THE EFFECTS OF PARCELING ON TESTING GROUP DIFFERENCES IN SECOND-ORDER CFA MODELS: A COMPARISON BETWEEN MULTI-GROUP CFA AND MIMIC MODELS

A Dissertation

by

YUANYUAN ZOU

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

DOCTOR OF PHILOSOPHY

August 2009

Major Subject: Educational Psychology

THE EFFECTS OF PARCELING ON TESTING GROUP DIFFERENCES IN SECOND-ORDER CFA MODELS: A COMPARISON BETWEEN MULTI-GROUP CFA AND MIMIC MODELS

A Dissertation

by

YUANYUAN ZOU

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

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chair of Committee,	Oi-man Kwok
	Victor L. Willson
Committee Members,	Robert J. Hall
	Dudley L. Poston, Jr.
Head of Department,	Victor L. Willson

August 2009

Major Subject: Educational Psychology

ABSTRACT

The Effects of Parceling on Testing Group Differences

in Second-Order CFA Models:

A Comparison between Multi-Group CFA and MIMIC Models. (August 2009)

Yuanyuan Zou, B.A., Nanjing University; M.S., Baylor University

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

Using multi-group confirmatory factor analysis (MCFA) and multiple-indicator-multiple-cause (MIMIC) to investigate group difference in the context of the second-order factor model with either the unparceled or parceled data had never been thoroughly examined. The present study investigated (1) the difference of MCFA and MIMIC in terms of Type I error rate and power when testing the mean difference of the higher-order latent factor ($\Delta\kappa$) in a second-order confirmatory factor analysis (CFA) model; and (2) the impact of data parceling on the test of $\Delta\kappa$ between groups by using the two approaches. The methods were introduced, including the design of the models, the design of Monte Carlo simulation, the calculation of empirical Type I Error and empirical power, the two parceling strategies, and the adjustment of the random error variance.

The results suggested that MCFA should be favored when the compared groups were when the different group sizes were paired with the different generalized variances, and MIMIC should be favored when the groups were balanced (i.e., have equal group sizes) in social science and education disciplines. This study also provided the evidence that parceling could improve the power for both MCFA and MIMIC when the facotr loadings were low without bringing bias into the solution when the first-order factors were collapsed. However, parceling strategies might not be necessary when the factor loadings were high. The results also indicated that the two approaches were equally favored when domain representative parceling strategy was applied.

DEDICATION

To my parents

and those who helped me get this far in life

ACKNOWLEDGEMENTS

I would like to extend very special thanks to my chair Dr. Oi-man Kwok for his tremendous help during the year I spent working on the dissertation. Deep appreciation is also extended to my committee members, Dr. Victor Willson, Dr. Bob Hall, and Dr. Dudley Poston. I would never have finished this dissertation without their advice and encouragement. A special thanks also goes to Dr. Michael Speed for giving me and others such a wonderful SAS class and correcting the SAS codes for running the data analysis in my dissertation.

To my parents, especially my mom, who never went to college and never learned English, I owe my gratitude. The education they've given me since I was born has made me an independent, conscientious, efficient, and tough person. With these characteristics they gave me, as well as their constant belief in me, I came out of depression, started to pursue a graduate education in the U.S. six years ago, and ended up where I am today. My dear parents, I want to tell you that I will never let you down.

Finally, my love and heartfelt thanks go to my fiancé, Christopher McBride, for his love, support, and understanding during the four years I spent at Texas A&M University.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

TABLE Page

CHAPTER I

INTRODUCTION

Structural equation modeling (SEM), the foundation of confirmatory factor analysis (CFA), is one of the major analytic tools in the social sciences. One of the advantages of using SEM is that it can separate measurement errors from estimates of the parameters in the model (e.g., Bollen, 1989; James, Mulaik, & Brett, 1982; Kline, 2005). This is realized by separately modeling latent constructs and latent error variables so that the estimates of the model parameters can be made based on the "error-free" latent constructs, instead of observed measures that are potentially contaminated by systematic and random errors.

Two types of CFA models are most widely used in empirical and analytical studies, i.e., the first-order factor model and the hierarchical factor model. In the first-order factor model, each latent factor is measured by the indicators, and the latent factors are assumed to be correlated with each other. An example of a first-order factor model of behavioral and emotional problems among adolescents (Windle & Mason, 2004, p. 54) is shown in Figure 1.1. In this model, each of the latent factors is measured by two or more indicators and the correlation between each pair of the latent factors is freely estimated. Note that a single-factor CFA model needs at least three indicators to be

---

This dissertation follows the style of *Structural Equation Modeling*.

