PARTIAL INVARIANCE IN LOADINGS AND INTERCEPTS — THEIR INTERPLAYS AND IMPLICATIONS FOR LATENT MEAN COMPARISONS

by

Yongwei Yang

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Psychological Studies in Education

Under the Supervision of Professor Deborah L. Bandalos

Lincoln, Nebraska

May, 2008

UMI Number: 3297716

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

PARTIAL INVARIANCE IN LOADINGS AND INTERCEPTS — THEIR INTERPLAYS AND IMPLICATIONS FOR LATENT MEAN COMPARISONS

Yongwei Yang, Ph.D.
University of Nebraska, 2008

Advisor: Deborah L. Bandalos

Mean and covariance structure modeling procedures have been widely used for assessing factorial invariance and the practice of partial factorial invariance (PFI) is common. This study investigates the extent to which PFI affects subsequent tests of intercept invariance and latent mean differences. The study implements a Monte Carlo experiment where factors of model size, severity of PFI, construct reliability, and intercept and latent mean difference pattern are manipulated. Results show that when partial loading invariance exists but is correctly specified, it did not negatively affect the power of the omnibus test of intercept invariance. However, results also suggest that relying on modification indices to determine noninvariant intercepts would be problematic — these indices are likely to lead to incorrect identifications of noninvariant parameters. With respect to latent mean comparisons, various forms of partial factorial invariance, when correctly specified, did not impact the test of latent mean differences. With the proliferation of educational research involving cross-group comparisons, researchers need to understand how PFI may affect analyses upon which substantive inferences are based. Findings from this study will provide a starting point for answering this very important question.

ACKNOWLEDGMENT

For something such as my doctoral studies that takes a long time to accomplish, there are naturally many people to thank. Since the day I arrived in Lincoln, Nebraska, I have never lacked mentors and friends. My advisor, Dr. Bandalos, has always been kind and supportive. More importantly, I have learned a great deal from her both on subjective matters and on how to become a competent and conscientious researcher. Other committee members, Drs. De Ayala, Bovaird, McCutcheon, and Garbin, have given me encouragement and advice that have guided me through this endeavor. I have also benefited from the teaching and advice from numerous professors I encountered at the University of Nebraska-Lincoln; they taught me not what to think but *how* to think. They cared and appreciated me as an individual.

It is also hard to imagine how I could get this far with my career without great friends around me. There are friends at Gallup, who genuinely cared about me when I struggled and genuinely cheered with me when I succeeded. There are friends who never hesitated to offer me help when I needed the most — such as Piotr, Joe, Jan, Sangeeta, Ted, Dennison, and Maribel. Then there are the people — such as the Surbers and the Tenneys — who make me feel I have a family here.

While writing this piece a few scenes popped into my mind. First, there was the scene in summer of 1996, Shanghai, when my grandfather waved goodbye to me. Then, there was the scene where he took me to the Shanghai Library and pointed to the pictures of great scholars on the wall and said he hoped one day I would have great achievements. Finally, there was the scene the day before I left China the first time, when in our small apartment my parents carefully packed whatever necessities that they could think of into my suitcases. I owe so much to those who love me so profoundly — my grandparents, my parents, my sister, and my dearest companion, Tzu-Yun. This dissertation is a small way to show my love to them — that I have achieved something that they can be proud of, be it in this world or in heaven.

To My Parents

TABLE OF CONTENTS

CHAPTER I. BACKGROUND 1
1. Introduction 1
2. The Concept and Importance of Measurement Invariance 4
3. MI in the Context of Mean and Covariance Structure Modeling 6
4. Evaluating Factorial Invariance with Mean and Covariance Structure Modeling 15
5. Summary 21

CHAPTER II. LITERATURE REVIEW AND RESEARCH QUESTIONS 23
1. A Closer Look at Partial Factorial Invariance 23
Arguments Supporting Partial Factorial Invariance 27
Types and Forms of PFI 29
Some Concerns about PFI 31
2. Impact of PFI — An Unresolved Issue 34
Impact of PFI in Factorial Invariance Analysis (FIA) Applications 37
Impact of PFI in Latent Mean Comparison (LMC) Applications 44
3. Purpose and Research Questions 56
4. Factors to Be Considered 59
Model Size 59
Severity of PFI in the Baseline Models 60
Severity of Noninvariance in the Target Models 61
Severity of PII 61
Effect Size of Latent Mean Difference 62
Construct Reliability and Observed Score Variance 63

