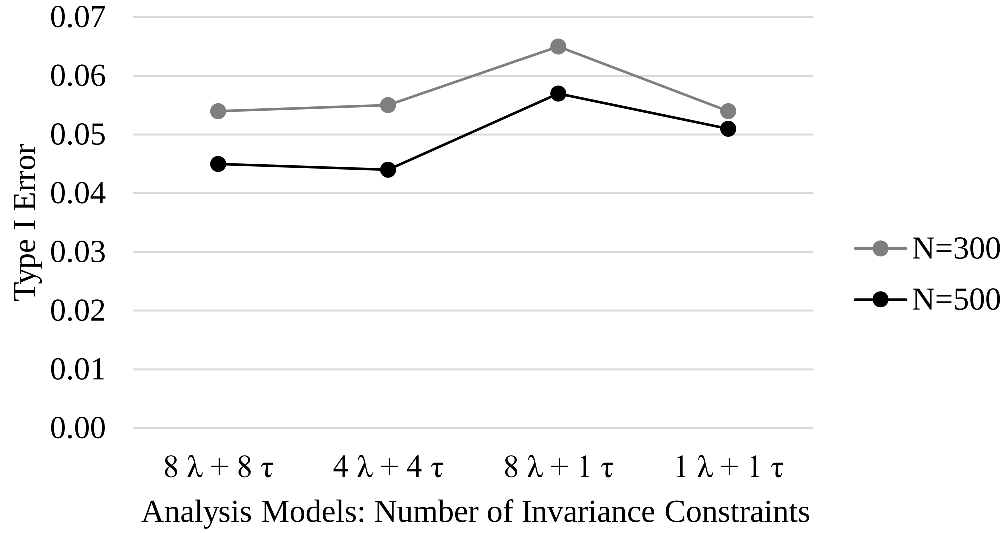
Results were presented based on the models used to generate the data: full invariance, uniform non-invariance, mixed non-invariance, uniform partial invariance, and mixed partial invariance. For each generation model, results from selected conditions were presented. Results from all conditions for the five generation models were included in Appendix A.

Full Invariance Conditions (F-IV)

The generation model for F-IV data had equal loadings and equal intercepts across groups. As described in Table 2, the generated data were analyzed using four models that varied in the number of between-group invariance constraints imposed on loadings and intercepts. All four models were *appropriate models*. Figure 2 presented the Type I error rates and power for the tests of factor mean differences based on the four analysis models. The means of three model fit indices from these analysis models were displayed in Figure 3.

As shown in panel (a) of Figure 2, Type I error rates fell between .04 and .06 for all but one of the analysis models. Conditions with a larger sample size tended to have more conservative Type I error rates. For conditions with the same sample size, the most inflated Type I error rates occurred when the analysis model had 8 loading constraints and 1 intercept constraints. Overall, Type I error rates from analysis models with different numbers of invariance constraints were deemed acceptable under F-IV conditions according to the liberal cutoff criterion of robustness ($\alpha \pm \alpha/2$) (Bradley, 1978).

(a) Generation Models: F-IV, $\underline{\Delta\kappa = 0}$



(b) Generation Models: F-IV, $\underline{\Delta\kappa = .20}$

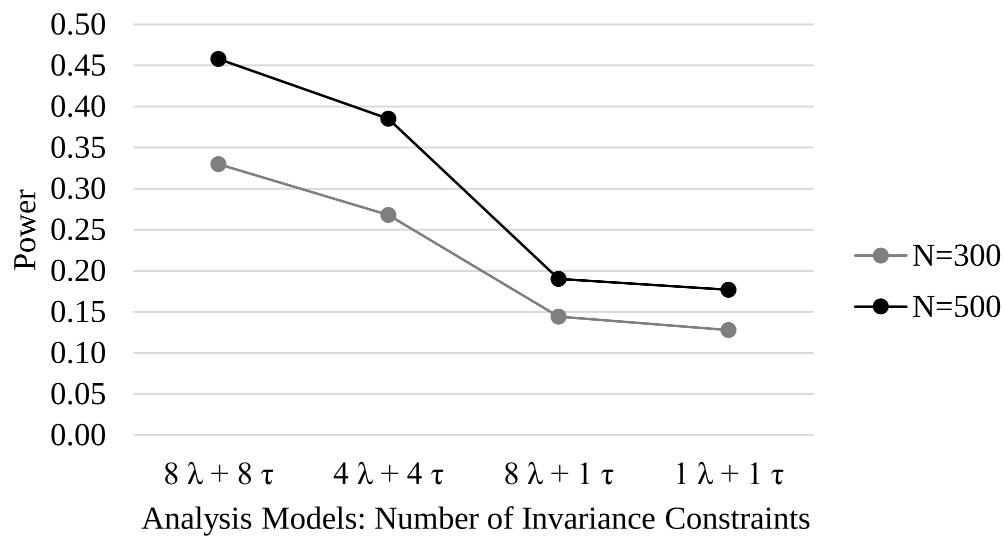


Figure 2. Type I Error Rates and Power Associated with Factor Mean Difference Testing from Various Analysis Models When Data Are Generated from F-IV Conditions.

In panel (b) of Figure 2, it was observed that power decreased as invariance constraints were removed in the analysis models. Invariance constraints on intercepts appeared to have a greater effect on power than constraints on factor loadings. For example, with $N = 500$, power dropped from .46 to .19 when 7 intercept invariance constraints were removed, whereas the decrease in power was from .19 to .18 when 7 loading constraints were removed. The greater effect of invariance constraints on intercepts was consistent with prior results (Chen, 2008; Wang et al., 2012), although not explicated.

Table 7 in Appendix A presented the average bias and relative bias in estimates of factor mean differences, as well as efficiency and estimated effect size. Bias was relatively small across all conditions under F-IV. Absolute value of the average bias ranged from .0004 to .0263, with most values being positive. Average relative bias ranged from 3.55% to 13.15%. Within each generation condition, removing appropriate invariance constraints in analysis models led to increasing in estimate bias. Similarly, the inflation in estimated effect size increased and the efficiency of estimation decreased when appropriate invariance constraints were removed. Again, it was found greater effect of constraining intercepts than loadings on all these quantities.

According to Figure 3, all analysis models fit the data adequately for F-IV conditions: the mean CFIs were above .98, the mean SRMRs were below .06, and the mean RMSEAs were below .02. The mean values for CFI and SRMR indicated better fit with fewer invariance constraints. On the other hand, the mean RMSEAs tended to increase slightly as fewer invariance constraints were imposed on the analysis model within each generation condition. Based on RMSEA, models with fewer appropriate constraints showed worse fit. In other words, changes in RMSEA indicated that more

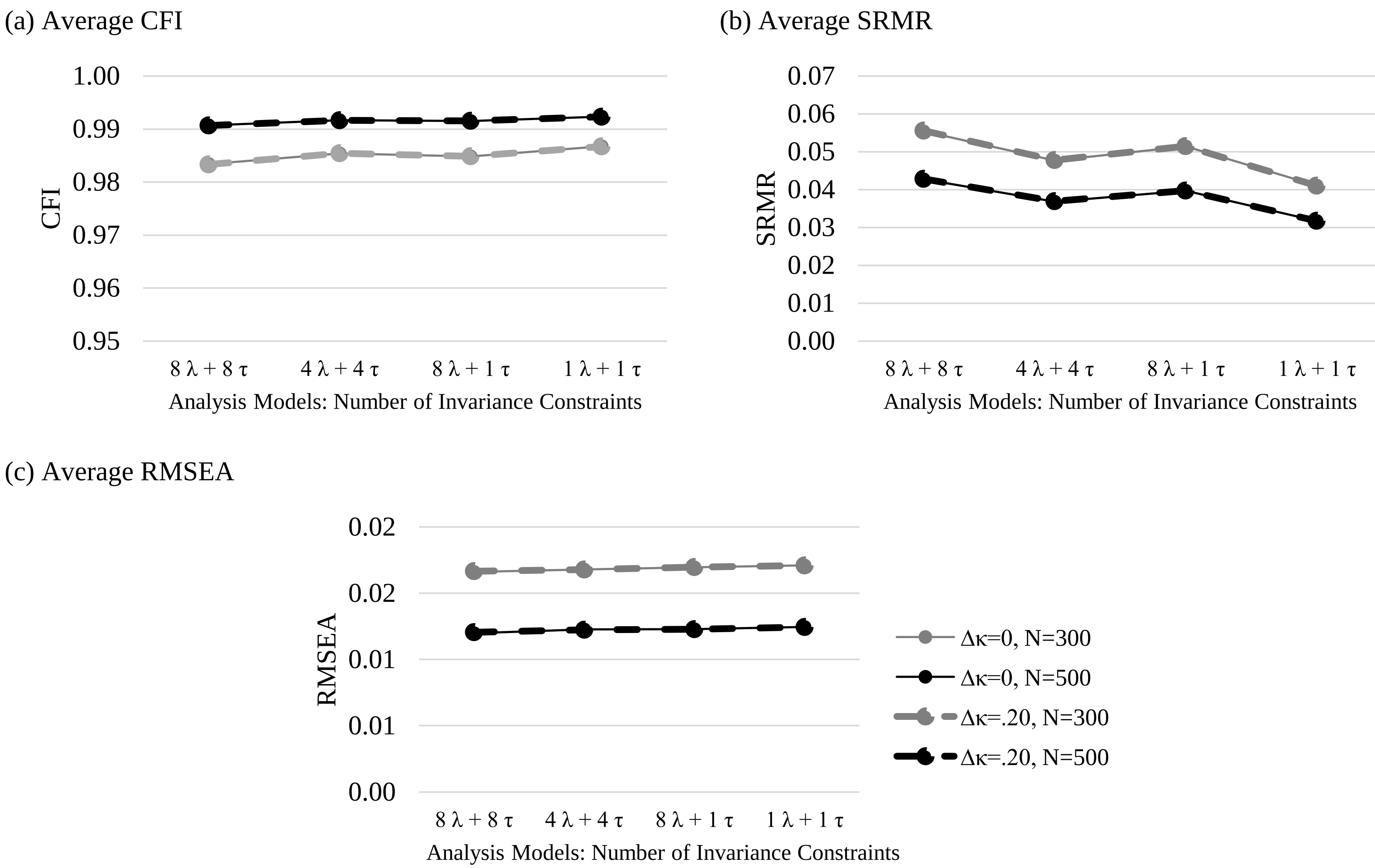


Figure 3. Model Fit Indices from Various Analysis Models When Data Are Generated from F-IV Conditions.

parsimonious models are preferred when all imposed constraints are appropriate. Generally, models fitting on data with larger sample size yield better fit

Uniform Non-Invariance Conditions (Uniform N-IV)

For uniform N-IV generated data, all loadings for indicators were greater in group 1and all intercepts for indicators were greater in group 2. Four analysis models were fit to each simulated dataset. All analysis models for uniform N-IV generated data were *100% unbalanced models*.

Type I error rates from all analysis models were highly inflated for uniform N-IV data, as shown in Figure 4. Conditions with higher DIF and/or larger sample size had higher inflated Type I error rates. Within each generation condition, the magnitude of inflation became smaller as more inappropriate invariance constraints were removed from analysis models. Type I error rates decreased more dramatically as a function of the number of intercept constraints in comparison with the number of loading constraints. Interestingly, this effect of intercept constraints on Type I error mimicked the effect of intercept constraints on power for F-IV generated data. Estimates and effect size of factor mean differences are highly biased when generated data were uniform N-IV (see Table 9 in Appendix A). Similarly with F-IV, average bias and inflation in estimated effect sized increased and efficiency became lower as invariance constraints were removed in analysis models, although they are inappropriate.

CFI, SRMR, and RMSEA from all analysis models under uniform N-IV conditions were presented in Figure 5. The performance of the three indices indicated that the models fitted uniform N-IV data fairly well, although they were all misspecified to some extent. Counterintuitively, CFI and SRMR indicate better model fit as more inappropriate invariance constraints are removed from the model. The plots of the three indices for uniform N-IV data were similar to those for F-IV data. For uniform N-IV data, the original non-invariant indicators are rescaled as “invariant” indicators in analysis models with invariance constraints because the group differences in loadings and intercepts are simulated to be exactly the same across all indicators. With this readjustment in scale, analysis models fit data under uniform N-IV conditions as well as they fit data under F-IV conditions. The scale readjustment was also observed and illustrated mathematically in Yoon & Millsap (2007).

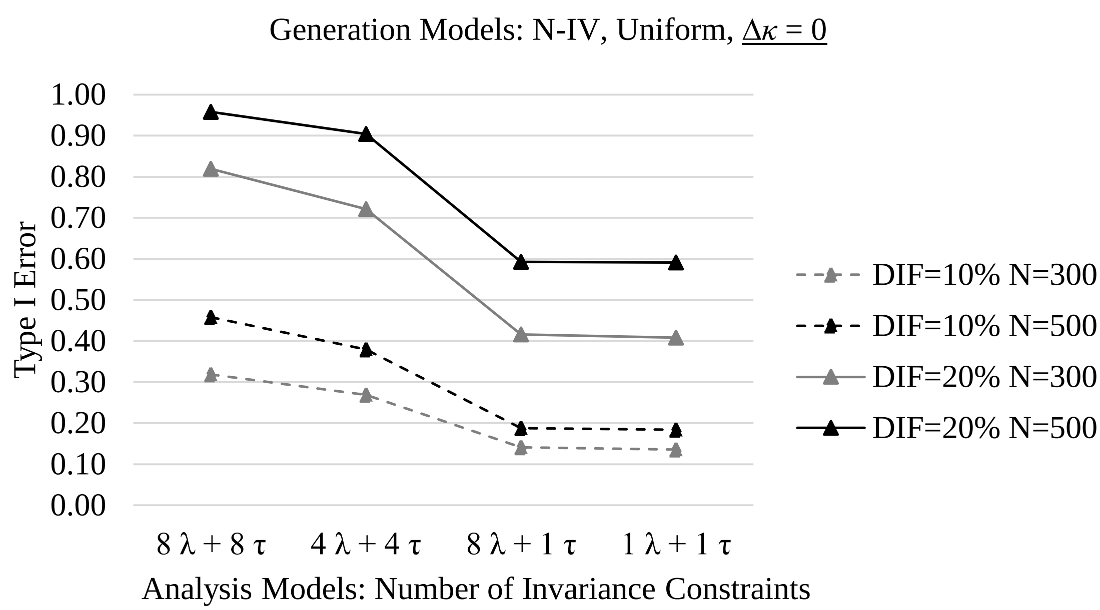


Figure 4. Type I Error Rates Associated with Factor Mean Difference Testing from Various Analysis Models When Data Are Generated from Uniform N-IV Conditions.