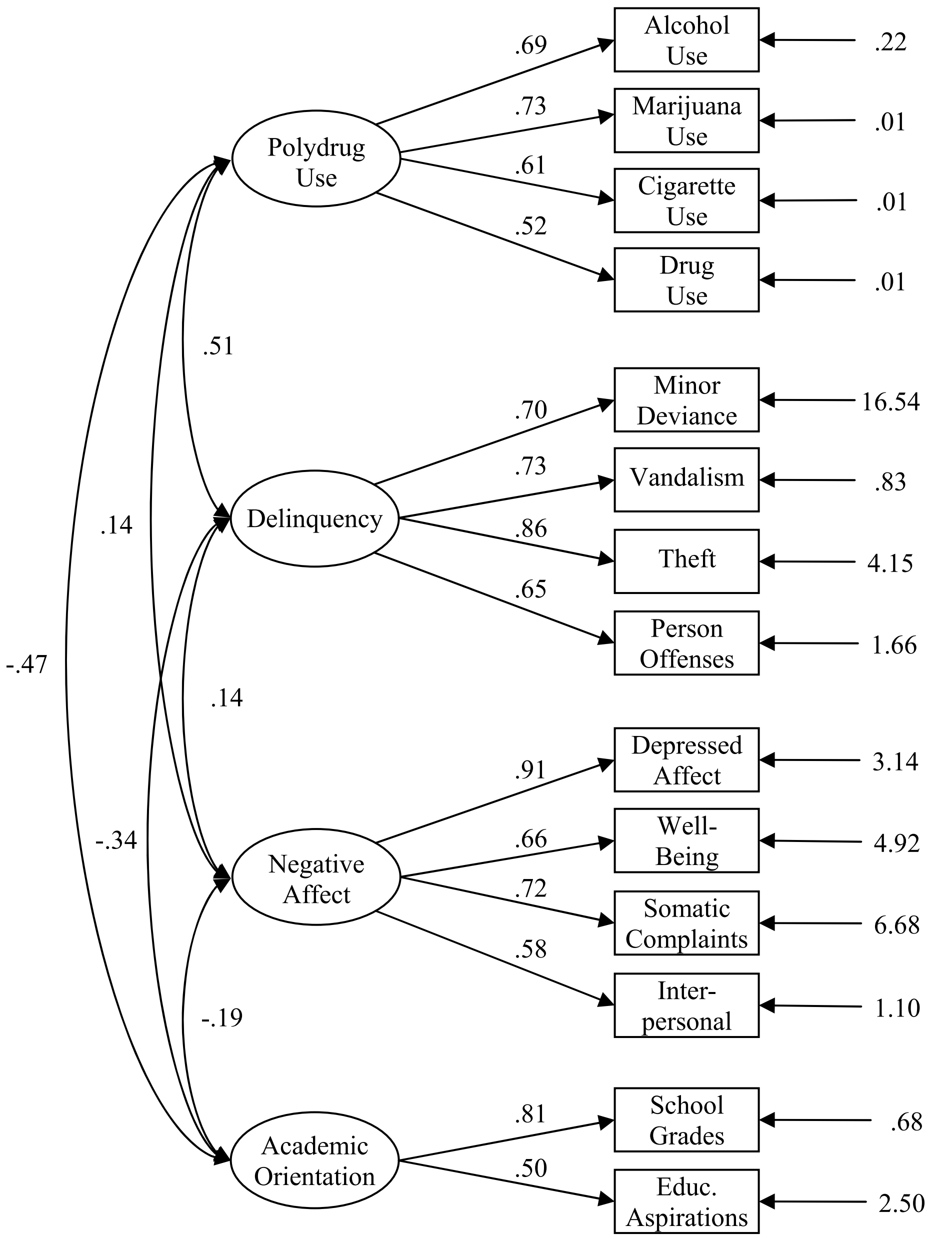


Fig. 1.1 An Empirical Example of the First-Order CFA Model

uniquely identified. Therefore the construct “academic orientation” is not identified as an individual model, but the whole model in Figure 1.1 is over-identified because the total degrees of freedom is greater than the total number of parameters that need to be estimated.

The second type of CFA model is termed the hierarchical CFA model, or second-order CFA model, in which the lower-order latent factors that are measured by the indicators have one or more common direct causes. Because a common cause is measured by the lower-order latent factors, it is also called a second-order factor, and the lower-order factors are called the first-order factors. Figure 1.2 shows an example of a second-order CFA model (Meuleners, Lee, Binns, & Lower, 2003, p. 287). In this model, each of the five first-order factors, namely, “Psychological”, “Environment”, “Social”, “Opportunities for growth”, and “Health” is measured by at least three items. Meanwhile, the five first-order factors are caused by a common factor “Quality of life”, which is the second-order factor in the model.