Consistencies in the Direction of Intercept and Latent Mean Differences 65
Summary 66

CHAPTER III. METHODS AND PROCEDURES 68
1. Fixed Design Elements 68
2. Independent Variables — Factors and Levels to be Manipulated 70
Model Size 70
Severity of PFI in the Baseline Models 70
Loading 70
Intercept 71
Construct Reliability and Observed Score Variance 72
Severity of Noninvariance in the Target Models 78
Intercept 78
Latent Mean 78
Consistencies between the Direction of Intercept and Latent Mean Differences 79
Summary 79
3. Number of Replications 81
4. Data Generation and Model Fitting 83
5. Convergence Check 85
6. Dependent Variables 86
7. Data Analysis 89

CHAPTER IV. RESULTS 90
1. Evaluating Baseline Design Conditions 91
Baseline Design Conditions for RQ1 91
Baseline Design Conditions for RQ2 95
2. RQ1 Results 96
Type I Error 96
Power 98

Transformation of the Dependent Variable 99
Accuracy in Identifying Noninvariant Intercepts 109
3. RQ2a Results 114
Type I Error 114
Power 116
Latent Mean Estimation Bias 121
4. RQ2b Results 123
Type I Error 123
Power 124
Latent Mean Estimation Bias 129
5. RQ2c Results 130
Type I Error 130
Power 136
Latent Mean Estimation Bias 148

CHAPTER V. DISCUSSION 169
1. Impact of Partial Loading Invariance on Intercept Invariance Analysis 170
2. Impact of Partial Factorial Invariance on Latent Mean Comparisons 172
3. General Discussion 173
Limitations 173
Importance and Implications 174
Suggestions for Future Research 178

REFERENCES 180

LIST OF FIGURES

Figure 1. Illustration of configural invariance 9

Figure 2. Illustration of loading invariance 10

Figure 3. Illustration of intercept invariance 12

Figure 4. Illustration of uniqueness (co)variance) invariance 14

Figure 5. Illustration of research questions 57

Figure 6. RQ1 — Observed Type I error rates by design conditions 98

Figure 7. RQ1 — Observed power by design conditions 101

Figure 8. RQ2a — Observed Type I error rates by design conditions 116

Figure 9. RQ2a — Observed power by design conditions 118

Figure 10. RQ2b — Observed Type I error rates by design conditions 124

Figure 11. RQ2b — Observed power by design conditions 126

Figure 12. RQ2c — Observed Type I error rates by design conditions 132

Figure 13. RQ2c — Observed Power by design conditions 139

LIST OF TABLES

Table 1a. Loading size, error variance, *H*, and V_x under various PLI conditions for MS_1 (6 indicators) 76

Table 1b. Loading size, error variance, *H*, and V_x under various PLI conditions for MS_2 (12 indicators) 77

Table 2. Design factors and levels 80

Table 3. Power from baseline conditions for RQ1 92

Table 4. Power from baseline conditions for RQ1: Performance of modification indices for identifying targeted noninvariant intercepts 94

Table 5. Observed power from baseline conditions for RQ2a, 2b, and 2c 96

Table 6. Type I error rates by design conditions for RQ1 97

Table 7. Observed power by design conditions for RQ1 100

Table 8. Descriptive statistics of LR test p-values (when the population models were intercept noninvariant) by design conditions for RQ1 106

Table 9. RQ1 ANOVA with LR test p-values effect sizes 108

Table 10. Power for RQ1: Performance of modification indices for identifying targeted noninvariant intercepts by design conditions 110

Table 11. Type I Error rates by design conditions for RQ2a 115

Table 12. Observed power by design conditions for RQ2a ………………… 117

Table 13. Descriptive statistics of LR test p-values (when the population models had unequal latent means) by design conditions for RQ2a ……. 120