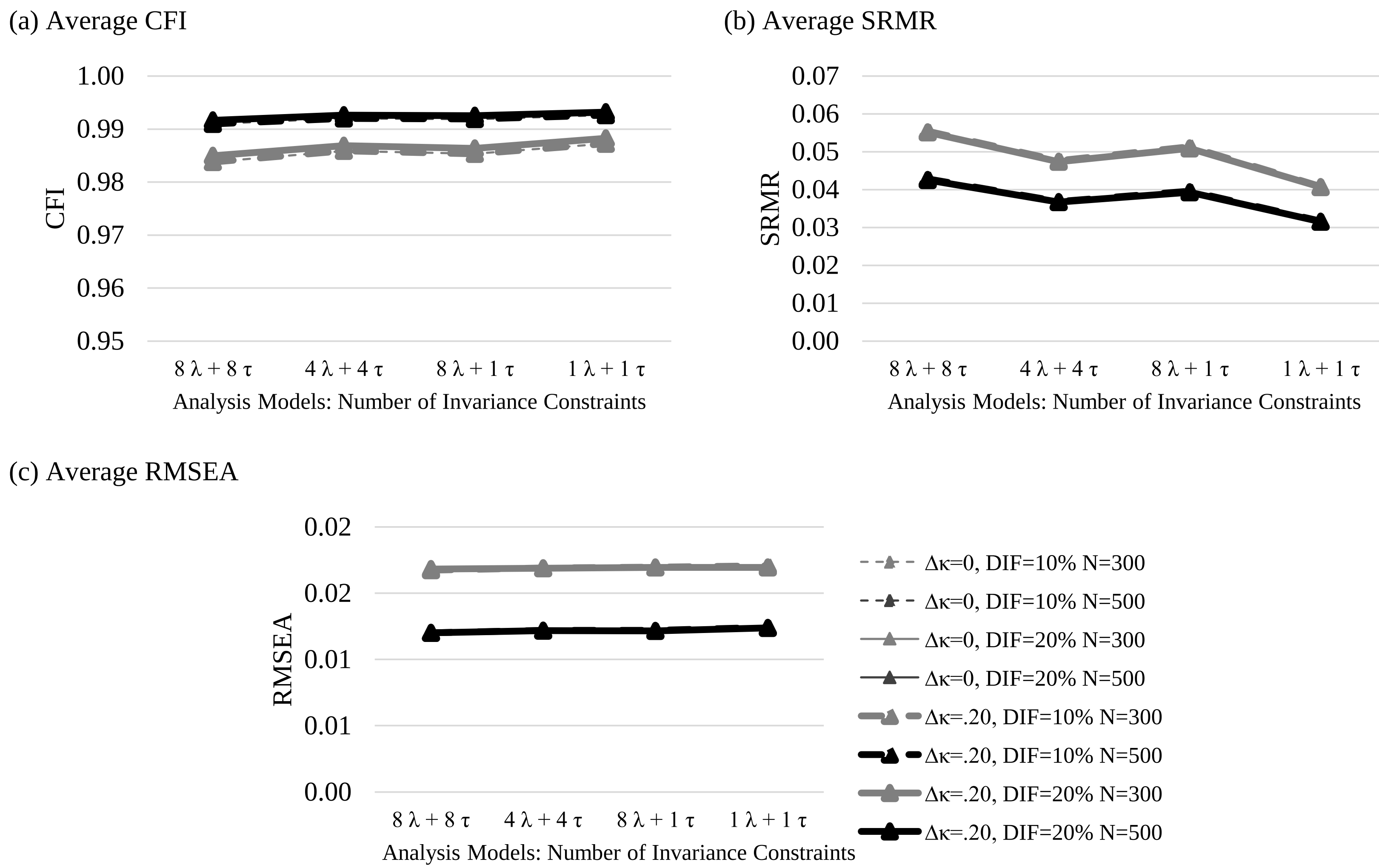


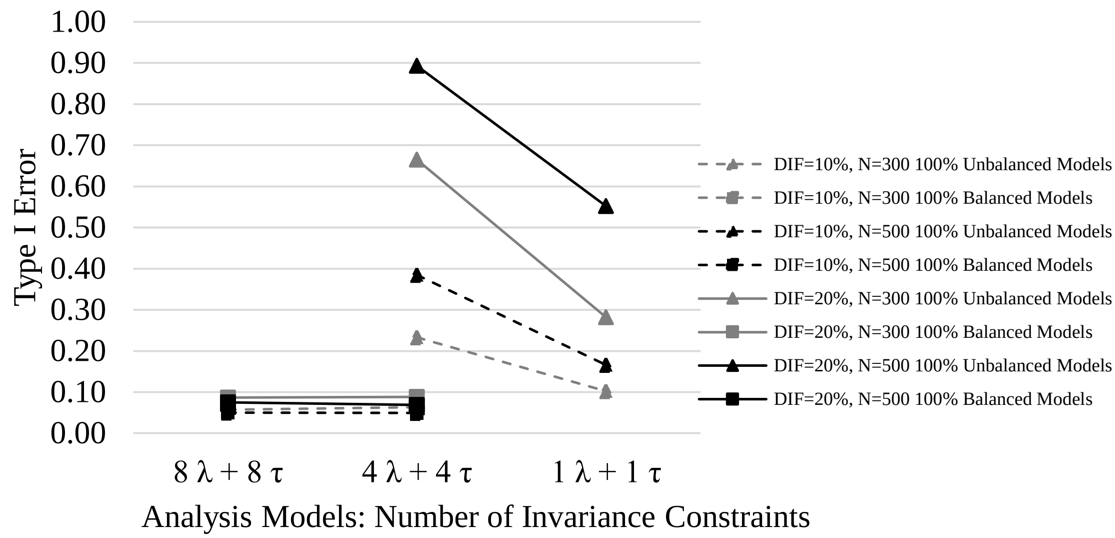
Figure 5. Model Fit Indices from Various Analysis Models When Data Are Generated from Uniform N-IV Conditions.

Mixed Non-Invariance Conditions (Mixed N-IV)

For mixed N-IV generated data, half of the non-invariant loadings (intercepts) were higher (lower) in group 1 and half were higher in group 2. Analysis models for mixed N-IV data were either *100% unbalanced models* or *100% balanced models* depending on the number and specification of invariance constraints (see Table 3). For ease of interpretation, the model with 8 loading constraints and 1 intercept constraint was excluded from the figures. As a result, for mixed N-IV data, there were four analysis models, two of them being *100% balanced models* and two being *100% unbalanced models*.

Figure 6 displayed Type I error rates and power for the different analysis models. In panel (a), Type I errors clustered around .05 for *100% balanced analysis models*, regardless of the number of invariance constraints. On the other hand, Type I error rates were inflated to varying degrees for *100% unbalanced analysis models*. For either *100% balanced* or *unbalanced models*, the magnitude of inflation of Type I error rate correlated positively with DIF. For *100% balanced models*, inflation of Type I error rates was negatively related to sample size; the correlation turned positive when *100% unbalanced models* were fitted. In panel (b), the powers for the *100% balanced analysis models* (with 4 or 8 constraints imposed on the loadings and intercept constraints) were comparable to the powers for the same models fitting the F-IV data. Powers for the *100% unbalanced analysis model* were much lower than those for the *100% balanced model*. The constrained indicators in the *100% unbalanced analysis models* have higher intercepts (although lower loadings) in group 1 than in group 2. But the factor means were simulated to be higher in group 2. As a result, difference in constrained intercepts were

(a) Generation Models: N-IV, Mixed, $\underline{\Delta\kappa = 0}$



(b) Generation Models: N-IV, Mixed, $\underline{\Delta\kappa = .20}$

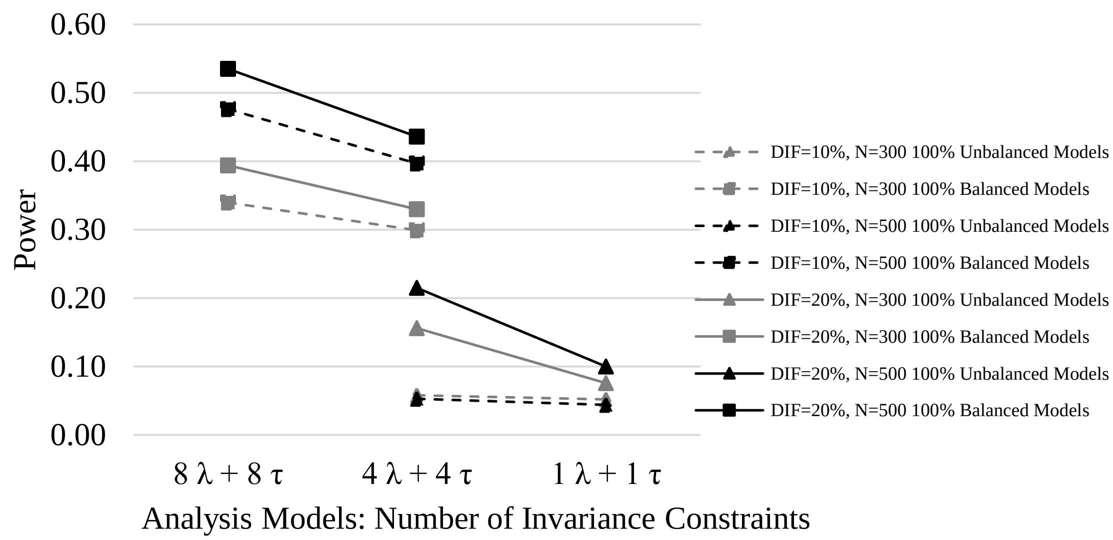


Figure 6. Type I Error Rates and Power Associated with Factor Mean Difference Testing from Various Analysis Models When Data Are Generated from Mixed N-IV Conditions.

transferred to factor means, but in an opposite direction, leading to loss in power. Still, conditions with higher DIF and sample size tended to have higher power. Table 11 in Appendix A showed how average bias, effect size, and efficiency performed with varying models. As expected, bias and effect size were relatively minimal when analysis models were *100% balanced models*; while the estimations were highly inflated when analysis models were *100% unbalanced*. With either *100% unbalanced* or *100% balanced models*, these quantities became less optimal as invariance constraints were removed, regardless of the appropriateness.

For simplicity, Figure 7 presents model fit indices for the four analysis models fitted to data generated with $\Delta\kappa = 0$, DIF = 20% N = 300. Complete results of fit indices for all conditions can be found in Table 10 in Appendix A. All three fit indices indicated adequate fit when analyzed with *100% unbalanced models*, regardless of the number of constraints. In contrast, the *100% balanced models* fit the data poorly. With 8 loading and 8 intercept constrained, the mean CFI was .82, the mean SRMR was .10, and the mean RMSEA was .08, all exceeding the commonly accepted cutoff criteria of good fit (Hu & Bentler, 1999). Also, the *100% balanced model* with 4 loading and intercept constraints fit significantly worse than the *100% unbalanced model* with 4 invariance loading and intercept constraints. It should be noted that these models had the same amount of misspecification. Consistent with previous conditions, for *100% unbalanced models*, removing inappropriate constraints led to increasing CFIs and decreasing SRMRs, with RMSEA remaining at approximately the same level. For *100% balanced models*, all three indices indicated better model fit as inappropriate constraints were removed.

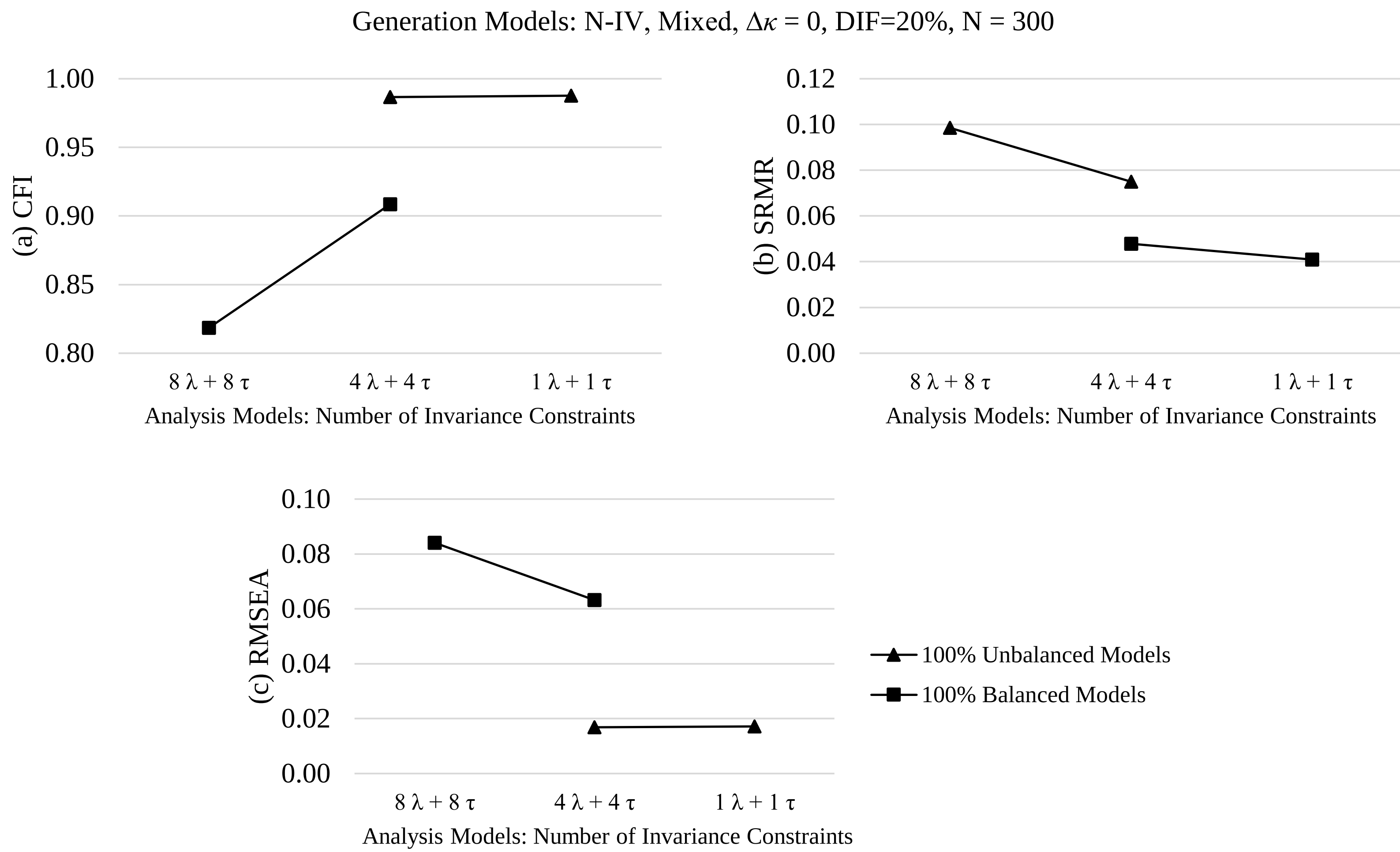


Figure 7. Model Fit Indices from Various Analysis Models When Data Are Generated from Mixed N-IV Conditions.

Uniform Partial Invariance Conditions (Uniform P-IV)

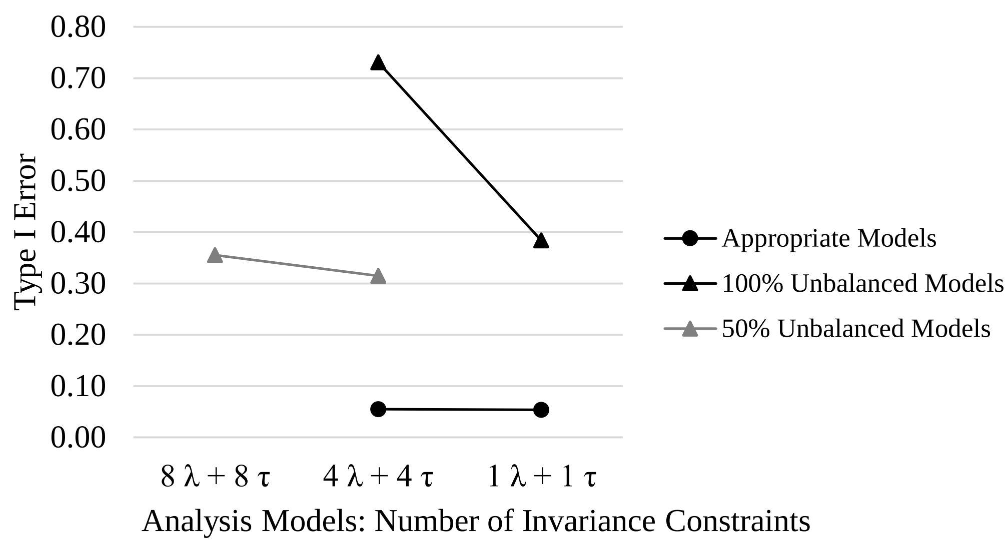
With uniform P-IV generated data, half of the indicators were invariant and half had greater loadings (smaller intercepts) in group 1. As specified in Table 4, a total of eight analysis models were conducted on each simulated dataset, including *appropriate models*, *50% unbalanced models* and *100% unbalanced models*. Results for six of the eight analysis models for these data were plotted in Figures 8 and 9. Complete results can be found in Table 12 and Table 13 in Appendix A.