First-order and second-order CFA models, as well as the other CFA models, are both developed based on factor analysis with the goal of explaining the common variance among a -large number of correlated variables by a small number of factors (Tinsley & Tinsley, 1987). Therefore, the two types of models are closely related and share many characteristics. Firstly, the observed variables are linear functions of the latent variables with the variable-factor regression coefficients or loading as weights. Secondly, all the latent factors in these two types of models should be interpreted in relation to the observed variables. This characteristic also holds for the second-order

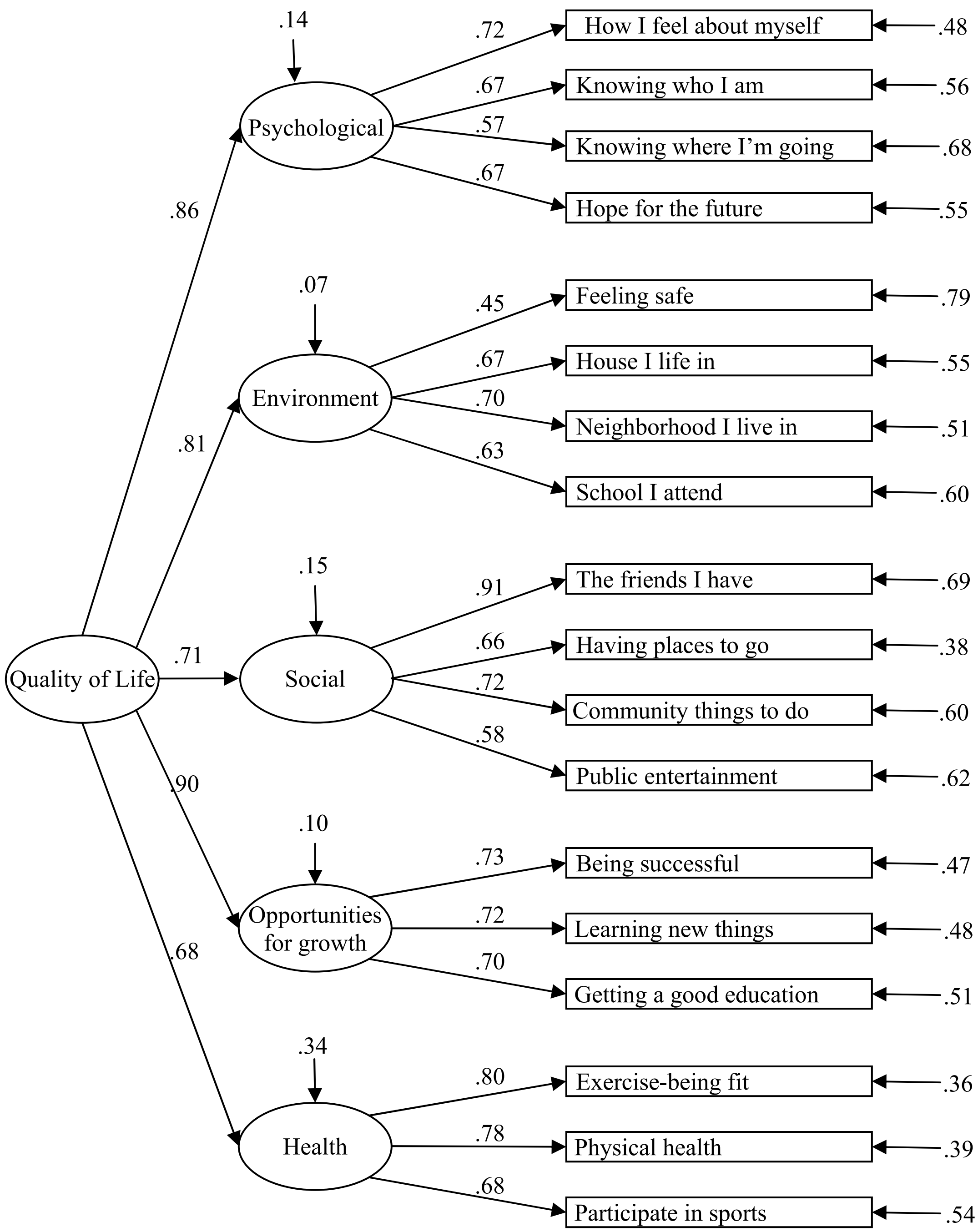


Fig. 1.2 An Empirical Example of the Second-Order CFA Model

factor in the second-order CFA model although it is derived from the first-order factors (Gorsuch, 1983). Moreover, the two models both usually assume that the residual variables are not correlated with each other or with the common latent factors. However, second-order CFA models are more easily understood than the first-order factor models with fewer second-order factors accounting for the common variances among the observed variables. In addition, the data may be interpreted by using the two types of CFA models from different perspectives. Thompson (1990) described this difference as:

> The first-order analysis is a close-up view that focuses on the details of the valleys and the peaks in mountains. The second-order analysis is like looking at the mountains at a greater distance, and yields a potentially different perspective on the mountains as constituents of a range. Both perspectives may be useful in facilitating understanding of data. (p. 579).