Table 14. RQ2a ANOVA with LR test p-values effect sizes ……………….. 121

Table 15. Latent mean estimation bias and efficiency for RQ2a …………… 122

Table 16. Type I Error by design conditions for RQ2b …………………….. 124

Table 17. Observed power by design conditions for RQ2b ………………… 125

Table 18. Descriptive statistics of LR test p-values (when the population models had unequal latent means) by design conditions for RQ2b ……. 128

Table 19. RQ2b ANOVA with LR test p-values effect sizes ……………….. 129

Table 20. Latent mean estimation bias and efficiency for RQ2b …………… 130

Table 21. Type I Error rates by design conditions for RQ2c ……………….. 131

Table 22. Observed power by design conditions for RQ2c ………………… 137

Table 23. Descriptive statistics of LR test p-values (when the population models had unequal latent means) by design conditions for RQ2c ……. 149

Table 24. RQ2c ANOVA with LR test p-values effect sizes ……………….. 157

Table 25. Latent mean estimation bias and efficiency for RQ2c …………… 161

ABBREVIATIONS

DIF	**D**ifferential **I**tem **F**unctioning
GOF	**G**oodness **O**f **F**it
FI	**F**actorial **I**nvariance
FIA	**F**actorial **I**nvariance **A**nalysis
LMC	**L**atent **M**ean **C**omparison
LR	**L**ikelihood **R**atio
MACS	**M**ean **A**nd **C**ovariance **S**tructure
MI	**M**easurement **I**nvariance
MIMIC	**M**ultiple **I**ndicators, **M**ultiple **C**auses
PCI	**P**artial **C**onfigural **I**nvariance
PFI	**P**artial **F**actorial **I**nvariance
PII	**P**artial **I**ntercept **I**nvariance
PLI	**P**artial **L**oading **I**nvariance
PMI	**P**artial **M**easurement **I**nvariance
SMM	**S**tructured **M**ean **M**odeling

ABBREVIATIONS USED TO DESCRIBE DESIGN FACTORS AND LEVELS

DIR	Consistency between the **Dir**ection of Intercept and Latent Mean Differences
G1	**G**roup **1**
G2	**G**roup **2**
HV	Construct Reliability (**H**) and Observed Total Score Variance (**V**)
IDP	**I**ntercept **D**ifference — **P**roportion
IDS	**I**ntercept **D**ifference — **S**ize
LDP	**L**oading **D**ifference — **P**roportion
LDS	**L**oading **D**ifference — **S**ize
LMD	**L**atent **M**ean **D**ifference
MS	**M**odel **S**ize

CHAPTER I. BACKGROUND

1. Introduction

Empirical research typically requires some quantification of phenomena or information, from as simple as recording the presence or absence of an event and counting to complex statistical modeling. Hence measurement, as broadly (yet succinctly) defined by Stevens[1] (1946) as the "assignment of numerals to objects or events according to rules" (p. 677), is a critical component of research. Six decades have elapsed since Stevens (1946) and the evolution of the "rules" has led to, for some applications, elegant "measurement models." Again broadly speaking, a measurement model can be viewed as a mathematical model that describes the relationships between observed phenomena (e.g., responses to tests or surveys) and the unobserved entities (true scores, factors, latent classes, etc.) that are believed to explain the former. Examples of these models include true score theory, common factor models, and various item response models. These models consist of "rules" that enable an investigator to, with some degrees of confidence, assign categories or values ("numerals") to objects or events based on the observed phenomena.

In research involving different populations, the same measurement model may be imposed on subjects in all relevant populations. When making inferences about the similarities and differences among populations with respect to the unobserved entities, an implicit assumption is that the measurement model works in the same way across

[1] Although this famous quote is usually credited to Stevens, he simply said it was a paraphrase of N. R. Campbell.

these populations. That is, researchers must first have confidence that the unobserved entities can be implied from the observed phenomena in the same way, before they can make inferences about these entities. This gives rise to the importance of the issue generally referred to as "measurement invariance" (MI) (Meredith, 1993). The investigation of MI concerns the generalizability of measurement relationships between latent variables and their manifest indicators across populations. Such populations could be naturally occurring (e.g., gender, country of residence), or derived based on some background information (e.g., socio-economic status, culture, academic achievement level), or formed for longitudinal comparisons (e.g., based on time points).