As expected, Type I error, power, and the model fit indices for *appropriate models* analyzing uniform P-IV data (1 or 4 invariance constrains) were similar to those for the *appropriate models* analyzing F-IV data. Type I errors were at acceptable levels and model fit indices indicated good fit. For *100% unbalanced models*, Type I error rates were highly inflated, and powers were artificially high due to the inappropriate constraints on non-invariant loadings and intercepts. Consistent with the results of model fit from F-IV and uniform N-IV data, *100% unbalanced models* fitted uniform P-IV data as well as *appropriate models* when they had the same number of constraints. Results from the *50% unbalanced analysis models* were similar to those for the *100% unbalanced models*, but with weaker effects in magnitude.

Mixed Partial Invariance Conditions (Mixed P-IV)

Results for mixed P-IV were consistent with those based on data for other generation models. For mixed P-IV, half of the indicators were invariant and half of them were non-invariant in a mixed pattern. Nine different models were applied to each mixed P-IV dataset, as detailed in Table 4. All five types of analysis models were included and compared for these data: *appropriate models*, *50%* and *100% unbalanced models*, and

(a) Generation Model: P-IV, Uniform, DIF = 20%, $\underline{\Delta\kappa = 0}$, N = 300



(b) Generation Model: P-IV, Uniform, DIF = 20%, $\underline{\Delta\kappa = .20}$, N = 300

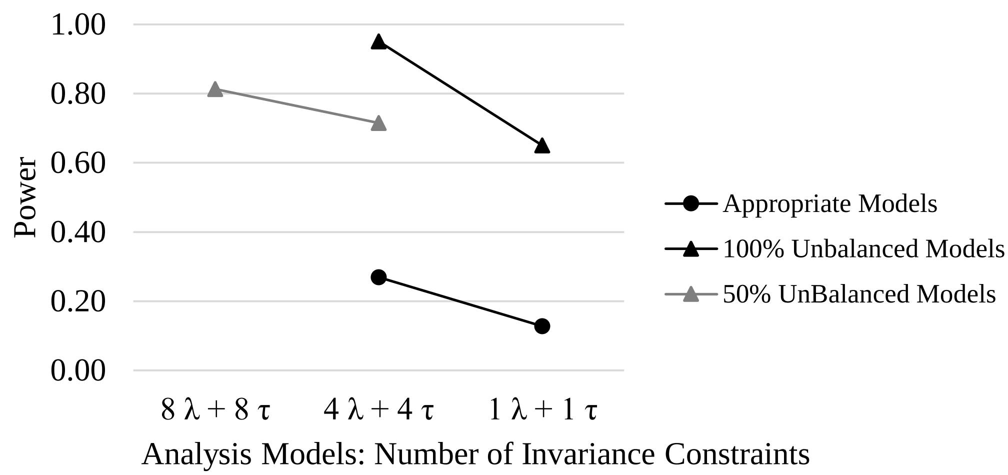


Figure 8. Type I Error Rates and Power Associated with Factor Mean Difference Testing from Various Analysis Models When Data Are Generated from Uniform P-IV Conditions.

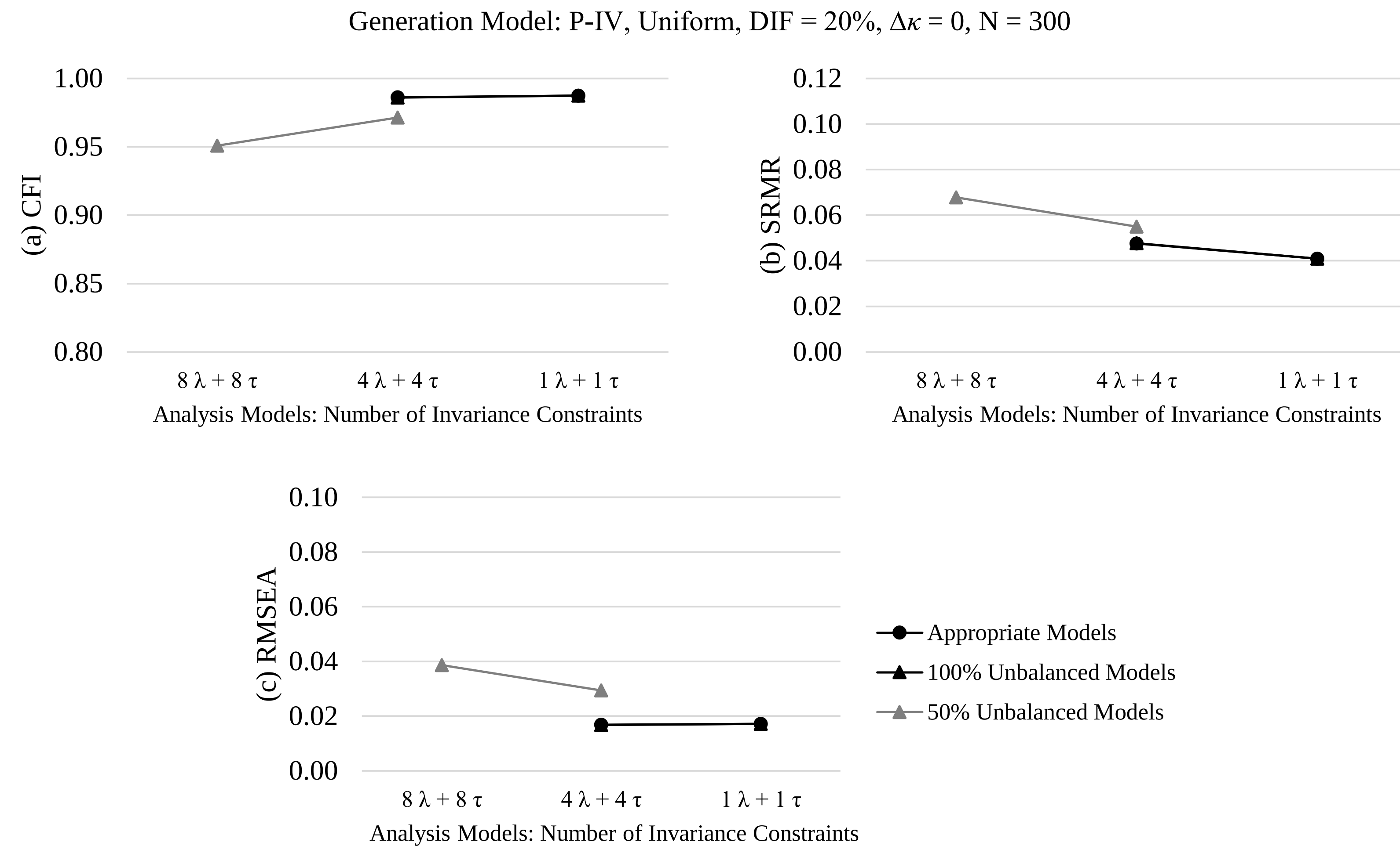


Figure 9. Model Fit Indices from Various Analysis Models When Data Are Generated from Uniform P-IV Conditions.

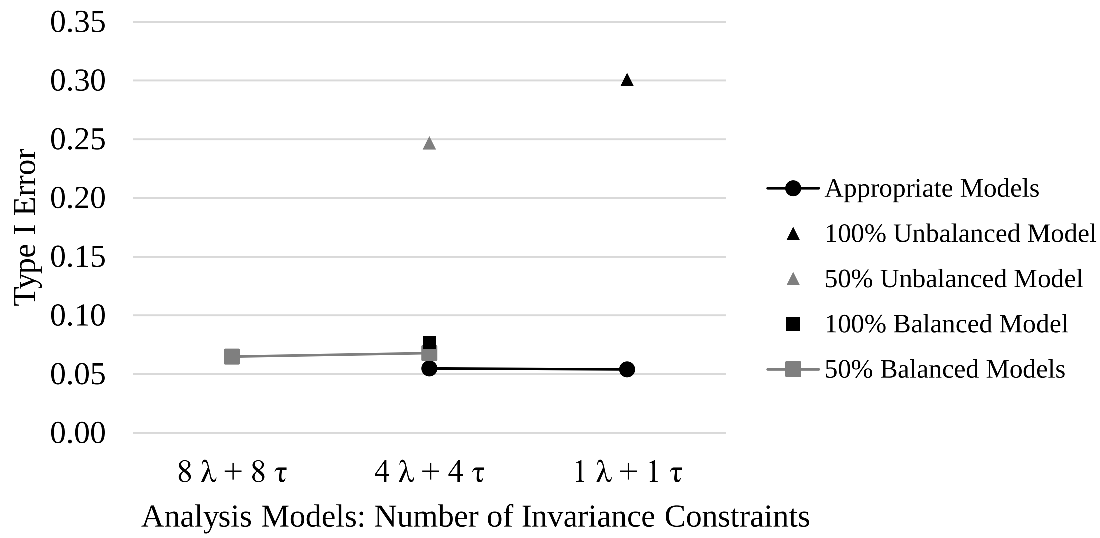
50% and *100% balanced models*. Figures 10 and 11 display results for varying analysis models when data were generated with DIF = 20% and N = 300. Table 14 and Table 15 in Appendix A included results from all generation conditions under mixed P-IV.

As shown in panel (a) of Figure 10, Type I error rates ranged from .054 to .077 when analysis models were *appropriate models* and *balanced models*. In contrast, *50%* and *100% unbalanced models* had inflated Type I error rates, with the alphas for *100% unbalanced model* showing greater inflation. With the same number of invariance constraints (4 constraints), Type I error rates were controlled best when analyses were conducted with an *appropriate model* (α = .055). Type I error rates for the *50%* and *100% balanced model* were somewhat inflated (α = .068 and .077, respectively), and Type I error for the *50% unbalanced model* was most highly inflated (α = .247).

Powers for *100% balanced models* were close to the powers for *appropriate model*s given the same number of constraints. The low power rates for *50%* and *100% unbalanced models* were due to the inappropriate constraints on the parameters for the non-invariant indicators, with higher intercepts (and smaller loadings) in group 1.

As illustrated in previous conditions, *unbalanced models* fit the data as well as appropriate models; and *balanced models* fit the data relatively poorly (see Figure 11). Specifically, given the same number of invariance constraints (1 or 4 constraints), *100% unbalance model* fit the data as well as the *appropriate model*; *50% unbalanced model* fitted worse than the *appropriate model* and the *100% unbalanced model*; the *50% balanced model* fit worse than the *50% unbalanced model*; and the *100% balanced model* fit worst. The results indicated that, with the same amount misspecification, a model has Type I error rates closer to the nominal level and more substantial power if the inappropriately constrained non-invariance have a mixed pattern, although the model will suffer from a poor fit

(a) Generation Model: P-IV, Mixed, DIF = 20%, <u>Δκ = 0</u>, N = 300



(b) Generation Model: P-IV, Mixed, DIF = 20%, <u>Δκ = .20</u>, N = 300

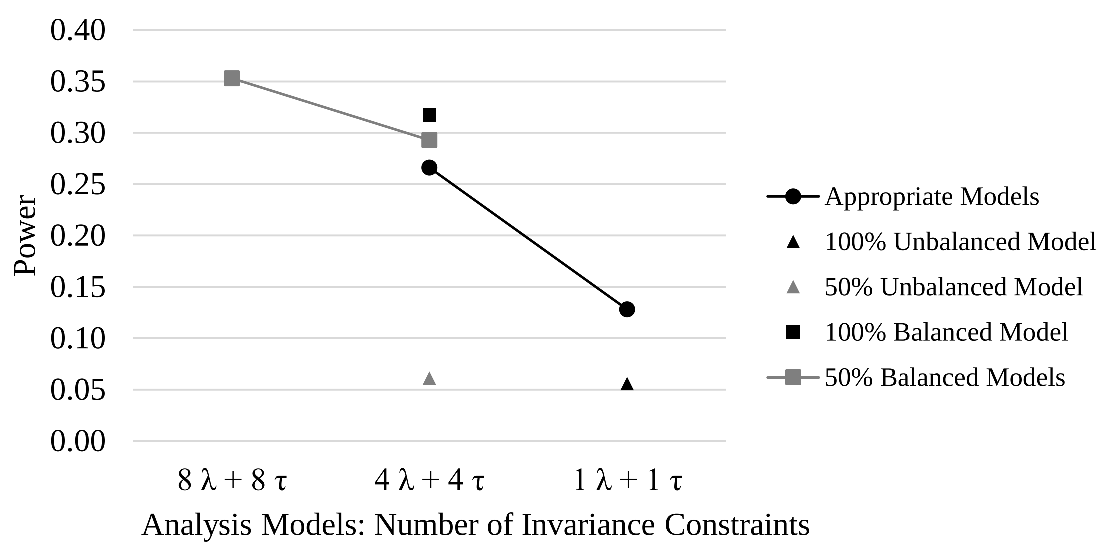


Figure 10. Type I Error Rates and Power Associated with Factor Mean Difference Testing from Various Analysis Models When Data Are Generated from Mixed P-IV Conditions.

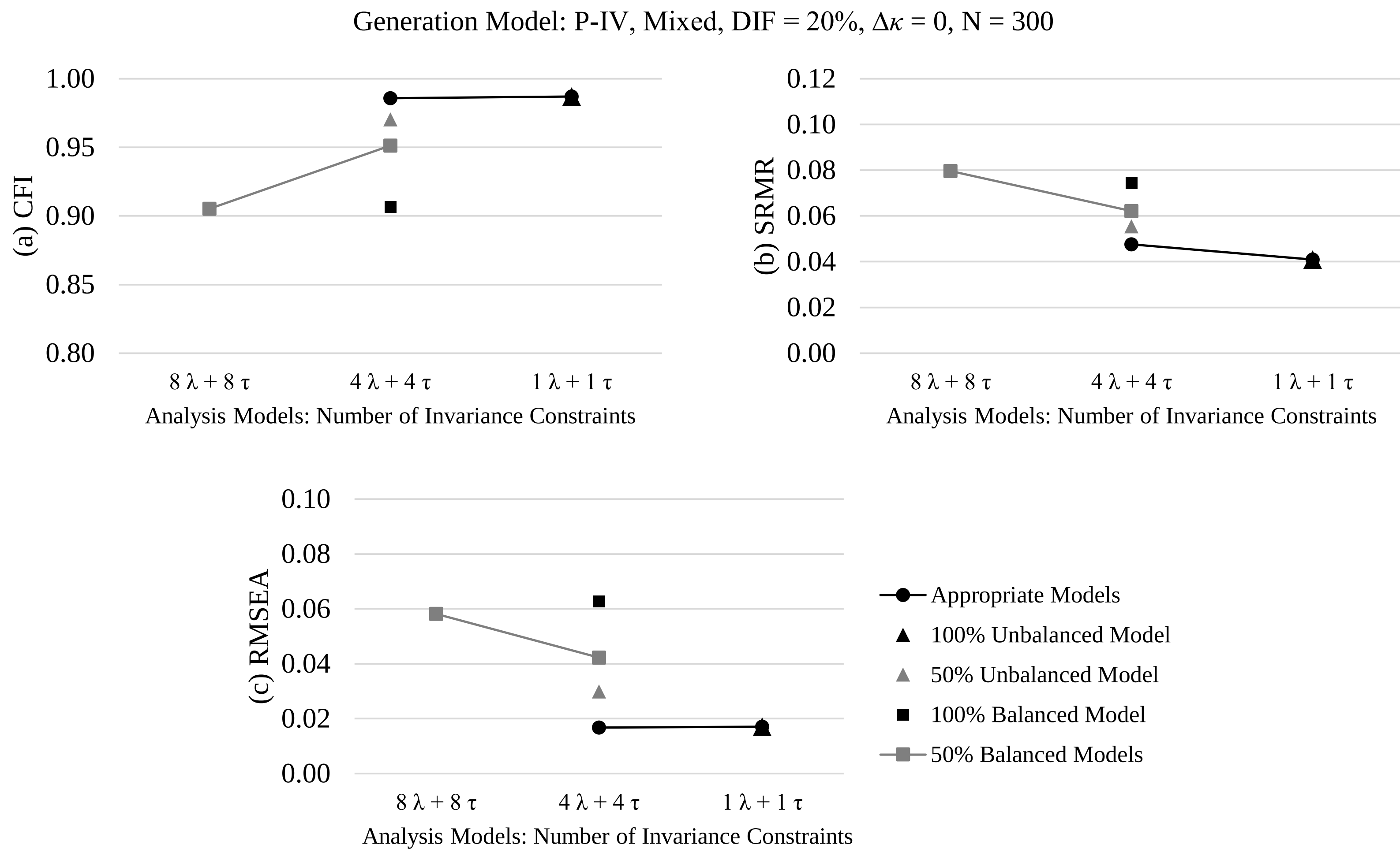


Figure 11. Model Fit Indices from Various Analysis Models When Data Are Generated from Mixed P-IV Conditions.