Regardless of the fact that the first-order and second-order CFA models are very commonly used in multivariate analysis in different social sciences, the second-order CFA model has been less employed than the first-order factor model, particularly in two areas. The first one involves group comparison. There are two major approaches for testing group difference in CFA models, multi-group CFA (MCFA) and multiple-indicator-multiple-cause (MIMIC). The advantage of using MCFA is that it can examine potential group differences in different parts of the factor model through the test of measurement invariance (Widaman & Reise, 1997). However, compared with the MIMIC model, the MCFA model requires a relatively larger sample size because more parameters may need to be estimated. The MIMIC model, on the other hand, requires a smaller number of observations but imposes more constraints (i.e., assuming a strict invariance condition) when testing group difference on the target construct. The two approaches are both widely used in psychology and education area to investigate the group difference across different population. Some examples among recent studies include the following: Byrne and Stewart's (2006) comparison of the responses to a depression inventory for Hong Kong and American nonclinical adolescents in a second-order factor model using MCFA; Anderson et al.'s (2005) MIMIC approach in a first-order factor model to investigate if some drinking behaviors are differentiated between genders and schools; Agrawal and Lynskey's (2007) use of the two approaches to investigate the gender difference about cannabis abuse and dependency in a first-order factor model.

Hancock, Lawrence, and Nevitt (2000) compared MCFA and MIMIC in the context of first-order factor models. They found that there was very little difference between the two approaches across groups in terms of power or Type I error under balanced design condition. On the other hand, MCFA was preferred as the disparity of sample sizes increased. However, the difference between the two approaches in the context of second-order factor models had not yet been well examined.

The second area in which the second-order CFA model has been less employed is related to data parceling. Parceling (i.e., combining items to form parcels/composites) is a very common technique in educational and psychological studies, and the method of creating parcels for SEM analysis has received increasing attention in recent years.

Parcel is a simple sum (or mean) of several items that are used to assess the same construct (Kishton & Widaman, 1994), which is also referred to as a *testlet* (Wainer & Kiely, 1987) or a *miniscale* (Prats, 1990). Compared with models based on item-level data, models based on parceled data have the advantages that (1) they are more parsimonious (i.e., have fewer parameters for estimation because one parcel is a single indicator that requires only one loading estimate) and therefore require a smaller sample size (Bagozzi & Edwards, 1998, Little, Cunningham, & Shahar, 2002); (2) they generally display better model fit than the solutions based on item-level data (i.e., disaggregated data), especially when either the assumption of normality or the continuity of item-level data is violated (Bandalos, 2002, Little et al., 2002, Bagozzi & Edwards, 1998); and (3) they have fewer chances for residuals to be correlated because the corresponding reliability is increased and the unique variances are smaller by parceling the items (Little et al., 2002). A major advantage of using parcelling for second-order CFA models is that it can reduce the complexity of the model and requires smaller sample size when conducting multiple group comparison.

The effects of parceling on model fit and parameter estimate have been thoroughly investigated in the context of single group first-order CFA models (e.g., Kishton & Widaman, 1994; Nasser & Takahashi, 2003; Nasser & Wisenbaker, 2006; Sass & Smith, 2006), particularly those with some variables loaded on a secondary factor (e.g., Hall, Snell, & Singer-Foust, 1999; Bandalos, 2002; Bandalos, 2008). However, very few studies have been conducted that examine the effects of parceling on group comparison in either first-order or second-order CFA model. Given the fact that second-order CFA models have gained popularity and that parceling strategy has been receiving increasing attention in education and other social science disciplines, examining the effects of parceling in the second-order CFA model on group comparison may provide important guidance on its use.

This article is organized as followed. In Chapter II, the theoretical framework and previous research including second-order CFA models, MCFA and MIMIC, and parceling were reviewed to set up a necessary backgroup for the current investigation. The current investigation based on these techniques, which included two studies (i.e., Study 1 and 2) were introduced in Chapter III. Chapter IV contained the method, results, and discussion of Study 1, and Chapter V contained the method, results and discussion of Study 2. Finally the conclusion was made in Chapter VI.