Various frameworks have been shown effective for investigating MI, including the applications of mean and covariance structure (MACS) modeling (Jöreskog, 1971; Sörbom, 1974; Little, 1997), MIMIC modeling (Muthén, 1989), item response theory (IRT) based analyses (Douglas, Roussos, & Stout, 1996; Lord, 1980; Raju, van der Linden, & Fleer, 1995; Thissen, Steinberg, & Wainer, 1988, 1993), loglinear and latent class modeling (Kelderman & Macready, 1990; McCutcheon, 2002; Mellenbergh, 1982; Saris, 2003), mixture modeling (De Ayala, Kim, Stapleton, & Dayton, 2002; Cohen & Bolt, 2005; Lubke & Muthén, 2005), multi-level modeling (Swanson, Clauser, Case, Nungester, & Featherman, 2002; Williams & Beretvas, 2006), etc. This dissertation research focused on a range of methodological issues and implications pertaining to MACS modeling, following the general framework developed by Jöreskog (1971). With this approach, MI is studied within the common factor model framework and entails the evaluation of cross-group invariance of the four measurement aspects implied by the common factor model, namely factor configuration, factor loadings, intercepts, and

uniquenesses. These aspects are evaluated in a sequence that starts with factor configuration, and followed by factor loadings, intercepts, and uniquenesses. Invariance in these aspects is generally regarded as the logical precursor to subsequent investigation of structural parameters of latent variables (factor (co)variances, factor means, structural regression weights, structural regression disturbances). In applied research, however, complete invariance may not be tenable with sample data on some or all of these aspects. Furthermore, some argue it may not be reasonable or necessary for subsequent analysis (e.g., Byrne, Shavelson, & Muthén, 1989; Labouvie & Ruetsch, 1995; Steenkamp & Baumgartner, 1998). Hence, the practice of allowing "partial measurement invariance" (PMI) is not uncommon. PMI refers to the situation where some but not all of the elements within measurement aspects are invariant across some or all populations.

Despite the practice of allowing for partial invariance in applied research, what is not clear is the impact of allowing for partial invariance in one or more measurement aspects on subsequent analyses. These subsequent analyses could be the evaluations of invariance in subsequent measurement aspects or the estimation and testing of the properties and relationships among latent variables. This dissertation study aimed at understanding the impact of partial invariance. Using a Monte Carlo experiment, the study investigated the impact of partial invariance in a measurement parameter on the analysis in subsequent measurement parameters. It investigated the impact of various forms of partial invariance on the estimation and testing of latent variable mean differences. The remainder of this chapter provides discussions of (1) the formal

concepts of MI, (2) the concepts of MI in the context of MACS, and (3) common procedures for evaluating MI with MACS modeling.

2. The Concept and Importance of Measurement Invariance

The formal definition of measurement invariance provided below is primarily based on Meredith (1993), Meredith and Millsap (1992), and Millsap and Everson (1993). Meredith (1993) is a particularly important reference in that it provided a series of theorems, corollaries, and proofs associated with various forms of measurement invariance. However, it should be noted that the relevant concepts and definition should also be, as Meredith and others noted, attributed to Mellenbergh (1989) (Lubke, Dolan, Kelderman, & Mellenbergh, 2003a, 2003b; Meredith, 1993; Millsap, 1995; Millsap & Meredith, 1992; Oort, 1992, 1998). Others, such as Lord (1980), Kelderman and Macready (1990), Mellenbergh and Kok (1991), Oort (1992), and Shealy and Stout (1993), have provided similar conceptualizations to Meredith (1993) or Mellenbergh (1989), but they are less general or detailed. Following these earlier works, the concepts of MI have been reiterated by various scholars, such as Millsap (1995, 2005), Borsboom, Mellenbergh, van Heerden (2002), Lubke, Dolan, Kelderman, and Mellenbergh (2003a, 2003b), DeShon (2004), Meredith and Teresi (2006), and Millsap and Meredith (2007).