Summary of Results

Analysis models were categorized into five types according to the amount and pattern of misspecification. It was found that analysis model type affected the error rates, estimation of factor mean differences, and model fit. A brief summary comparing the results from different model types were presented in Table 5. The 50% and 100% models were collapsed because they had similar results patterns but only differed in magnitude.

Table 5

Results Summary by Model Types

		Type I Error	Power	Estimate Bias	Model Fit
Appropriate Models	Highly Constrained	Acceptable	Sufficient	Not biased	Adequate
	Remove Constraints	Relatively unaffected	Decreased substantially	Slightly increased	Relatively unaffected
Unbalanced Models	Highly Constrained	Inflated	-	Biased	Adequate
	Remove Constraints	Inflation decreased	-	Increased	Relatively unaffected
Balanced Models	Highly Constrained	Acceptable	Sufficient	Not biased	Inadequate
	Remove Constraints	Relatively unaffected	Decreased substantially	Slightly increased	Improved

As shown in the first section of Table 5, *appropriate models* had well-controlled Type I error rates, sufficient power, non-biased estimates, and adequate model fit when the model is highly constrained. As invariance constraints were removed in analysis model, these properties stay relatively unaffected except that power rates decreased substantially.

Unbalanced models had inflated Type I error rates, spuriously high or low level of powers, and highly biased estimates of factor mean differences. Removing inappropriate constraints led to decreases in inflation of Type I error rates but more inflated bias. The inflation cannot be eliminated unless the measures were correctly scaled. Model fit of *unbalanced models* was as good as *appropriate models* given the same number of invariance constraints.

Balanced models also had well-controlled Type I error rates, sufficient level of power, and relatively unbiased estimates, as for *appropriate models*. However model fit for *balanced models* was relatively poor. As inappropriate invariance constraints were removed, model fit improved; Type I error, power, and bias behaved the same as for *appropriate models*.

Across all generation models, invariance constraints on intercepts had a greater effect on factor mean differences than invariance constraints on loadings. Also, for uniform non-invariant or partial non-invariant data, greater DIF was correlated with higher inflated Type I error rates, greater bias, and greater inflation of effect size. Increase in sample size led to higher power and stable estimates when models were appropriate or balanced. When analysis models were unbalanced, increase in sample size exacerbated problem in estimating and testing factor mean differences.

CHAPTER 4

DISCUSSION

Previous studies on testing factor mean differences using SEM have been focusing on the impact of different levels and patterns of non-invariance in generated data (Chen, 2008; Kaplan & George, 1995; Wang et al., 2012; Whittaker, 2013). One common feature of these studies is that the analysis models fit to data were fixed in terms of their specifications of invariance constraints. Results from the studies informed researchers the robustness of factor mean difference testing when the invariance assumption is violated. The current study differed from previous ones in that it focused on the impact of different specifications of invariance constraints for analysis models, given various patterns of non-invariance in generated data. The generation models and analysis models investigated in the study mimicked the typical model misspecification types in practice. Therefore, the results should inform researchers who are conducting specification searches and are facing with choosing among different model specifications for their decision making.

Researchers who wish to test between-group differences in means on factors need to know whether the indicators of these factors are invariant across groups so that they can make informed choices about model specification (Green & Thompson, 2010; Horn & McArdle, 1992; Meredith, 1993). SEM can be conducted to assess measurement invariance, and the statistical results from this assessment can be combined with substantive knowledge to determine the specification of the model that is used to test factor mean differences. Potentially though, absence of related theory and potential problems with statistical analyses involving measurement invariance might discourage researchers from conducting tests of factor means, as a limited number of studies were found to apply such strategies on real data analysis (as opposed to simulation data) (Marsh & Grayson, 1994; Reise et al., 1993; Widaman & Reise, 1997). As an alternative, they may choose to conduct MANOVA and multiple ANOVAs because these methods are "less controversial" and thus are "more comfortable". However, the use of MANOVA and multiple ANOVAs is even more problematic in that researchers are likely to face similar issues, knowingly or unknowingly, when they are reaching conclusions about the constructs underlying their measures, but in an ad hoc manner. Accordingly, the goal of my study was to develop guidelines that would be helpful to researchers when they conduct factor mean differences so that they find the SEM approach less onerous and will choose it over MANOVA and multiple ANOVAs. Based on the results of my study, several suggestions are offered for assessing measurement invariance in the test of factor mean differences across groups.

First, factor indicators should be carefully selected when designing a study to minimize potential problems with equivalency in loadings and intercepts across groups. Based on the current study and the previous ones (Chen, 2008; Kaplan & George, 1995; Wang et al., 2012; Whittaker, 2013), non-invariant loadings and/or intercepts is likely to lead to biased estimates of factor mean differences and other undesired model estimation properties if researchers fail to adequately model non-invariance in parameters. The effect of misspecification of models with non-invariance of indicator intercepts is particularly problematic on the assessment of factor mean differences. The implication is that a few "good" indicators (i.e., with minimal differences in loadings and intercepts) is preferable to a few "good" indicators and a bunch of "mediocre or bad" indicators. When such indicators are not available in practice, it is suggested to select indicators based on a broad spectrum of cross-group differences, rather than to use indicators that uniformly favor one of the groups over the other(s). Researchers might choose a collection of indicators that are balanced across groups in the sense that positive differences in parameters with respect to one group are offset by negative differences in the parameters with respect to the same group. If successful, the benefit is maintaining the nominal Type I error rate associated with testing factor mean differences and minimizing bias in mean difference estimates.

Second, decisions about measurement invariance should not be based only on statistical results. Results from this study indicate that a pure empirically driven decision process is dangerous in two ways. First, adequate model fit does not guarantee the appropriateness of a model. As shown by the current study and Yoon & Millsap (2007), the inappropriateness of invariance constraints will not reflected by model fit when the constrained indicators are uniformly non-invariant. Consequently, applied researchers are likely to identify uniformly non-invariant indicators as invariant under these conditions. Unfortunately, this type of misspecification due to scale readjustment cannot be detected by using purely modeling strategies, as illustrated by Hancock et al. (2009). Researchers must rely on their understanding of substantive theories to make judgment about indicators. Second, inadequate model fit does not lead necessarily to appropriate model modification either. As shown, appearance of inadequate model fit indicated that at least two of the three types of indicator are present: invariant indicators, indicators with non-invariance in one direction, indicators with non-invariance in the opposite direction. Correct identification of non-invariance relies heavily on whether or not one has a truly invariant referent indicator. Thus, researchers are recommended to consider all potential conditions of indicator compositions that might lead to the observed results, and compare these considerations with their substantive knowledge and experience to make a specification judgment.

In summary, we suggest researchers to integrate the thought processes of choosing measurement indicators and specifying invariance constraints when testing factor mean differences across groups. Before gathering measures to form an instrument and data collection, the invariance properties of all available indicators should be thoroughly studied in the context theory and based on previous related studies. With a preliminary understanding of the indicators, the assessment of measurement invariance becomes a tool to verify researchers' assumptions rather than post hoc decisions about model respecification (MacCallum, 1986; Vandenberg & Lance, 2000). Observed results should be compared to the researchers' hypotheses about the composition of invariant and non-invariant indicators to see if they are consistent. For example, if the indicators were purposely selected to have a mixture direction of non-invariance, relatively poor model fit should be expected and invariance constraints should be respecified with fully knowing what statistical consequences might occur. On the other hand, it is possible in practice that all available indicators uniformly favor one population group relative to others, such as in many cross-cultural studies (Chen, 2008; van de Vijver & Leung, 2000; Vandenberg & Lance, 2000). In this situation, a deceivingly good model fit should be expected, and researchers should be able to realize that it is not an assurance of measurement invariance.

Researchers are also recommended to try models with different specifications of invariance constraints along the process of testing measurement invariance, especially when the indicators' invariance properties are not well known. In such cases, switching referent indicators might be helpful in identifying the patterns of invariance and non-invariance in a relative sense. Also, trying out different specifications on invariance constraints enables researchers to compare the results with their assumptions about the indicators so that decisions based on a systematic thinking process could be made.

In the future this study could be extended in a variety of ways. First, fitting models with decreasing numbers of invariance constraints to datasets is not the same as conducting a model search process. Comparison of statistics and parameter estimates from the independent models in my study may not reflect changes that occur in empirical specification searches. Future simulation studies should include specification searches to explore more fully decisions that are made in the modification process. In addition, future studies should include a more appropriate CFI in their analyses rather than the one generated automatically by the software. Calculation of CFI is based on the increment of the target model in comparison with the null model. The null model is specified to allow variances and means of manifest variables to be freely estimated in most software. However the default standard null model is inappropriate because this null model is not nested in models with invariance constraints on both loadings and intercepts (Widaman & Thompson, 2003). As a result, these incorrect CFIs are likely to be slightly liberal. Finally, the study only investigated generation conditions with equal sample sizes and small non-invariance between groups. Future studies should consider conditions with unequal sample sizes and greater non-invariance in loadings and intercepts.

REFERENCES

Bagozzi, R. P., & Edwards, J. R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods*, *1*(1), 45-87.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMIPLIS: Basic concepts, applications, and programming.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456.

Byrne, B. M., & Stewart, S. M. (2006). Teacher's corner: the MACS approach to testing for multigroup invariance of a second-order structure: a walk through the process. *Structural Equation Modeling*, *13*(2), 287-321.

Carle, A. C., Millsap, R. E., & Cole, D. A. (2008). Measurement Bias across Gender on the Children's Depression Inventory Evidence for Invariance from Two Latent Variable Models. *Educational and Psychological Measurement*, *68*(2), 281-303.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of personality and social psychology*, *95*(5), 1005-1018.

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

Cole, D. A., Maxwell, S. E., Arvey, R., & Salas, E. (1993). Multivariate group comparisons of variable systems: MANOVA and structural equation modeling. *Psychological Bulletin, 114*, 174-184.

De Beuckelaer, A., & Swinnen, G. (2011). Biased latent variable mean comparisons due to measurement non-invariance: A simulation study. *Methods and applications in cross-cultural analysis*, 117-148.

French, B. F., & Finch, W. H. (2008). Multigroup confirmatory factor analysis: Locating the invariant referent sets. *Structural Equation Modeling*, *15*(1), 96-113.

Green, S. B., & Babyak, M. A. (1997). Control of Type I errors with multiple tests of constraints in structural equation modeling. *Multivariate Behavioral Research*, *32*(1), 39-51.

Green, S.B., & Thompson, S.M. (2012). A flexible structural equation modeling approach for analyzing means. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 393-416). New York: Guilford Press.

Hancock, G. R. (1997). Structural equation modeling methods of hypothesis testing of latent variable means. *Measurement and Evaluation in Counseling and Development*, *30*, 91-105.

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*(3), 373-388.

Hancock, G. R. (2004). Experimental, quasi-experimental, and nonexperimental design and analysis with latent variables. In D. Kaplan (Ed.), *The SAGE Handbook of Quantitative Methodology for the Social Sciences.* Thousand Oaks, CA: SAGE Publications.

Hancock, G. R., Lawrence, F. R., & Nevitt, J. (2000). Type I error and power of latent mean methods and MANOVA in factorially invariant and noninvariant latent variable systems. *Structural Equation Modeling: A Multidisciplinary Journal, 7*, 534-556.

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*(3), 117-144.

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

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 70*, 631-639.

Kaplan, D. (1989). Power of the likelihood ratio test in multiple group confirmatory factor analysis under partial measurement invariance. *Educational and psychological measurement*, *49*(3), 579-586.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions.* Thousand Oaks, CA: Sage Publications.

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

MacCallum, R. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin, 100*(1), 107.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*(3), 490-504.

Marsh, H. W., & Grayson, D. (1994). Longitudinal stability of latent means and individual differences: A unified approach. *Structural Equation Modeling: A Multidisciplinary Journal*, *1*(4), 317-359.

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First-and higher-order factor models and their invariance across groups. *Psychological bulletin*, *97*(3), 562-582.

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

Millsap, R. E. (1997). Invariance in measurement and prediction: Their relationship in the single-factor case. *Psychological Methods*, *2*(3), 248.

Millsap, R. E. (1998). Group differences in regression intercepts: Implications for factorial invariance. *Multivariate Behavioral Research*, *33*(3), 403-424.

Millsap, R. E., & Hartog, S. B. (1988). Alpha, beta, and gamma change in evaluation research: A structural equation approach. *Journal of Applied Psychology*, *73*(3), 574.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological bulletin*, *114*(3), 552-566.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27*, 229-239.

Steenkamp, J-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25,* 78-90.

Van de Vijver, F. J., & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of cross-cultural psychology*, *31*(1), 33-51.

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*(1), 4-70.

Wang, D., Whittaker, T.A., & Beretvas, S.N. (2012). The impact of violating factor scaling method assumptions on latent mean difference testing in structured means models. *Journal of Modern Applied Statistical Methods, 11*(1), 24-41.

Whittaker, T. A. (2013). The Impact of Noninvariant Intercepts in Latent Means Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(1), 108-130.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. *The science of prevention: Methodological advances from alcohol and substance abuse research*, 281-324.

Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological methods*, *8*(1), 16.