CHAPTER II

LITERATURE REVIEW

In this chapter, the theoretical framework and previous research on second-order factor models, the two model comparison approaches (i.e., Multi-group CFA and MIMIC), and parceling strategies were reviewed separately.

SECOND-ORDER CFA MODEL

Second-order CFA models is applicable when (1) the first-order factors are highly correlated, and (2) there is a hypothetical second-order factor that can account for the associations among the first-order factors (Chen, Sousa, & West, 2005). Compared with the first-order CFA model, the second-order CFA model has advantages that it is more parsimonious when there are four or more first-order factors[1], and it can be used for investigating reliability and validity by separating specific and unique variance estimates (Rindskopf & Rose 1988). The most famous example of second-order factor models would be the theory of general intelligence factor (abbreviated *g*), which was proposed by Spearman (1927). In his study of schoolchildren's academic achievement, he claimed that there was a dominant intelligence factor (i.e., *g*) that could account for the correlations between the seemingly unrelated school subjects. Spearman's *g* factor model can be reformulated by a first-order factor model with *g* as the latent factor and all the school subjects are loaded on this *g* factor directly. Spearman's *g* theory was

[1] Assuming the residuals are all uncorrelated with each other and with the factors, the first-order factor model with three first-order factors is as parsimonious as the corresponding second-order factor model.

criticized in that it ignored the potential group/specific factors corresponding to various abilities, e.g., perceptual organization, memory, and verbal ability. The development of cognitive tests and analytical techniques improved the conception of *g* and led to a second-order factor model with *g* as the second-order factor and the specific intelligence-related abilities (each of which is assessed by multiple items) as the first-order factors, which has been widely although not universally accepted.

In a second-order factor model, each first-order factor is represented as a linear function of the second-order factor and the residual, and each observed variable is represented as a linear function of the corresponding first order factor plus the measurement error. Suppose there is a second-order factor model with one second-order factor and four first-order factors, and each of the first-order factors is associated with three observed variables (Figure 2.1). Then, the twelve observed variables $\underline{y}$ for individuals can be expressed as:

$$\underline{y} = \underline{\mu} + \Lambda\underline{\eta} + \underline{\varepsilon} \tag{1}$$

where both $\underline{y}$ and $\underline{\mu}$ are a 12×1 vector with $\underline{y}$ contains the 12 observed scores and $\underline{\mu}$ contains the intercept values for the 12 observed variables, Λ is a 12×4 matrix containing the loadings of the observed variables on the first order factors, $\underline{\eta}$ is a 4×1 vector containing the theoretical scores for the individual on the first-order factor, and $\underline{\varepsilon}$ is a 12×1 vector of unique variables/errors of the observed variables. The 4×1 vector