Let $\boldsymbol{X}$ be a set of observed variables x (e.g., items, subscales, tests), $\boldsymbol{W}$ be a set of latent variables w that underlying $\boldsymbol{X}$ (i.e., "$\boldsymbol{X}$ measures $\boldsymbol{W}$"), and $\boldsymbol{V}$ be a set of population indicators v (e.g., gender, race, SES, developmental state), and $F(\cdot)$ denotes the density (for continuous $\boldsymbol{X}$) or probability distribution (for discrete $\boldsymbol{X}$) function in the population.

Following Meredith (1993), $\boldsymbol{X}$ possesses measurement invariance (or a lack of bias) with respect to $\boldsymbol{V}$ if, for any x, w, and v,

$$F(x \mid w, v) = F(x \mid w) \tag{1}$$

This definition states that given subjects' scores on the latent $\boldsymbol{W}$, the subjects' scores on the observed $\boldsymbol{X}$ do not depend on their group membership on $\boldsymbol{V}$ (DeShon, 2004). In other words, the distribution of the observed scores is not affected by group membership, after controlling for the latent variables.

From this general definition, Meredith (1993) further derived *weak measurement invariance* as the situation in which only the first two moments (i.e., mean and (co)variances) of the distributions of $\boldsymbol{X}$ are not affected by group membership, after controlling for the latent variables. It should be noted that if the distribution of $F(x \mid w, v)$ is multivariate normal, then weak measurement invariance will be the same as measurement invariance; because the multivariate normal distribution is determined by its first and second moments (Lubke et al., 2003a; Millsap, 2005).

The above conceptualization of measurement invariance is important because, when satisfied, it implies the between-group differences in the observed variables' means and (co)variances will only be due to the group differences in latent variable means and (co)variances. On the other hand if measurement invariance is not satisfied, as Cheung and Rensvold (2002) stated, "finding of a between-group difference cannot be unambiguously interpreted" (p. 233). It cannot be unambiguously interpreted because one would not be able to know for sure whether the observed between-group difference is due to real differences with respect to the latent variables or to the differences in how they are measured.

3. MI in the Context of Mean and Covariance Structure Modeling

The conceptualization of measurement invariance presented in the previous section is a general one. It does not imply any specific distribution form for $F(x \mid w, v)$ or any specific form of relationships between $\boldsymbol{X}$ and $\boldsymbol{W}$. In order to empirically evaluate and establish measurement invariance in different situations, this general definition can be specialized. One such specialization is through the common factor model. A common factor model first states that the latent variables that explain the observed indicators can be divided into two uncorrelated groups — common factors and unique factors (Gorsuch, 1983; Meredith, 1993). Common factors represent the latent variables that are explicitly modeled as underlying more than one observed indicators (e.g., items or tests), whereas unique factors represent both random and specific measurement errors that are unique to a single indicator. With continuous $\boldsymbol{X}$, the common factor model can be expressed as the linear regression of the observed indicators on the latent factors[1]. That is

$$X^g = \tau^g + \Lambda^g \xi^g + \delta^g \tag{2}$$

Here, $\boldsymbol{X}$ is now formally defined as a $q^g \times 1$ vector of observed indicators (e.g., q^g tests or items), $\boldsymbol{W}$ is replaced by $\boldsymbol{\xi}$ to distinguish from equation (1) and $\boldsymbol{\xi}$ is an $n^g \times 1$ vector of common factors, $\boldsymbol{\tau}$ is a $q^g \times 1$ vector of measurement intercepts, $\boldsymbol{\Lambda}$ is a $q^g \times n^g$ matrix of factor loadings, and $\boldsymbol{\delta}$ is a $q^g \times 1$ vector of unique factors (Jöreskog, 1971; Meredith, 1993). To facilitate subsequent discussions about between-group invariance, the

[1] Lubke et al. (2003a) used a stricter definition in that they define common factor model as one "where observed multivariate normally distributed variables are linearly regressed on underlying multivariate normally distributed factors" (p. 232).

superscript "*g*" indicates that the parameters in equation (2) correspond to the g^{th} population, and that another population may have a different set of values. Following equation (2), the expected value (mean), $E^g(X)$, and model-implied (co)variances, $\Sigma^g(X)$, of the observed indicators are

$$E^g(X) = \tau^g + \Lambda^g \kappa^g \text{ and } \Sigma^g(X) = \Lambda^g \Phi^g \Lambda^{g}{}' + \Theta^g \tag{3}$$

where $\boldsymbol{\kappa^g}$ is an $n^g \times 1$ vector of common factor means, $\boldsymbol{\Phi^g}$ is an $n^g \times n^g$ common factor covariance matrix, and $\boldsymbol{\Theta^g}$ is a $q^g \times q^g$ diagonal matrix of unique factor variances.