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

APPENDIX A

COMPLETE RESULTS FROM ALL CONDITIONS

Table 6

Model Fit Indices, Type I and Type II Error Rates Associated with Factor Mean Difference Testing, and Estimates of Factor Mean Differences When Generation Models are F-IV (All Analysis Models are Appropriate Models)

1. Generation Model: F-IV, $\Delta\kappa = 0$, $N = 300$

	Total Invariance Constraints	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Type I Error Rates
Analysis Models	$8\ \lambda + 8\ \tau$	54	.9834	.0166	.0554	.0039	.0540
	$4\ \lambda + 4\ \tau$	46	.9854	.0168	.0477	.0053	.0550
	$1\ \lambda + 1\ \tau$	40	.9867	.0171	.0410	.0168	.0540
	$8\ \lambda + 1\ \tau$	47	.9848	.0170	.0514	.0167	.0650

2. Generation Model: F-IV, $\Delta\kappa = .20$, $N = 300$

	Total Invariance Constraints	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Power Rates
Analysis Models	$8\ \lambda + 8\ \tau$	54	.9833	.0166	.0555	.2071	.3300
	$4\ \lambda + 4\ \tau$	46	.9854	.0168	.0478	.2093	.2680
	$1\ \lambda + 1\ \tau$	40	.9867	.0171	.0410	.2263	.1280
	$8\ \lambda + 1\ \tau$	47	.9848	.0170	.0515	.2231	.1440

3. Generation Model: F-IV, $\Delta\kappa = 0$, $N = 500$

	Total Invariance Constraints	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Type I Error Rates
Analysis Models	$8\ \lambda + 8\ \tau$	54	.9907	.0120	.0427	-.0006	.0450
	$4\ \lambda + 4\ \tau$	46	.9917	.0123	.0369	.0012	.0440
	$1\ \lambda + 1\ \tau$	40	.9923	.0124	.0317	-.0004	.0510
	$8\ \lambda + 1\ \tau$	47	.9915	.0123	.0397	.0006	.0570

4. Generation Model: F-IV, $\Delta\kappa = .20$, $N = 500$

	Total Invariance Constraints	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Power Rates
Analysis Models	$8\ \lambda + 8\ \tau$	54	.9907	.0120	.0429	.2012	.4580
	$4\ \lambda + 4\ \tau$	46	.9917	.0122	.0369	.2033	.3850
	$1\ \lambda + 1\ \tau$	40	.9923	.0124	.0317	.2045	.1770
	$8\ \lambda + 1\ \tau$	47	.9915	.0123	.0398	.2032	.1900

Table 7

Estimates of Factor Mean Differences, Bias, Relative Bias, Efficiency, and Effect Size of Estimated Factor Mean Differences When Generation Models are F-IV (All Analysis Models are Appropriate Models)

1. Generation Model: F-IV, $\Delta\kappa = 0$, $N = 300$							
	Total Invariance Constraints	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	$8\ \lambda + 8\ \tau$	54	.0039	.0039	-	.1373	.0031
	$4\ \lambda + 4\ \tau$	46	.0053	.0053	-	.1560	.0041
	$1\ \lambda + 1\ \tau$	40	.0168	.0168	-	.2454	.0121
	$8\ \lambda + 1\ \tau$	47	.0167	.0167	-	.2392	.0120

2. Generation Model: F-IV, $\Delta\kappa = .20$, $N = 300$							
	Total Invariance Constraints	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	$8\ \lambda + 8\ \tau$	54	.2071	.0071	3.5470	.1383	.1454
	$4\ \lambda + 4\ \tau$	46	.2093	.0093	4.6615	.1575	.1463
	$1\ \lambda + 1\ \tau$	40	.2263	.0263	13.1505	.2493	.1552
	$8\ \lambda + 1\ \tau$	47	.2231	.0231	11.5260	.2413	.1564

3. Generation Model: F-IV, $\Delta\kappa = 0$, $N = 500$							
	Total Invariance Constraints	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	$8\ \lambda + 8\ \tau$	54	-.0006	-.0006	-	.1056	-.0004
	$4\ \lambda + 4\ \tau$	46	.0012	.0012	-	.1196	.0009
	$1\ \lambda + 1\ \tau$	40	-.0004	-.0004	-	.1848	.0003
	$8\ \lambda + 1\ \tau$	47	.0006	.0006	-	.1818	.0002

4. Generation Model: F-IV, $\Delta\kappa = .20$, $N = 500$							
	Total Invariance Constraints	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	$8\ \lambda + 8\ \tau$	54	.2012	.0012	.5930	.1063	.1417
	$4\ \lambda + 4\ \tau$	46	.2033	.0033	1.6605	.1207	.1429
	$1\ \lambda + 1\ \tau$	40	.2045	.0045	2.2715	.1872	.1422
	$8\ \lambda + 1\ \tau$	47	.2032	.0032	1.6120	.1831	.1429

Table 8

Model Fit Indices, Type I and Type II Error Rates Associated with Factor Mean Difference Testing, and Estimates of Factor Mean Differences When Generation Models are Uniform N-IV (All Analysis Models are 100% Unbalanced Models)

1. Generation Model: Uniform N-IV, $\Delta\kappa = 0$, DIF = 10%, $N = 300$							
	Total Invariance Constraints	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Type I Error Rates
Analysis Models	$8\ \lambda + 8\ \tau$	54	.9838	.0167	.0552	.1852	.3190
	$4\ \lambda + 4\ \tau$	46	.9859	.0169	.0475	.1868	.2690
	$1\ \lambda + 1\ \tau$	40	.9873	.0171	.0408	.2002	.1360
	$8\ \lambda + 1\ \tau$	47	.9853	.0170	.0512	.1988	.1410

2. Generation Model: Uniform N-IV, $\Delta\kappa = .20$, DIF = 10%, $N = 300$							
	Total Invariance Constraints	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Power Rates
Analysis Models	$8\ \lambda + 8\ \tau$	54	.9838	.0167	.0555	.3482	.7900
	$4\ \lambda + 4\ \tau$	46	.9859	.0169	.0476	.3502	.7000
	$1\ \lambda + 1\ \tau$	40	.9873	.0171	.0408	.3668	.3720
	$8\ \lambda + 1\ \tau$	47	.9853	.0170	.0513	.3641	.3840

3. Generation Model: Uniform N-IV, $\Delta\kappa = 0$, DIF = 20%, $N = 300$							
	Total Invariance Constraints	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Type I Error Rates
Analysis Models	$8\ \lambda + 8\ \tau$	54	.9850	.0168	.0549	.3284	.8190
	$4\ \lambda + 4\ \tau$	46	.9869	.0169	.0473	.3298	.7210
	$1\ \lambda + 1\ \tau$	40	.9883	.0169	.0407	.3431	.4080
	$8\ \lambda + 1\ \tau$	47	.9864	.0169	.0509	.3419	.4160

4. Generation Model: Uniform N-IV, $\Delta\kappa = .20$, DIF = 20%, $N = 300$							
	Total Invariance Constraints	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Power Rates
Analysis Models	$8\ \lambda + 8\ \tau$	54	.9849	.0168	.0552	.4581	.9740
	$4\ \lambda + 4\ \tau$	46	.9869	.0169	.0474	.4596	.9440
	$1\ \lambda + 1\ \tau$	40	.9883	.0169	.0407	.4748	.6750
	$8\ \lambda + 1\ \tau$	47	.9864	.0169	.0509	.4731	.6800

5. Generation Model: Uniform N-IV, $\Delta\kappa = 0$, DIF = 10%, $N = 500$							
	Total Invariance	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates	Type I Error

	Constraints					of Δκ	Rates
Analysis Models	8 λ + 8 τ	54	.9910	.0120	.0426	.1795	.4580
	4 λ + 4 τ	46	.9920	.0122	.0368	.1813	.3790
	1 λ + 1 τ	40	.9926	.0124	.0316	.1816	.1840
	8 λ + 1 τ	47	.9919	.0122	.0395	.1811	.1880

6. Generation Model: Uniform N-IV, Δκ = .20, DIF = 10%, *N* = 500							
	Total Invariance Constraints	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of Δκ	Power Rates
Analysis Models	8 λ + 8 τ	54	.9910	.0120	.0428	.3415	.9520
	4 λ + 4 τ	46	.9920	.0122	.0369	.3434	.8840
	1 λ + 1 τ	40	.9926	.0124	.0316	.3454	.5430
	8 λ + 1 τ	47	.9919	.0122	.0396	.3437	.5510

7. Generation Model: Uniform N-IV, Δκ = 0, DIF = 20%, *N* = 500							
	Total Invariance Constraints	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of Δκ	Type I Error Rates
Analysis Models	8 λ + 8 τ	54	.9917	.0120	.0423	.3219	.9580
	4 λ + 4 τ	46	.9926	.0122	.0366	.3235	.9040
	1 λ + 1 τ	40	.9932	.0124	.0315	.3245	.5910
	8 λ + 1 τ	47	.9925	.0122	.0392	.3237	.5930

8. Generation Model: Uniform N-IV, Δκ = .20, DIF = 20%, *N* = 500							
	Total Invariance Constraints	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of Δκ	Power Rates
Analysis Models	8 λ + 8 τ	54	.9917	.0120	.0425	.4509	1.0000
	4 λ + 4 τ	46	.9926	.0122	.0367	.4524	.9950
	1 λ + 1 τ	40	.9932	.0124	.0315	.4543	.8590
	8 λ + 1 τ	47	.9925	.0122	.0393	.4530	.8630

Table 9

Estimates of Factor Mean Differences, Bias, Relative Bias, Efficiency, and Effect Size of Estimated Factor Mean Differences When Generation Models are Uniform N-IV (All Analysis Models are 100% Unbalanced Models)

1. Generation Model: Uniform N-IV, $\Delta\kappa$ = 0, DIF = 10%, N = 300							
	Total Invariance Constraints	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	54	.1852	.1852	-	.1245	.1435
	4 λ + 4 τ	46	.1868	.1868	-	.1410	.1443
	1 λ + 1 τ	40	.2002	.2002	-	.2189	.1527
	8 λ + 1 τ	47	.1988	.1988	-	.2154	.1539

2. Generation Model: Uniform N-IV, $\Delta\kappa$ = .20, DIF = 10%, N = 300							
	Total Invariance Constraints	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	54	.3482	.1482	74.1040	.1265	.2697
	4 λ + 4 τ	46	.3502	.1502	75.0790	.1437	.2704
	1 λ + 1 τ	40	.3668	.1668	83.3945	.2250	.2794
	8 λ + 1 τ	47	.3641	.1641	82.0325	.2195	.2818

3. Generation Model: Uniform N-IV, $\Delta\kappa$ = 0, DIF = 20%, N = 300							
	Total Invariance Constraints	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	54	.3284	.3284	-	.1157	.2753
	4 λ + 4 τ	46	.3298	.3298	-	.1305	.2759
	1 λ + 1 τ	40	.3431	.3431	-	.1999	.2840
	8 λ + 1 τ	47	.3419	.3419	-	.1975	.2864

4. Generation Model: Uniform N-IV, $\Delta\kappa$ = .20, DIF = 20%, N = 300							
	Total Invariance Constraints	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	54	.4581	.2581	129.0675	.1180	.3839
	4 λ + 4 τ	46	.4596	.2596	129.8200	.1334	.3844
	1 λ + 1 τ	40	.4748	.2748	137.3845	.2057	.3929
	8 λ + 1 τ	47	.4731	.2731	136.5595	.2022	.3962

5. Generation Model: Uniform N-IV, $\Delta\kappa$ = 0, DIF = 10%, N = 500							
	Total Invariance	*df*	Average Estimates	Average Bias	Average Relative	Efficiency	Effect Size

	Constraints		of $\Delta\kappa$		Bias (%)		
Analysis Models	8 λ + 8 τ	54	.1795	.1795	-	.0958	.1396
	4 λ + 4 τ	46	.1813	.1813	-	.1083	.1407
	1 λ + 1 τ	40	.1816	.1816	-	.1658	.1399
	8 λ + 1 τ	47	.1811	.1811	-	.1638	.1405

6. Generation Model: Uniform N-IV, $\Delta\kappa$ = .20, DIF = 10%, N = 500							
	Total Invariance Constraints	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	54	.3415	.1415	70.7520	.0973	.2655
	4 λ + 4 τ	46	.3434	.1434	71.7010	.1103	.2665
	1 λ + 1 τ	40	.3454	.1454	72.6755	.1700	.2656
	8 λ + 1 τ	47	.3437	.1437	71.8600	.1667	.2669

7. Generation Model: Uniform N-IV, $\Delta\kappa$ = 0, DIF = 20%, N = 500							
	Total Invariance Constraints	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	54	.3219	.3219	-	.0891	.2706
	4 λ + 4 τ	46	.3235	.3235	-	.1003	.2716
	1 λ + 1 τ	40	.3245	.3245	-	.1519	.2705
	8 λ + 1 τ	47	.3237	.3237	-	.1504	.2719

8. Generation Model: Uniform N-IV, $\Delta\kappa$ = .20, DIF = 20%, N = 500							
	Total Invariance Constraints	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	54	.4509	.2509	125.4280	.0909	.3790
	4 λ + 4 τ	46	.4524	.2524	126.2160	.1025	.3799
	1 λ + 1 τ	40	.4543	.2543	127.1740	.1560	.3786
	8 λ + 1 τ	47	.4530	.2530	126.5175	.1537	.3806

Table 10

Model Fit Indices, Type I and Type II Error Rates Associated with Factor Mean Difference Testing, and Estimates of Factor Mean Differences When Generation Models are <u>*Mixed N-IV*</u>

1. Generation Model: Mixed N-IV, $\Delta\kappa$ = 0, DIF = 10%, N = 300

	Total Invariance Constraints	Constraints on non-invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Type I Error Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*100% Balanced*	54	.9479	.0396	.0685	.0027	.0570
	4 λ + 4 τ	4 λ and 4 τ	-	*100% Unbalanced*	46	.9859	.0168	.0477	-.2216	.2330
		2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.9698	.0300	.0554	.0084	.0630
	1 λ + 1 τ	1 λ and 1 τ	-	*100% Unbalanced*	40	.9871	.0172	.0409	-.2166	.1020
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.9710	.0287	.0591	-.1882	.1120

2. Generation Model: Mixed N-IV, $\Delta\kappa$ = .20, DIF = 10%, N = 300

	Total Invariance Constraints	Constraints on non-invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Power Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*100% Balanced*	54	.9528	.0371	.0678	.2097	.3400
	4 λ + 4 τ	4 λ and 4 τ	-	*100% Unbalanced*	46	.9859	.0168	.0477	.0330	.0580
		2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.9721	.0284	.0550	.2165	.2990
	1 λ + 1 τ	1 λ and 1 τ	-	*100% Unbalanced*	40	.9871	.0172	.0409	.0468	.0520
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.9710	.0287	.0591	.0412	.0630

3. Generation Model: Mixed N-IV, $\Delta\kappa$ = 0, DIF = 20%, N = 300

	Total Invariance Constraints	Constraints on non-invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Type I Error Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*100% Balanced*	54	.8185	.0841	.0985	-.0017	.0870
	4 λ + 4 τ	4 λ and 4 τ	-	*100% Unbalanced*	46	.9866	.0169	.0478	-.5068	.6650
		2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.9086	.0632	.0750	.0045	.0880
	1 λ + 1 τ	1 λ and 1 τ	-	*100% Unbalanced*	40	.9877	.0172	.0409	-.5157	.2820
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.9145	.0604	.0786	-.3924	.3420

4. Generation Model: Mixed N-IV, $\Delta\kappa$ = .20, DIF = 20%, *N* = 300

	Total Invariance Constraints	Constraints on non-invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Power Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*100% Balanced*	54	.8399	.0788	.0963	.2203	.3940
	4 λ + 4 τ	4 λ and 4 τ	-	*100% Unbalanced*	46	.9866	.0168	.0478	-.1861	.1560
		2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.9190	.0591	.0735	.2264	.3300
	1 λ + 1 τ	1 λ and 1 τ	-	*100% Unbalanced*	40	.9877	.0172	.0409	-.1804	.0760
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.9145	.0604	.0788	-.1368	.0900

5. Generation Model: Mixed N-IV, $\Delta\kappa$ = 0, DIF = 10%, *N* = 500

	Total Invariance Constraints	Constraints on non-invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Type I Error Rates
Analysis	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*100% Balanced*	54	.9520	.0402	.0588	-.0014	.0500

Models	4 λ + 4 τ	4 λ and 4 τ	-	*100% Unbalanced*	46	.9919	.0123	.0368	-.2235	.3850
		2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.9742	.0293	.0466	-.0031	.0490
	1 λ + 1 τ	1 λ and 1 τ	-	*100% Unbalanced*	40	.9925	.0125	.0316	-.2293	.1660
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.9762	.0274	.0493	-.2007	.1790