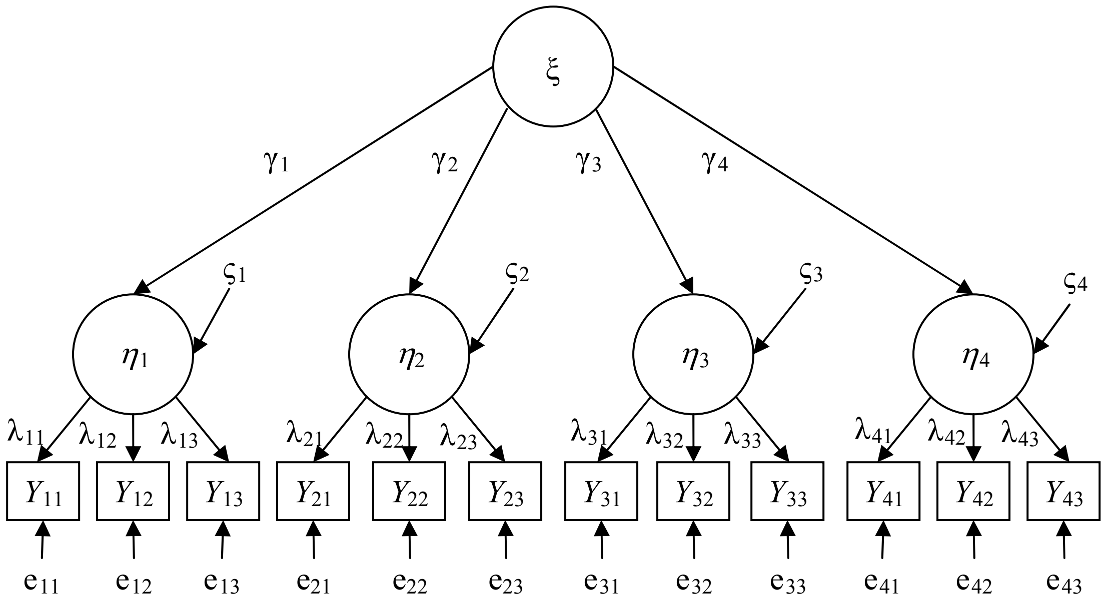


Fig. 2.1 A Conceptual Second-Order Factor Model

Note. ξ is the second-order factor, γ_i is the loading of the first-order factor on the second-order factor, η_i is the first-order factor, ς_i is the residual of the first-order factor, λ_{ij} is the loading of the observed variable on the first-order factor, Y_{ij} is the observed variables, and e_{ij} is the random error of the observed variable (i = 1 - 4, j = 1-3).

of the first-order factors $\underline{\eta}$ that are loaded on the second-order factor can also be expressed as:

$$\underline{\eta} = \underline{\nu} + \underline{\Gamma}\xi + \underline{\zeta} \tag{2}$$

where $\underline{\nu}$ is a 4 × 1 vector of intercept values for the first order factors, $\underline{\Gamma}$ is a 4 × 1 vector containing the loadings of the first-order factors on the second-order factor, ξ represents the score of the second-order factor, and $\underline{\zeta}$ represents the 4 × 1 vector of the residuals in the first-order factors. The intercepts $\underline{\mu}$ and $\underline{\nu}$ are usually constrained as zero for simplicity and identification purpose. The corresponding variance-covariance matrices for the four first-order factors (i.e., η in Equation 2) and the 12 observed variables (y in Equation 1) are:

$$\Sigma(\eta) = \Sigma = \underline{\Gamma}\Phi\underline{\Gamma}'+\Psi \tag{3}$$

and

$$\Sigma(y) = \Sigma = \Lambda(\underline{\Gamma}\Phi\underline{\Gamma}'+\Psi)\Lambda'+\Theta_{\varepsilon}\,. \tag{4}$$

In both equations, Φ , Ψ and Θ_{ε} represent the variance-covariance matrix of the second-order factor, residuals of the first-order factor and errors of the observed variables, respectively.

Rindskopf & Rose (1988) provided a theoretical framework for using a second-order factor model to investigate reliability and validity of the observed variables. The variance-covariance matrix of the observed variables (Equation 4) can be rewritten as:

$$\Sigma = \Lambda\underline{\Gamma}\Phi\underline{\Gamma}'\,\Lambda'+\Lambda\Psi\Lambda'+\Theta_{\varepsilon} \tag{5}$$

which displays explicitly that the variance of each observed variables consists of three parts. The first part $\Lambda\underline{\Gamma}\Phi\underline{\Gamma}'\,\Lambda'$ is the common variance of the 12 observed variables. The second part $\Lambda\Psi\Lambda'$ is the specific variance or systematic error variance from the unknown factors that are specified as the residuals of the first-order factors in the model, typically assumed to be independent of each other, although potentially forming additional factors if the model is misspecified. And the third part Θ_ε, as stated before, is the variance from the random error. Note that for the model depicted as above, Θ_ε is a diagonal matrix under the assumption that measurement errors are mutually uncorrelated.

In the classic test theory, reliability can be mathematically defined as the ratio of the true score variance to the total variance, in which the true score variance is the sum of the common variance and specific variance. That is, the reliability of the *i*th observed variable can be estimated as the *i*th element of (Rindskopf & Rose, 1988, p. 64, Equation 5):

$$Diag(\Lambda\underline{\Gamma}\Phi\underline{\Gamma}'\Lambda' + \Lambda\Psi\Lambda') / Diag(\Sigma) \tag{6}$$

Rindskopf and Rose (1988) also defined “measure validity” as the ratio of the common variance to the total variance and “method validity” as the ratio of variance of the first-order factors due to the second-order factor to the total variance of the first-order factor. The two validities can be expressed as (Rindskopf & Rose, 1988, p. 64, Equation 6):

$$\text{Measure validity: } Diag(\Lambda\underline{\Gamma}\Phi\underline{\Gamma}'\Lambda') / Diag(\Sigma) \tag{7}$$

and (Rindskopf & Rose, 1988, p. 65, Equation 7)

$$\text{Method validity: } Diag(\underline{\Gamma}\Phi\underline{\Gamma}') / Diag(\underline{\Gamma}\Phi\underline{\Gamma}'+\Psi) \tag{8}$$

respectively. They further explained that measure validity answered the question "how well the observed variables measure the second-order factors" (p. 68) and method validity answered the question "how well the first-order factors measure the second-order factors" (p. 66).