Within the common factor model framework, measurement invariance can then be evaluated in terms of aspects and levels of *factorial invariance* (FI) (Horn & McArdle, 1992; Meredith, 1993; Millsap, 1995, 2005; Widaman & Reise, 1997). Equations (2) and (3) facilitate the distinctions of various aspects and levels of FI. Note that the invariance of the parameters characterizing $\boldsymbol{\xi^g}$ (i.e., $\boldsymbol{\kappa^g}$ and $\boldsymbol{\Phi^g}$) is not considered as a part of the factorial invariance investigation. This is because it does not concern the relationships between $\boldsymbol{X}$ and ξ. Populations can be different along these parameters even when the measurement model is fully invariant. The invariance of $\boldsymbol{\kappa^g}$ and $\boldsymbol{\Phi^g}$ is about the similarity in the structure or nature of the latent variables (Byrne et al., 1989; Vandenberg & Lance, 2000). The measurement relationship between $\boldsymbol{X}$ and $\boldsymbol{\xi}$, on the other hand, is characterized by $\boldsymbol{\Lambda^g}$, $\boldsymbol{\tau^g}$, $\boldsymbol{\Theta^g}$ and, less obviously, the configuration of factor patterns.

Invariance of this last measurement aspect across populations is usually termed *configural invariance*. Configural invariance requires the same pattern of free and fixed factor loadings across populations (Vandenberg & Lance, 2000). According to Meredith (1993), this means that, across populations, "the same simple structure exists … in the

sense that zero elements are found in the same locations" in $\boldsymbol{\Lambda}^{g}$ and that corresponding non-zero loadings have the same sign (p. 540). Simply put, it requires the same configuration of zero and non-zero factor loadings (Bollen, 1989; Cheung & Rensvold, 2002; Horn & McArdle, 1992; Widaman & Reise, 1997; Vandenberg & Lance, 2000) and can be illustrated using Figure 1. Figure 1 represents the factor loading matrix ($\boldsymbol{\Lambda}^{g}$) of a 2-factor, 6-indicator common factor model between two groups (g = 1, 2). It represents a configural invariance situation because both groups have the same number of factors (i.e., 2), the same three indicators have non-zero free loadings (as represented by the λs) on the same factor, and all fixed zero loadings are also at the same location. Configural invariance is also called invariant cogeneric measurement properties (Vandenberg and Lance, 2000), invariance in factor forms (Bollen, 1989; Cheung & Rensvold, 1999; Rensvold & Cheung, 1998), or invariance of factor structures (Cole & Maxwell, 1985). It should be noted that in configural invariance no equality constraints are imposed between groups on the size of the "salient" loadings (loadings on a target factor); they are free to vary across groups. There are, however, equality constraints on the non-salient loadings, in all groups, because of the requirement of the "invariance in the zeros" (Horn & McArdle, 1992; Steenkamp & Baumgartner, 1998). It should also be noted that there are examples where configural invariance is evaluated on complex structure. For example, Widaman and Reise (1997) used an example in which the factor patterns, although the same across groups, included cross loadings (i.e., an indicator loaded on more than one factor). In any case, because the size of the salient loadings may still vary under configural invariance, between-group comparisons based on measurements that are only configural invariant is problematic.

$$\begin{array}{cc} \underline{\text{Group 1}} & \underline{\text{Group 2}} \\ \begin{bmatrix} \lambda_{11}^{1} & 0 \\ \lambda_{21}^{1} & 0 \\ \lambda_{31}^{1} & 0 \\ 0 & \lambda_{42}^{1} \\ 0 & \lambda_{52}^{1} \\ 0 & \lambda_{62}^{1} \end{bmatrix} & \begin{bmatrix} \lambda_{11}^{2} & 0 \\ \lambda_{21}^{2} & 0 \\ \lambda_{31}^{2} & 0 \\ 0 & \lambda_{42}^{2} \\ 0 & \lambda_{52}^{2} \\ 0 & \lambda_{62}^{2} \end{bmatrix} \end{array}$$

Figure 1. Illustration of configural invariance.