6. Generation Model: Mixed N-IV, Δ*κ* = .20, DIF = 10%, *N* = 500

	Total Invariance Constraints	Constraints on non-invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of Δ*κ*	Power Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*100% Balanced*	54	.9572	.0374	.0580	.2041	.4760
	4 λ + 4 τ	4 λ and 4 τ	-	*100% Unbalanced*	46	.9919	.0123	.0369	.0283	.0530
		2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.9766	.0274	.0461	.2032	.3970
	1 λ + 1 τ	1 λ and 1 τ	-	*100% Unbalanced*	40	.9925	.0125	.0316	.0272	.0440
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.9762	.0274	.0494	.0247	.0560

7. Generation Model: Mixed N-IV, Δ*κ* = 0, DIF = 20%, *N* = 500

	Total Invariance Constraints	Constraints on non-invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of Δ*κ*	Type I Error Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*100% Balanced*	54	.8197	.0841	.0924	-.0045	.0750
	4 λ + 4 τ	4 λ and 4 τ	-	*100% Unbalanced*	46	.9922	.0123	.0370	-.5053	.8930
		2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.9092	.0641	.0689	-.0071	.0690
	1 λ + 1 τ	1 λ and 1 τ	-	*100% Unbalanced*	40	.9928	.0126	.0316	-.5192	.5520
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.9158	.0611	.0719	-.4007	.5790

8. Generation Model: Mixed N-IV, Δ*κ* = .20, DIF = 20%, *N* = 500

	Total Invariance Constraints	Constraints on non-invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Power Rates
Analysis Models	$8\lambda + 8\tau$	4λ and 4τ	4λ and 4τ	*100% Balanced*	54	.8411	.0789	.0899	.2159	.5350
	$4\lambda + 4\tau$	4λ and 4τ	-	*100% Unbalanced*	46	.9922	.0123	.0369	-.1886	.2150
		2λ and 2τ	2λ and 2τ	*100% Balanced*		.9198	.0600	.0673	.2135	.4360
	$1\lambda + 1\tau$	1λ and 1τ	-	*100% Unbalanced*	40	.9928	.0126	.0316	-.1947	.1000
	$8\lambda + 1\tau$	4λ and 1τ	4λ and 0τ	-	47	.9158	.0611	.0721	-.1499	.1130

Table 11

Estimates of Factor Mean Differences, Bias, Relative Bias, Efficiency, and Effect Size of Estimated Factor Mean Differences When Generation Models are Mixed N-IV

1. Generation Model: Mixed N-IV, $\Delta\kappa$ = 0, DIF = 10%, N = 300

	Total Invariance Constraints	Constraints on non-invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*100% Balanced*	54	.0027	.0027	-	.1381	.0027
	4 λ + 4 τ	4 λ and 4 τ	-	*100% Unbalanced*	46	-.2216	-.2216	-	.1775	-.1359
		2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.0084	.0084	-	.1574	.0065
	1 λ + 1 τ	1 λ and 1 τ	-	*100% Unbalanced*	40	-.2166	-.2166	-	.2822	-.1290
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	-.1882	-.1882	-	.2398	-.1316

2. Generation Model: Mixed N-IV, $\Delta\kappa$ = .20, DIF = 10%, N = 300

	Total Invariance Constraints	Constraints on non-invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*100% Balanced*	54	.2097	.0097	4.8400	.1385	.1478
	4 λ + 4 τ	4 λ and 4 τ	-	*100% Unbalanced*	46	.0330	-.1670	-83.4945	.1757	.0207
		2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.2165	.0165	8.2420	.1580	.1519
	1 λ + 1 τ	1 λ and 1 τ	-	*100% Unbalanced*	40	.0468	-.1532	-76.6140	.2779	.0289
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.0412	-.1588	-79.3925	.2384	.0291

3. Generation Model: Mixed N-IV, $\Delta\kappa$ = 0, DIF = 20%, N = 300

	Total Invariance Constraints	Constraints on non-invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*100% Balanced*	54	-.0017	-.0017	-	.1420	.0020
	4 λ + 4 τ	4 λ and 4 τ	-	*100% Unbalanced*	46	-.5068	-.5068	-	.2119	-.2669
		2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.0045	.0045	-	.1635	.0068
	1 λ + 1 τ	1 λ and 1 τ	-	*100% Unbalanced*	40	-.5157	-.5157	-	.3482	-.2614
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	-.3924	-.3924	-	.2462	-.2743

4. Generation Model: Mixed N-IV, $\Delta\kappa$ = .20, DIF = 20%, N = 300										
	Total Invariance Constraints	Constraints on non-invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*100% Balanced*	54	.2203	.0203	10.1740	.1396	.1581
	4 λ + 4 τ	4 λ and 4 τ	-	*100% Unbalanced*	46	-.1861	-.3861	-193.0375	.2038	-.0977
		2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.2264	.0264	13.2150	.1605	.1621
	1 λ + 1 τ	1 λ and 1 τ	-	*100% Unbalanced*	40	-.1804	-.3804	-190.1855	.3256	-.0907
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	-.1368	-.3368	-168.3775	.2405	-.0955

5. Generation Model: Mixed N-IV, $\Delta\kappa$ = 0, DIF = 10%, N = 500										
	Total Invariance Constraints	Constraints on non-invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*100% Balanced*	54	-.0014	-.0014	-	.1061	-.0008

Models	4 λ + 4 τ	4 λ and 4 τ	-	*100% Unbalanced*	46	-.2235	-.2235	-	.1358	-.1387
		2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		-.0031	-.0031	-	.1206	-.0021
	1 λ + 1 τ	1 λ and 1 τ	-	*100% Unbalanced*	40	-.2293	-.2293	-	.2110	-.1393
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	-.2007	-.2007	-	.1824	-.1415

6. Generation Model: Mixed N-IV, $\Delta\kappa$ = .20, DIF = 10%, N = 500

	Total Invariance Constraints	Constraints on non-invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*100% Balanced*	54	.2041	.0041	2.0445	.1064	.1442
	4 λ + 4 τ	4 λ and 4 τ	-	*100% Unbalanced*	46	.0283	-.1717	-85.8500	.1344	.0177
		2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.2032	.0032	1.6200	.1210	.1430
	1 λ + 1 τ	1 λ and 1 τ	-	*100% Unbalanced*	40	.0272	-.1728	-86.4020	.2077	.0172
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.0247	-.1753	-87.6370	.1811	.0172

7. Generation Model: Mixed N-IV, $\Delta\kappa$ = 0, DIF = 20%, N = 500

	Total Invariance Constraints	Constraints on non-invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*100% Balanced*	54	-.0045	-.0045	-	.1090	-.0018
	4 λ + 4 τ	4 λ and 4 τ	-	*100% Unbalanced*	46	-.5053	-.5053	-	.1615	-.2694
		2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		-.0071	-.0071	-	.1253	-.0034
	1 λ + 1 τ	1 λ and 1 τ	-	*100% Unbalanced*	40	-.5192	-.5192	-	.2562	-.2700
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	-.4007	-.4007	-	.1870	-.2820

8. Generation Model: Mixed N-IV, $\Delta\kappa$ = .20, DIF = 20%, N = 500

	Total Invariance Constraints	Constraints on non-invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*100% Balanced*	54	.2159	.0159	7.9530	.1071	.1545
	4 λ + 4 τ	4 λ and 4 τ	-	*100% Unbalanced*	46	-.1886	-.3886	-194.2925	.1555	-.1005
		2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.2135	.0135	6.7455	.1228	.1523
	1 λ + 1 τ	1 λ and 1 τ	-	*100% Unbalanced*	40	-.1947	-.3947	-197.3555	.2408	-.1009
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	-.1499	-.3499	-174.9475	.1826	-.1057

Table 12

Model Fit Indices, Type I and Type II Error Rates Associated with Factor Mean Difference Testing, and Estimates of Factor Mean Differences When Generation Models are Uniform P-IV

1. Generation Model: Uniform P-IV, $\Delta\kappa$ = 0, DIF = 10%, N = 300

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Type I Error Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*50% Unbalanced*	54	.9762	.0221	.0588	.1009	.1260
	4 λ + 4 τ	4 λ and 4 τ	-	*Appropriate*	46	.9857	.0168	.0476	.0054	.0550
		-	4 λ and 4 τ	*100% Unbalanced*		.9854	.0169	.0477	.1847	.2590
		2 λ and 2 τ	2 λ and 2 τ	*50% Unbalanced*		.9824	.0196	.0496	.1063	.1120
	1 λ + 1 τ	1 λ and 1 τ	-	*Appropriate*	40	.9870	.0171	.0409	.0168	.0540
		-	1 λ and 1 τ	*100% Unbalanced*		.9870	.0171	.0409	.1876	.1250
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.9818	.0197	.0534	.0159	.0660
		4 λ and 0 τ	4 λ and 1 τ	-		.9818	.0197	.0534	.1960	.1340

2. Generation Model: Uniform P-IV, $\Delta\kappa$ = .20, DIF = 10%, N = 300

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Power Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*50% Unbalanced*	54	.9772	.0214	.0588	.2832	.5640
	4 λ + 4 τ	4 λ and 4 τ	-	*Appropriate*	46	.9857	.0168	.0477	.2093	.2680
		-	4 λ and 4 τ	*100% Unbalanced*		.9854	.0169	.0478	.3486	.6950
		2 λ and 2 τ	2 λ and 2 τ	*50% Unbalanced*		.9829	.0193	.0496	.2895	.5090
	1 λ + 1 τ	1 λ and 1 τ	-	*Appropriate*	40	.9870	.0171	.0409	.2260	.1280
		-	1 λ and 1 τ	*100% Unbalanced*		.9870	.0171	.0409	.3547	.3410
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.9818	.0197	.0534	.2123	.1440

		4 λ and 0 τ	4 λ and 1 τ	-		.9818	.0197	.0534	.3701	.3550

3. Generation Model: Uniform P-IV, $\Delta\kappa$ = 0, DIF = 20%, N = 300										
	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Type I Error Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*50% Unbalanced*	54	.9508	.0387	.0678	.1932	.3550
	4 λ + 4 τ	4 λ and 4 τ	-	*Appropriate*	46	.9861	.0168	.0476	.0054	.0550
		-	4 λ and 4 τ	*100% Unbalanced*		.9859	.0168	.0477	.3286	.7310
		2 λ and 2 τ	2 λ and 2 τ	*50% Unbalanced*		.9713	.0294	.0550	.1986	.3150
	1 λ + 1 τ	1 λ and 1 τ	-	*Appropriate*	40	.9874	.0171	.0409	.0168	.0540
		-	1 λ and 1 τ	*100% Unbalanced*		.9874	.0171	.0409	.3327	.3840
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.9717	.0285	.0588	.0154	.0660
		4 λ and 0 τ	4 λ and 1 τ	-		.9717	.0285	.0588	.3586	.3890

4. Generation Model: Uniform P-IV, $\Delta\kappa$ = .20, DIF = 20%, N = 300										
	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Power Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*50% Unbalanced*	54	.9564	.0356	.0670	.3552	.8130
	4 λ + 4 τ	4 λ and 4 τ	-	*Appropriate*	46	.9861	.0168	.0476	.2092	.2690
		-	4 λ and 4 τ	*100% Unbalanced*		.9859	.0168	.0479	.4589	.9510
		2 λ and 2 τ	2 λ and 2 τ	*50% Unbalanced*		.9738	.0275	.0545	.3613	.7150
	1 λ + 1 τ	1 λ and 1 τ	-	*Appropriate*	40	.9874	.0171	.0409	.2258	.1280
		-	1 λ and 1 τ	*100% Unbalanced*		.9874	.0171	.0409	.4649	.6500
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.9717	.0285	.0589	.2045	.1450
		4 λ and 0 τ	4 λ and 1 τ	-		.9717	.0285	.0589	.5010	.6560

5. Generation Model: Uniform P-IV, $\Delta\kappa$ = 0, DIF = 10%, N = 500

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Type I Error Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*50% Unbalanced*	54	.9831	.0195	.0471	.0959	.1590
	4 λ + 4 τ	4 λ and 4 τ	-	*Appropriate*	46	.9918	.0123	.0368	.0013	.0440
		-	4 λ and 4 τ	*100% Unbalanced*		.9917	.0123	.0368	.1786	.3830
		2 λ and 2 τ	2 λ and 2 τ	*50% Unbalanced*		.9881	.0164	.0395	.0949	.1260
	1 λ + 1 τ	1 λ and 1 τ	-	*Appropriate*	40	.9924	.0125	.0317	-.0003	.0520
		-	1 λ and 1 τ	*100% Unbalanced*		.9924	.0125	.0317	.1887	.1930
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.9883	.0160	.0423	.0005	.0570
		4 λ and 0 τ	4 λ and 1 τ	-		.9883	.0160	.0422	.1959	.2000

6. Generation Model: Uniform P-IV, $\Delta\kappa$ = .20, DIF = 10%, *N* = 500

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Power Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*50% Unbalanced*	54	.9842	.0185	.0470	.2770	.7910
	4 λ + 4 τ	4 λ and 4 τ	-	*Appropriate*	46	.9918	.0122	.0369	.2033	.3880
		-	4 λ and 4 τ	*100% Unbalanced*		.9917	.0123	.0369	.3412	.8840
		2 λ and 2 τ	2 λ and 2 τ	*50% Unbalanced*		.9886	.0159	.0394	.2765	.6850
	1 λ + 1 τ	1 λ and 1 τ	-	*Appropriate*	40	.9924	.0125	.0317	.2045	.1770
		-	1 λ and 1 τ	*100% Unbalanced*		.9924	.0125	.0317	.3545	.5680
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.9883	.0160	.0423	.1935	.1910
		4 λ and 0 τ	4 λ and 1 τ	-		.9883	.0160	.0423	.3678	.5750

7. Generation Model: Uniform P-IV, $\Delta\kappa$ = 0, DIF = 20%, *N* = 500

	Total Invariance Constraints	Constraints on invariant parameters	Constraints on non-invariant	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Type I Error Rates

		(V1 - V4)	parameters (V5 - V8)							
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*50% Unbalanced*	54	.9551	.0390	.0580	.1879	.4960
	4 λ + 4 τ	4 λ and 4 τ	-	*Appropriate*	46	.9921	.0122	.0368	.0013	.0420
		-	4 λ and 4 τ	*100% Unbalanced*		.9919	.0123	.0369	.3218	.9060
		2 λ and 2 τ	2 λ and 2 τ	*50% Unbalanced*		.9756	.0286	.0462	.1870	.4160
	1 λ + 1 τ	1 λ and 1 τ	-	*Appropriate*	40	.9927	.0125	.0317	-.0002	.0540
		-	1 λ and 1 τ	*100% Unbalanced*		.9927	.0125	.0317	.3328	.6020
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.9770	.0272	.0490	.0003	.0570
		4 λ and 0 τ	4 λ and 1 τ	-		.9770	.0272	.0490	.3562	.6030

8. Generation Model: Uniform P-IV, $\Delta\kappa$ = .20, DIF = 20%, *N* = 500

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Power Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*50% Unbalanced*	54	.9611	.0357	.0570	.3488	.9570
	4 λ + 4 τ	4 λ and 4 τ	-	*Appropriate*	46	.9921	.0122	.0368	.2032	.3870
		-	4 λ and 4 τ	*100% Unbalanced*		.9919	.0124	.0369	.4512	.9940
		2 λ and 2 τ	2 λ and 2 τ	*50% Unbalanced*		.9784	.0263	.0456	.3484	.9030
	1 λ + 1 τ	1 λ and 1 τ	-	*Appropriate*	40	.9927	.0125	.0317	.2044	.1770
		-	1 λ and 1 τ	*100% Unbalanced*		.9927	.0125	.0317	.4642	.8730
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.9770	.0272	.0491	.1864	.1910
		4 λ and 0 τ	4 λ and 1 τ	-		.9770	.0272	.0491	.4968	.8730