However, using second-order factor models to estimate the reliability and validity as demonstrated by Rindskopft has some deficiencies. Firstly, there are many validities, e.g., content validity, concurrent validity, and predictive validity, discriminant validity, or convergent validity. The two validities specified by Rinskopft and his colleague (i.e., measure validity and method validity) belong to construct validity. The validity calculated by using Equation 7 actually provides the proportion of the variation/variance in the empirical observed variables that is due to the underlying second-order factor. However, whether the empirically identified construct is consistent with the theoretical construct is not determined by such computations. Secondly, the specific variance that is separated out by the second-order factor model is actually the common variance for the observed variables that have the same first-order factor. Each observed variable under the same first-order factor, however, may have its own specific variance that is not shared with the others. This variance is mixed with the true random error and cannot be identified by the second-order factor model. Therefore, the reliability calculated by using Equation 6 might be smaller than the total true score reliability.

MULTI-GROUP CFA

Multi-group confirmatory factor analysis (MCFA) is the most commonly used technique for testing measurement invariance (MI) and particularly, factor invariance between different populations. It can be used for comparing not only the covariance structures but also mean structures across the groups (Meredith, 1993; Widaman & Reise, 1997). Because the theoretical frameworks for applying MCFA in first- and second-order factor models are analogous to each other, the theory of testing invariance using MCFA on first-order factor models will be introduced first, which is the basis of the theory of second-order factor models.

First-order CFA Models

Widaman and Reise (1997) provided a theoretical framework for applying MCFA to test MI in first-order factor models. This theoretical framework, however, is still applicable in the second-order factor model context. They distinguished four types of different MIs with increasingly strict constraints imposed: 1) configural invariance, (i.e., identical patterns of fixed and free factor loadings across groups), 2) weak invariance (i.e., configural invariance plus the equality of factor loadings across groups), 3) strong invariance (i.e., weak invariance plus the equality of intercepts across groups), and 4) strict invariance (strong invariance plus equality of residual variances across groups).

Widaman and Reise (1997), as well as Mererdith (1993), argued that the covariance model in Equations 3 and 4 was only applicable when testing weak invariance. If a higher-order level of MI needs to be tested, the model for a first-order CFA model should be modified as (Widaman & Reise, 1997, p. 288, Equation 8):

$$M = \tau\tau' + \Lambda(\alpha\alpha' + \Psi)\Lambda' + \Theta_{\varepsilon}. \tag{9}$$

In this model, M is a p × p moment matrix (p = the number of observed variables), τ is the intercept, α is the mean of each latent factor, Ψ is an m × m matrix of covariance among latent factor scores (m = the number of latent factors), and Θ_{ε} is a p × p matrix of covariance among the measurement residuals. Note that the moment matrix *M* is similar to but not exactly a variance-covariance matrix. Each element in *M* also contains the information about the means of the two variables.

In the moment matrix specified in Equation 9 (or the corresponding variance-covariance matrix), if the Λ matrix holds for each group, weak MI is achieved, which indicates that the correlation between the latent variables and the observed variables in one group is the same as the corresponding correlations in the other groups. If the intercept matrices $\tau\tau'$ equalities hold in addition to equality of the Λ matrices across the groups, then the groups exhibit strong MI. Under this condition, because the scores from the different groups have the same weights (factor loadings) and origin (intercepts), the group mean difference ($\Delta\kappa$)on the latent variables can be tested. If constraints are invoked on θ_{ε} in addition to Λ and $\tau\tau'$, the groups have strict MI. When strict MI holds, the variances of the latent factors as well as the means, are comparable across the groups. Nevertheless, strict MI is rarely achieved in educational and psychological research.

Second-order CFA Models

For testing MI in second-order CFA models, additional constraints are required, including the constraints on the loadings and intercepts of all the second-order factors, and the specific variances (as defined earlier) of the first-order factors. One approach for comparing groups on the second-order CFA models with increasingly strict constraints is: (1) configural invariance; (2) constraining loadings of all the first-order factors; (3) constraining loadings of all the second-order factors; (4) constraining latent intercepts of the observed variables; (5) latent intercepts of the first-order factors, (6) residual variances of the first-order factors (i.e., specific variances), and (7) measurement error variances of the observed variables.

Because the models of these seven hierarchical constraint steps are nested from 1 to 7, chi-square difference test can be conducted to test each level of MI. A non-significant chi-square difference test indicates that the constraints of a specific tested step are not statistically different between groups and more restricted constraint step can further be examined. On the other hand, a significant chi-square difference test indicates that at least one or more constraints are not equal between groups and no further test step should be performed. Theoretically, a strong MI (with both loadings and latent intercepts of the observed variables to be the same between groups) is the prerequisite for obtaining a valid test on the mean difference of the latent factor(s) between groups in first-order CFA. Similarly, this is also the required condition for testing valid group difference in second-order factors: the loadings of the second-order factors and the intercepts of the first-order factors should also be invariant across the groups in addition to the strong MI in the first-order part. Note that invariant specific variances or measurement error variances are not necessary when comparing the means of the second-order factors, although these (residual) variances, combined with other factors such as group size, might impact the Type I error rate and the empirical power.