The next type of factorial invariance is the invariance in factor loadings across groups. That is, $\boldsymbol{\Lambda}^{1} = \boldsymbol{\Lambda}^{2} = \cdots = \boldsymbol{\Lambda}^{m}$ across m populations. An illustration is given in Figure 2 using the same model of Figure 1. Here, the superscripts are dropped indicating for each loading element only one value is needed for both groups. Because in the common factor model a factor loading is a regression slope and indicates the change in the observed indicator given one unit change in the latent factor, loading invariance indicates that a same amount of change in the latent factor will result in the same amount of change in the observed indicators across all groups. Loading invariance is also called metric invariance (Millsap & Kwok, 2004; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000), invariant factor pattern or factor pattern invariance (Alwin & Jackson, 1981; Jöreskog, 1971; Millsap, 1995), invariance in scaling units (Cole & Maxwell, 1985; Vandenberg & Lance, 2000), or weak factorial invariance (Widaman & Reise, 1997). However, the term metric invariance has also been used by some (e.g., Widaman & Reise, 1997) to refer to boarder concepts and the term factor pattern invariance can be mistaken as configural invariance. Thus for clarity and

because this research is limited within the common factor model framework, the term *loading invariance* will be used to refer to this scenario in the remainder of the dissertation.

$$\begin{array}{cc} \underline{\text{Group 1}} & \underline{\text{Group 2}} \\ \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} & \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} \end{array}$$

Figure 2. Illustration of loading invariance.

Intercept invariance refers to the condition in which the regression intercepts ($\boldsymbol{\tau}^g$) are equivalent across populations. This can be illustrated by Figure 3, in which the dropped superscript in the $\boldsymbol{\tau}^g$ matrix indicates that for each intercept element the value is the same in the two groups. The regression intercept is the mean of the corresponding observed indicator when the latent factor(s) is at zero and is termed by some the origins of the scale of the observed indicators (e.g., Alwin & Jackson, 1981). Substantively, non-zero intercept values have been viewed as an indication of systematic measurement bias (Cole, Maxwell, Arvey, & Salas, 1993; Hancock, 2001; Lubke & Dolan, 2003; Lubke et al., 2003a, 2003b; Mellenbergh, 1989; Oort, 1992). Lubke and Dolan (2003) and Lubke et al. (2003b) showed that, under the common factor model, group differences in intercepts can be attributed to the differences in the means of some unique factors. Group differences in the intercepts, then, imply a measurement is biased

toward one group over another because after controlling for latent factors, scores on the observed indicators would still be different. In other words, the observed scores are influenced differentially across groups by factors they are not designed to measure. If groups only differ in terms of measurement intercepts but not loadings, the observed scores of one group will be consistently higher than that of another, resulting in uniform bias. However, if both sets of parameters are different, the pattern of observed score differences will not be consistent, resulting in nonuniform bias (Lubke et al., 2003a, 2003b). Conversely, if both loadings and measurement intercepts are the same across groups, it amounts to *strong factorial invariance* (Meredith, 1993; Millsap, 2005; Widaman & Reise, 1997). Strong factorial invariance under the common factor model indicates that members from different groups will receive the same score on the observed indicators if they are equivalent on the latent factors. That is, the between group differences in observed indicators' means will unambiguously reflect the difference in the latent factor means. Strong factorial invariance is also called equality in scale metric or units of measurement (Rock, Werts, & Flaugher, 1978) or scalar invariance (e.g., Millsap & Kwok, 2004; Steenkamp & Baumgartner, 1998). In the remainder of this dissertation, the invariance in measurement intercepts will simply be referred to as *intercept invariance* and the combined condition of invariance in both intercepts and factor loadings as *strong factorial invariance*.