Table 13

Estimates of Factor Mean Differences, Bias, Relative Bias, Efficiency, and Effect Size of Estimated Factor Mean Differences When Generation Models are Uniform P-IV

1. Generation Model: Uniform P-IV, $\Delta\kappa = 0$, DIF = 10%, $N = 300$

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*50% Unbalanced*	54	.1009	.1009	-	.1303	.0749
	4 λ + 4 τ	4 λ and 4 τ	-	*Appropriate*	46	.0054	.0054	-	.1559	.0041
		-	4 λ and 4 τ	*100% Unbalanced*		.1847	.1847	-	.1410	.1431
		2 λ and 2 τ	2 λ and 2 τ	*50% Unbalanced*		.1063	.1063	-	.1481	.0786
	1 λ + 1 τ	1 λ and 1 τ	-	*Appropriate*	40	.0168	.0168	-	.2447	.0121
		-	1 λ and 1 τ	*100% Unbalanced*		.1876	.1876	-	.2192	.1438
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.0159	.0159	-	.2275	.0120
		4 λ and 0 τ	4 λ and 1 τ	-		.1960	.1960	-	.2266	.1452

2. Generation Model: Uniform P-IV, $\Delta\kappa = .20$, DIF = 10%, $N = 300$

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*50% Unbalanced*	54	.2832	.0832	41.6050	.1319	.2097
	4 λ + 4 τ	4 λ and 4 τ	-	*Appropriate*	46	.2093	.0093	4.6340	.1574	.1463
		-	4 λ and 4 τ	*100% Unbalanced*		.3486	.1486	74.3060	.1438	.2698
		2 λ and 2 τ	2 λ and 2 τ	*50% Unbalanced*		.2895	.0895	44.7485	.1503	.2136

	1 λ + 1 τ	1 λ and 1 τ	-	*Appropriate*	40	.2260	.0260	13.0035	.2484	.1552
		-	1 λ and 1 τ	*100% Unbalanced*		.3547	.1547	77.3615	.2252	.2712
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.2123	.0123	6.1480	.2295	.1570
		4 λ and 0 τ	4 λ and 1 τ	-		.3701	.1701	85.0740	.2310	.2738

3. Generation Model: Uniform P-IV, $\Delta\kappa = 0$, DIF = 20%, $N = 300$

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*50% Unbalanced*	54	.1932	.1932	-	.1242	.1512
	4 λ + 4 τ	4 λ and 4 τ	-	*Appropriate*	46	.0054	.0054	-	.1559	.0041
		-	4 λ and 4 τ	*100% Unbalanced*		.3286	.3286	-	.1306	.2754
		2 λ and 2 τ	2 λ and 2 τ	*50% Unbalanced*		.1986	.1986	-	.1413	.1550
	1 λ + 1 τ	1 λ and 1 τ	-	*Appropriate*	40	.0168	.0168	-	.2442	.0121
		-	1 λ and 1 τ	*100% Unbalanced*		.3327	.3327	-	.2004	.2764
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.0154	.0154	-	.2191	.0122
		4 λ and 0 τ	4 λ and 1 τ	-		.3586	.3586	-	.2143	.2791

4. Generation Model: Uniform P-IV, $\Delta\kappa = .20$, DIF = 20%, $N = 300$

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*50% Unbalanced*	54	.3552	.1552	77.6050	.1260	.2776
	4 λ + 4 τ	4 λ and 4 τ	-	*Appropriate*	46	.2092	.0092	4.6040	.1573	.1463
		-	4 λ and 4 τ	*100% Unbalanced*		.4589	.2589	129.4410	.1336	.3844
		2 λ and 2 τ	2 λ and 2 τ	*50% Unbalanced*		.3613	.1613	80.6255	.1435	.2818

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of Δ*κ*	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
	1 λ + 1 τ	1 λ and 1 τ	-	*Appropriate*	40	.2258	.0258	12.8810	.2477	.1553
		-	1 λ and 1 τ	*100% Unbalanced*		.4649	.2649	132.4255	.2064	.3860
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.2045	.0045	2.2630	.2210	.1593
		4 λ and 0 τ	4 λ and 1 τ	-		.5010	.3010	150.5135	.2198	.3898

5. Generation Model: Uniform P-IV, Δ*κ* = 0, DIF = 10%, *N* = 500

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of Δ*κ*	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*50% Unbalanced*	54	.0959	.0959	-	.1002	.0713
	4 λ + 4 τ	4 λ and 4 τ	-	*Appropriate*	46	.0013	.0013	-	.1195	.0010
		-	4 λ and 4 τ	*100% Unbalanced*		.1786	.1786	-	.1083	.1388
		2 λ and 2 τ	2 λ and 2 τ	*50% Unbalanced*		.0949	.0949	-	.1136	.0702
	1 λ + 1 τ	1 λ and 1 τ	-	*Appropriate*	40	-.0003	-.0003	-	.1845	.0004
		-	1 λ and 1 τ	*100% Unbalanced*		.1887	.1887	-	.1679	.1448
	8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.0005	.0005	-	.1731	.0002
		4 λ and 0 τ	4 λ and 1 τ	-		.1959	.1959	-	.1730	.1453

6. Generation Model: Uniform P-IV, Δ*κ* = .20, DIF = 10%, *N* = 500

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of Δ*κ*	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	4 λ and 4 τ	*50% Unbalanced*	54	.2770	.0770	38.5080	.1014	.2059
	4 λ + 4 τ	4 λ and 4 τ	-	*Appropriate*	46	.2033	.0033	1.6330	.1206	.1429
		-	4 λ and 4 τ	*100% Unbalanced*		.3412	.1412	70.6140	.1104	.2650
		2 λ and 2 τ	2 λ and 2 τ	*50% Unbalanced*		.2765	.0765	38.2480	.1151	.2049

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
	$1\ \lambda + 1\ \tau$	$1\ \lambda$ and $1\ \tau$	-	*Appropriate*	40	.2045	.0045	2.2310	.1868	.1423
		-	$1\ \lambda$ and $1\ \tau$	*100% Unbalanced*		.3545	.1545	77.2300	.1725	.2718
	$8\ \lambda + 1\ \tau$	$4\ \lambda$ and $1\ \tau$	$4\ \lambda$ and $0\ \tau$	-	47	.1935	-.0065	-3.2530	.1743	.1435
		$4\ \lambda$ and $0\ \tau$	$4\ \lambda$ and $1\ \tau$	-		.3678	.1678	83.8950	.1763	.2729

7. Generation Model: Uniform P-IV, $\Delta\kappa = 0$, DIF = 20%, $N = 500$

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	$8\ \lambda + 8\ \tau$	$4\ \lambda$ and $4\ \tau$	$4\ \lambda$ and $4\ \tau$	*50% Unbalanced*	54	.1879	.1879	-	.0956	.1473
	$4\ \lambda + 4\ \tau$	$4\ \lambda$ and $4\ \tau$	-	*Appropriate*	46	.0013	.0013	-	.1195	.0010
		-	$4\ \lambda$ and $4\ \tau$	*100% Unbalanced*		.3218	.3218	-	.1004	.2704
		$2\ \lambda$ and $2\ \tau$	$2\ \lambda$ and $2\ \tau$	*50% Unbalanced*		.1870	.1870	-	.1084	.1462
	$1\ \lambda + 1\ \tau$	$1\ \lambda$ and $1\ \tau$	-	*Appropriate*	40	-.0002	-.0002	-	.1843	.0004
		-	$1\ \lambda$ and $1\ \tau$	*100% Unbalanced*		.3328	.3328	-	.1541	.2770
	$8\ \lambda + 1\ \tau$	$4\ \lambda$ and $1\ \tau$	$4\ \lambda$ and $0\ \tau$	-	47	.0003	.0003	-	.1668	.0001
		$4\ \lambda$ and $0\ \tau$	$4\ \lambda$ and $1\ \tau$	-		.3562	.3562	-	.1637	.2781

8. Generation Model: Uniform P-IV, $\Delta\kappa = .20$, DIF = 20%, $N = 500$

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	$8\ \lambda + 8\ \tau$	$4\ \lambda$ and $4\ \tau$	$4\ \lambda$ and $4\ \tau$	*50% Unbalanced*	54	.3488	.1488	74.4230	.0969	.2735
	$4\ \lambda + 4\ \tau$	$4\ \lambda$ and $4\ \tau$	-	*Appropriate*	46	.2032	.0032	1.6095	.1205	.1429
		-	$4\ \lambda$ and $4\ \tau$	*100% Unbalanced*		.4512	.2512	125.5760	.1027	.3791
		$2\ \lambda$ and $2\ \tau$	$2\ \lambda$ and $2\ \tau$	*50% Unbalanced*		.3484	.1484	74.1915	.1101	.2726

1 λ + 1 τ	1 λ and 1 τ	-	*Appropriate*	40	.2044	.0044	2.2120	.1865	.1423
	-	1 λ and 1 τ	*100% Unbalanced*		.4642	.2642	132.1095	.1587	.3864
8 λ + 1 τ	4 λ and 1 τ	4 λ and 0 τ	-	47	.1864	-.0136	-6.7990	.1680	.1455
	4 λ and 0 τ	4 λ and 1 τ	-		.4968	.2968	148.4075	.1678	.3879

Table 14

Model Fit Indices, Type I and Type II Error Rates Associated with Factor Mean Difference Testing, and Estimates of Factor Mean Differences When Generation Models are Mixed P-IV

1. Generation Model: Mixed P-IV, $\Delta\kappa$ = 0, DIF = 10%, N = 300

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Type I Error Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	2 λ and 2 τ	2 λ and 2 τ	*50% Balanced*	54	.9680	.0279	.0621	.0035	.0560
	4 λ + 4 τ	4 λ and 4 τ	-	-	*Appropriate*	46	.9856	.0168	.0477	.0053	.0540
		-	2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.9696	.0297	.0553	.0009	.0580
		2 λ and 2 τ	2 λ and 2 τ	-	*50% Unbalanced*		.9824	.0199	.0496	-.0999	.0960
		2 λ and 2 τ	1 λ and 1 τ	1 λ and 1 τ	*50% Balanced*		.9788	.0228	.0514	.0046	.0560
	1 λ + 1 τ	1 λ and 1 τ	-	-	*Appropriate*	40	.9869	.0171	.0410	.0169	.0540
		-	1 λ and 1 τ	-	*100% Unbalanced*		.9869	.0171	.0410	-.2337	.0990
	8 λ + 1 τ	4 λ and 1 τ	2 λ and 0 τ	2 λ and 0 τ	-	47	.9785	.0226	.0553	.0167	.0650
		4 λ and 0 τ	2 λ and 1 τ	2 λ and 0 τ	-		.9785	.0226	.0553	-.2034	.1180

2. Generation Model: Mixed P-IV, $\Delta\kappa$ = .20, DIF = 10%, N = 300

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Power Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	2 λ and 2 τ	2 λ and 2 τ	*50% Balanced*	54	.9701	.0265	.0618	.2083	.3370
	4 λ + 4 τ	4 λ and 4 τ	-	-	*Appropriate*	46	.9856	.0168	.0477	.2093	.2670
		-	2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.9718	.0281	.0549	.2092	.2690
		2 λ and 2 τ	2 λ and 2 τ	-	*50% Unbalanced*		.9828	.0195	.0495	.1297	.1290
		2 λ and 2 τ	1 λ and 1 τ	1 λ and 1 τ	*50% Balanced*		.9798	.0220	.0511	.2116	.2770
	1 λ + 1 τ	1 λ and 1 τ	-	-	*Appropriate*	40	.9869	.0171	.0410	.2263	.1280

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of Δ*κ*	Type I Error Rates
		-	1 λ and 1 τ	-	*100% Unbalanced*		.9869	.0171	.0410	.0297	.0410
	8 λ + 1 τ	4 λ and 1 τ	2 λ and 0 τ	2 λ and 0 τ	-	47	.9785	.0226	.0554	.2229	.1440
		4 λ and 0 τ	2 λ and 1 τ	2 λ and 0 τ	-		.9785	.0226	.0554	.0268	.0500

3. Generation Model: Mixed P-IV, Δ*κ* = 0, DIF = 20%, *N* = 300

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of Δ*κ*	Type I Error Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	2 λ and 2 τ	2 λ and 2 τ	*50% Balanced*	54	.9052	.0581	.0796	.0021	.0650
	4 λ + 4 τ	4 λ and 4 τ	-	-	*Appropriate*	46	.9859	.0168	.0476	.0052	.0550
		-	2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.9066	.0627	.0743	-.0041	.0770
		2 λ and 2 τ	2 λ and 2 τ	-	*50% Unbalanced*		.9704	.0299	.0554	-.2303	.2470
		2 λ and 2 τ	1 λ and 1 τ	1 λ and 1 τ	*50% Balanced*		.9512	.0422	.0622	.0029	.0680
	1 λ + 1 τ	1 λ and 1 τ	-	-	*Appropriate*	40	.9871	.0171	.0410	.0170	.0540
		-	1 λ and 1 τ	-	*100% Unbalanced*		.9871	.0171	.0410	-.5347	.3010
	8 λ + 1 τ	4 λ and 1 τ	2 λ and 0 τ	2 λ and 0 τ	-	47	.9532	.0407	.0661	.0166	.0650
		4 λ and 0 τ	2 λ and 1 τ	2 λ and 0 τ	-		.9532	.0407	.0661	-.4080	.3630

4. Generation Model: Mixed P-IV, Δ*κ* = .20, DIF = 20%, *N* = 300

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of Δ*κ*	Power Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	2 λ and 2 τ	2 λ and 2 τ	*50% Balanced*	54	.9160	.0542	.0783	.2127	.3530
	4 λ + 4 τ	4 λ and 4 τ	-	-	*Appropriate*	46	.9859	.0168	.0477	.2092	.2660
		-	2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.9174	.0585	.0728	.2190	.3170
		2 λ and 2 τ	2 λ and 2 τ	-	*50% Unbalanced*		.9722	.0286	.0549	.0338	.0610
		2 λ and 2 τ	1 λ and 1 τ	1 λ and 1 τ	*50% Balanced*		.9562	.0394	.0613	.2159	.2930
	1 λ + 1 τ	1 λ and 1 τ	-	-	*Appropriate*	40	.9871	.0171	.0410	.2263	.1280
		-	1 λ and 1 τ	-	*100% Unbalanced*		.9871	.0171	.0410	-.2000	.0560

	8 λ + 1 τ	4 λ and 1 τ	2 λ and 0 τ	2 λ and 0 τ	-	47	.9532	.0407	.0663	.2228	.1440
		4 λ and 0 τ	2 λ and 1 τ	2 λ and 0 τ	-		.9532	.0407	.0663	-.1517	.0890

5. Generation Model: Mixed P-IV, $\Delta\kappa = 0$, DIF = 10%, $N = 500$

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Type I Error Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	2 λ and 2 τ	2 λ and 2 τ	*50% Balanced*	54	.9732	.0274	.0514	-.0008	.0500
	4 λ + 4 τ	4 λ and 4 τ	-	-	*Appropriate*	46	.9918	.0123	.0369	.0013	.0440
		-	2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.9739	.0294	.0465	-.0032	.0450
		2 λ and 2 τ	2 λ and 2 τ	-	*50% Unbalanced*		.9882	.0164	.0394	-.1100	.1360
		2 λ and 2 τ	1 λ and 1 τ	1 λ and 1 τ	*50% Balanced*		.9834	.0211	.0422	.0002	.0390
	1 λ + 1 τ	1 λ and 1 τ	-	-	*Appropriate*	40	.9924	.0125	.0317	-.0004	.0510
		-	1 λ and 1 τ	-	*100% Unbalanced*		.9924	.0125	.0317	-.2267	.1710
	8 λ + 1 τ	4 λ and 1 τ	2 λ and 0 τ	2 λ and 0 τ	-	47	.9846	.0199	.0447	.0006	.0570
		4 λ and 0 τ	2 λ and 1 τ	2 λ and 0 τ	-		.9846	.0199	.0447	-.1980	.1860

6. Generation Model: Mixed P-IV, $\Delta\kappa = .20$, DIF = 10%, $N = 500$

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Power Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	2 λ and 2 τ	2 λ and 2 τ	*50% Balanced*	54	.9756	.0256	.0510	.2026	.4620
	4 λ + 4 τ	4 λ and 4 τ	-	-	*Appropriate*	46	.9918	.0123	.0369	.2034	.3860
		-	2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.9764	.0274	.0460	.2031	.3960
		2 λ and 2 τ	2 λ and 2 τ	-	*50% Unbalanced*		.9887	.0159	.0393	.1173	.1440
		2 λ and 2 τ	1 λ and 1 τ	1 λ and 1 τ	*50% Balanced*		.9845	.0202	.0419	.2049	.4040
	1 λ + 1 τ	1 λ and 1 τ	-	-	*Appropriate*	40	.9924	.0125	.0317	.2045	.1770
		-	1 λ and 1 τ	-	*100% Unbalanced*		.9924	.0125	.0317	.0332	.0430
	8 λ + 1 τ	4 λ and 1 τ	2 λ and 0 τ	2 λ and 0 τ	-	47	.9846	.0199	.0448	.2032	.1900

		4 λ and 0 τ	2 λ and 1 τ	2 λ and 0 τ	-		.9846	.0199	.0448	.0294	.0500

7. Generation Model: Mixed P-IV, $\Delta\kappa$ = 0, DIF = 20%, *N* = 500

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Type I Error Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	2 λ and 2 τ	2 λ and 2 τ	*50% Balanced*	54	.9060	.0591	.0719	-.0016	.0610
	4 λ + 4 τ	4 λ and 4 τ	-	-	*Appropriate*	46	.9919	.0124	.0368	.0013	.0440
		-	2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.9066	.0638	.0684	-.0060	.0650
		2 λ and 2 τ	2 λ and 2 τ	-	*50% Unbalanced*		.9752	.0289	.0466	-.2391	.4020
		2 λ and 2 τ	1 λ and 1 τ	1 λ and 1 τ	*50% Balanced*		.9523	.0439	.0553	-.0009	.0500
	1 λ + 1 τ	1 λ and 1 τ	-	-	*Appropriate*	40	.9926	.0125	.0317	-.0005	.0520
		-	1 λ and 1 τ	-	*100% Unbalanced*		.9926	.0125	.0317	-.5206	.5140
	8 λ + 1 τ	4 λ and 1 τ	2 λ and 0 τ	2 λ and 0 τ	-	47	.9561	.0415	.0579	.0006	.0570
		4 λ and 0 τ	2 λ and 1 τ	2 λ and 0 τ	-		.9561	.0415	.0579	-.4003	.5540

8. Generation Model: Mixed P-IV, $\Delta\kappa$ = .20, DIF = 20%, *N* = 500

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average CFI	Average RMSEA	Average SRMR	Average Estimates of $\Delta\kappa$	Power Rates
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	2 λ and 2 τ	2 λ and 2 τ	*50% Balanced*	54	.9171	.0553	.0703	.2076	.4870
	4 λ + 4 τ	4 λ and 4 τ	-	-	*Appropriate*	46	.9919	.0124	.0369	.2034	.3880
		-	2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.9178	.0597	.0667	.2146	.4340
		2 λ and 2 τ	2 λ and 2 τ	-	*50% Unbalanced*		.9771	.0273	.0460	.0222	.0530
		2 λ and 2 τ	1 λ and 1 τ	1 λ and 1 τ	*50% Balanced*		.9577	.0408	.0542	.2096	.4240
	1 λ + 1 τ	1 λ and 1 τ	-	-	*Appropriate*	40	.9926	.0125	.0317	.2044	.1770
		-	1 λ and 1 τ	-	*100% Unbalanced*		.9926	.0125	.0317	-.1915	.0990
	8 λ + 1 τ	4 λ and 1 τ	2 λ and 0 τ	2 λ and 0 τ	-	47	.9561	.0415	.0580	.2031	.1900
		4 λ and 0 τ	2 λ and 1 τ	2 λ and 0 τ	-		.9561	.0415	.0580	-.1470	.1160

Table 15

Estimates of Factor Mean Differences, Bias, Relative Bias, Efficiency, and Effect Size of Estimated Factor Mean Differences When Generation Models are Mixed P-IV

1. Generation Model: Mixed P-IV, $\Delta\kappa$ = 0, DIF = 10%, N = 300											
	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	$8\lambda + 8\tau$	4λ and 4τ	2λ and 2τ	2λ and 2τ	*50% Balanced*	54	.0035	.0035	-	.1376	.0030
	$4\lambda + 4\tau$	4λ and 4τ	-	-	*Appropriate*	46	.0053	.0053	-	.1560	.0040
		-	2λ and 2τ	2λ and 2τ	*100% Balanced*		.0009	.0009	-	.1572	.0017
		2λ and 2τ	2λ and 2τ	-	*50% Unbalanced*		-.0999	-.0999	-	.1663	-.0654
		2λ and 2τ	1λ and 1τ	1λ and 1τ	*50% Balanced*		.0046	.0046	-	.1573	.0039
	$1\lambda + 1\tau$	1λ and 1τ	-	-	*Appropriate*	40	.0169	.0169	-	.2452	.0121
		-	1λ and 1τ	-	*100% Unbalanced*		-.2337	-.2337	-	.2824	-.1397
	$8\lambda + 1\tau$	4λ and 1τ	2λ and 0τ	2λ and 0τ	-	47	.0167	.0167	-	.2392	.0120
		4λ and 0τ	2λ and 1τ	2λ and 0τ	-		-.2034	-.2034	-	.2406	-.1420

2. Generation Model: Mixed P-IV, $\Delta\kappa$ = .20, DIF = 10%, N = 300											
	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	$8\lambda + 8\tau$	4λ and 4τ	2λ and 2τ	2λ and 2τ	*50% Balanced*	54	.2083	.0083	4.1490	.1384	.1466
	$4\lambda + 4\tau$	4λ and 4τ	-	-	*Appropriate*	46	.2093	.0093	4.6385	.1575	.1462
		-	2λ and 2τ	2λ and 2τ	*100% Balanced*		.2092	.0092	4.5880	.1578	.1474
		2λ and 2τ	2λ and 2τ	-	*50% Unbalanced*		.1297	-.0703	-35.1315	.1662	.0856
		2λ and 2τ	1λ and 1τ	1λ and 1τ	*50% Balanced*		.2116	.0116	5.7890	.1584	.1480
	$1\lambda + 1\tau$	1λ and 1τ	-	-	*Appropriate*	40	.2263	.0263	13.1650	.2491	.1551

		-	1 λ and 1 τ	-	*100% Unbalanced*		.0297	-.1703	-85.1585	.2775	.0188
	8 λ + 1 τ	4 λ and 1 τ	2 λ and 0 τ	2 λ and 0 τ	-	47	.2229	.0229	11.4600	.2413	.1563
		4 λ and 0 τ	2 λ and 1 τ	2 λ and 0 τ	-		.0268	-.1732	-86.6005	.2389	.0194

3. Generation Model: Mixed P-IV, $\Delta\kappa = 0$, DIF = 20%, $N = 300$

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	2 λ and 2 τ	2 λ and 2 τ	*50% Balanced*	54	.0021	.0021	-	.1391	.0029
	4 λ + 4 τ	4 λ and 4 τ	-	-	*Appropriate*	46	.0052	.0052	-	.1560	.0040
		-	2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		-.0041	-.0041	-	.1638	.0011
		2 λ and 2 τ	2 λ and 2 τ	-	*50% Unbalanced*		-.2303	-.2303	-	.1815	-.1397
		2 λ and 2 τ	1 λ and 1 τ	1 λ and 1 τ	*50% Balanced*		.0029	.0029	-	.1599	.0038
	1 λ + 1 τ	1 λ and 1 τ	-	-	*Appropriate*	40	.0170	.0170	-	.2450	.0122
		-	1 λ and 1 τ	-	*100% Unbalanced*		-.5347	-.5347	-	.3486	-.2722
	8 λ + 1 τ	4 λ and 1 τ	2 λ and 0 τ	2 λ and 0 τ	-	47	.0166	.0166	-	.2391	.0119
		4 λ and 0 τ	2 λ and 1 τ	2 λ and 0 τ	-		-.4080	-.4080	-	.2468	-.2855

4. Generation Model: Mixed P-IV, $\Delta\kappa = .20$, DIF = 20%, $N = 300$

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	2 λ and 2 τ	2 λ and 2 τ	*50% Balanced*	54	.2127	.0127	6.3425	.1069	.1507
	4 λ + 4 τ	4 λ and 4 τ	-	-	*Appropriate*	46	.2092	.0092	4.5910	.1196	.1461
		-	2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.2190	.0190	9.5225	.1252	.1575
		2 λ and 2 τ	2 λ and 2 τ	-	*50% Unbalanced*		.0338	-.1662	-83.1085	.1389	.0215
		2 λ and 2 τ	1 λ and 1 τ	1 λ and 1 τ	*50% Balanced*		.2159	.0159	7.9275	.1222	.1522
	1 λ + 1 τ	1 λ and 1 τ	-	-	*Appropriate*	40	.2263	.0263	13.1535	.1846	.1551
		-	1 λ and 1 τ	-	*100% Unbalanced*		-.2000	-.4000	-199.9785	.2595	-.1011
	8 λ + 1 τ	4 λ and 1 τ	2 λ and 0 τ	2 λ and 0 τ	-	47	.2228	.0228	11.4005	.1816	.1561

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
		4 λ and 0 τ	2 λ and 1 τ	2 λ and 0 τ	-		-.1517	-.3517	-175.8735	.1885	-.1058

5. Generation Model: Mixed P-IV, $\Delta\kappa$ = 0, DIF = 10%, N = 500

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	2 λ and 2 τ	2 λ and 2 τ	*50% Balanced*	54	-.0008	-.0008	-	.1058	-.0004
	4 λ + 4 τ	4 λ and 4 τ	-	-	*Appropriate*	46	.0013	.0013	-	.1196	.0010
		-	2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		-.0032	-.0032	-	.1205	-.0019
		2 λ and 2 τ	2 λ and 2 τ	-	*50% Unbalanced*		-.1100	-.1100	-	.1274	-.0730
		2 λ and 2 τ	1 λ and 1 τ	1 λ and 1 τ	*50% Balanced*		.0002	.0002	-	.1203	.0002
	1 λ + 1 τ	1 λ and 1 τ	-	-	*Appropriate*	40	-.0004	-.0004	-	.1847	.0003
		-	1 λ and 1 τ	-	*100% Unbalanced*		-.2267	-.2267	-	.2135	-.1375
	8 λ + 1 τ	4 λ and 1 τ	2 λ and 0 τ	2 λ and 0 τ	-	47	.0006	.0006	-	.1817	.0002
		4 λ and 0 τ	2 λ and 1 τ	2 λ and 0 τ	-		-.1980	-.1980	-	.1838	-.1396

6. Generation Model: Mixed P-IV, $\Delta\kappa$ = .20, DIF = 10%, N = 500

	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	2 λ and 2 τ	2 λ and 2 τ	*50% Balanced*	54	.2026	.0026	1.3105	.1063	.1430
	4 λ + 4 τ	4 λ and 4 τ	-	-	*Appropriate*	46	.2034	.0034	1.6835	.1207	.1429
		-	2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.2031	.0031	1.5650	.1209	.1435
		2 λ and 2 τ	2 λ and 2 τ	-	*50% Unbalanced*		.1173	-.0827	-41.3445	.1272	.0776
		2 λ and 2 τ	1 λ and 1 τ	1 λ and 1 τ	*50% Balanced*		.2049	.0049	2.4395	.1211	.1439
	1 λ + 1 τ	1 λ and 1 τ	-	-	*Appropriate*	40	.2045	.0045	2.2370	.1871	.1422
		-	1 λ and 1 τ	-	*100% Unbalanced*		.0332	-.1668	-83.4105	.2104	.0204
	8 λ + 1 τ	4 λ and 1 τ	2 λ and 0 τ	2 λ and 0 τ	-	47	.2032	.0032	1.5940	.1831	.1429
		4 λ and 0 τ	2 λ and 1 τ	2 λ and 0 τ	-		.0294	-.1706	-85.3115	.1825	.0207

7. Generation Model: Mixed P-IV, $\Delta\kappa = 0$, DIF = 20%, $N = 500$											
	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	2 λ and 2 τ	2 λ and 2 τ	*50% Balanced*	54	-.0016	-.0016	-	.1069	-.0005
	4 λ + 4 τ	4 λ and 4 τ	-	-	*Appropriate*	46	.0013	.0013	-	.1196	.0010
		-	2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		-.0060	-.0060	-	.1252	-.0023
		2 λ and 2 τ	2 λ and 2 τ	-	*50% Unbalanced*		-.2391	-.2391	-	.1389	-.1469
		2 λ and 2 τ	1 λ and 1 τ	1 λ and 1 τ	*50% Balanced*		-.0009	-.0009	-	.1222	< .0001
	1 λ + 1 τ	1 λ and 1 τ	-	-	*Appropriate*	40	-.0005	-.0005	-	.1846	.0003
		-	1 λ and 1 τ	-	*100% Unbalanced*		-.5206	-.5206	-	.2595	-.2694
	8 λ + 1 τ	4 λ and 1 τ	2 λ and 0 τ	2 λ and 0 τ	-	47	.0006	.0006	-	.1816	.0002
		4 λ and 0 τ	2 λ and 1 τ	2 λ and 0 τ	-		-.4003	-.4003	-	.1885	-.2822

8. Generation Model: Mixed P-IV, $\Delta\kappa = .20$, DIF = 20%, $N = 500$											
	Total Invariance Constraints	Constraints on invariant parameters (V1 - V4)	Constraints on non-invariant parameters (V5 - V6)	Constraints on non-invariant parameters (V7 - V8)	Model Type	*df*	Average Estimates of $\Delta\kappa$	Average Bias	Average Relative Bias (%)	Efficiency	Effect Size
Analysis Models	8 λ + 8 τ	4 λ and 4 τ	2 λ and 2 τ	2 λ and 2 τ	*50% Balanced*	54	.2076	.0076	3.7780	.1066	.1473
	4 λ + 4 τ	4 λ and 4 τ	-	-	*Appropriate*	46	.2034	.0034	1.7060	.1206	.1429
		-	2 λ and 2 τ	2 λ and 2 τ	*100% Balanced*		.2146	.0146	7.3235	.1226	.1539
		2 λ and 2 τ	2 λ and 2 τ	-	*50% Unbalanced*		.0222	-.1778	-88.9155	.1364	.0138
		2 λ and 2 τ	1 λ and 1 τ	1 λ and 1 τ	*50% Balanced*		.2096	.0096	4.8090	.1219	.1481
	1 λ + 1 τ	1 λ and 1 τ	-	-	*Appropriate*	40	.2044	.0044	2.1785	.1870	.1422
		-	1 λ and 1 τ	-	*100% Unbalanced*		-.1915	-.3915	-195.7645	.2440	-.0989
	8 λ + 1 τ	4 λ and 1 τ	2 λ and 0 τ	2 λ and 0 τ	-	47	.2031	.0031	1.5320	.1829	.1427
		4 λ and 0 τ	2 λ and 1 τ	2 λ and 0 τ	-		-.1470	-.3470	-173.5090	.1841	-.1037