MULTIPLE-INDICATOR-MULTIPLE-CAUSE (MIMIC)

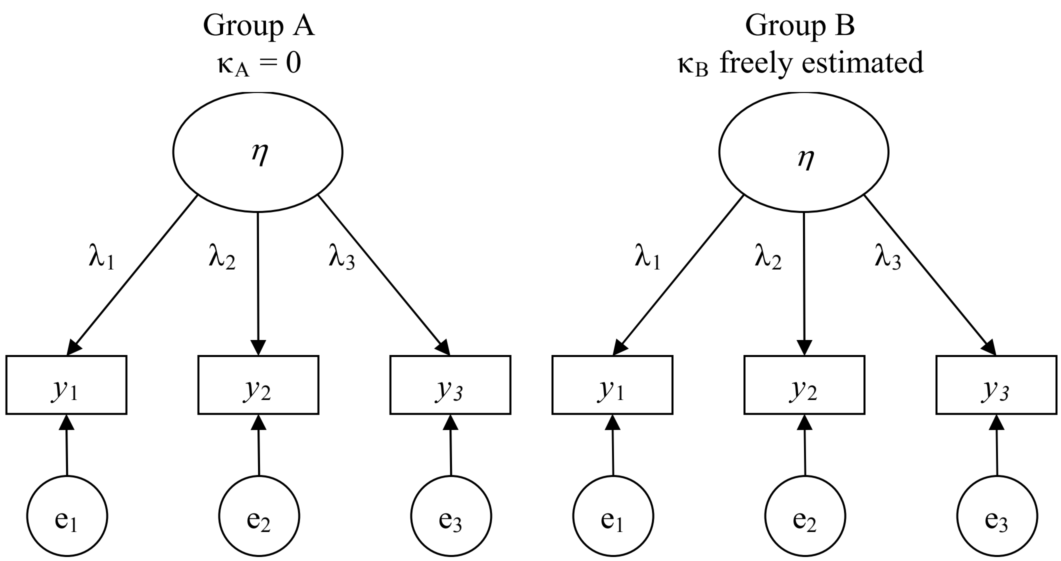
A Multiple-Indicator-Multiple-Cause (MIMIC) model is defined as a measurement model with both cause indicators and effect indicators (Kline, 2005). Instead of fitting a model to different groups separately like MCFA, MIMIC combines them and incorporates the membership variables as the cause indicators into the model. For example, suppose there is a simple first-order factor model with one latent variable and three observed variables, and the goal is to compare the factor mean of this model between two groups. The grouping/membership variable, X, is coded as 0 and 1, and used as the exogenous variable/cause indicator. The overall model can be presented as:

$$\eta = \nu + \gamma X + \zeta \tag{10}$$

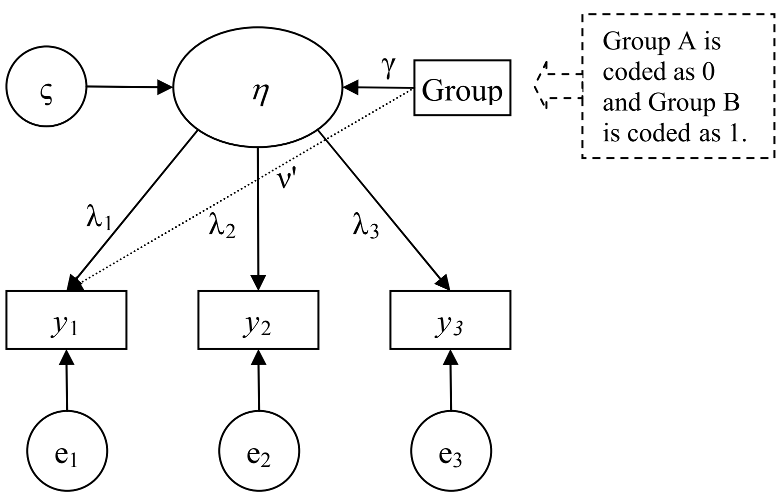
where η is the first-order factor, ν is the latent intercept (or the mean of the group with X coded as 0), ζ is the residual of the latent factor and γ is actually the mean difference of the latent factor between the two groups, which can be examined by the Wald z test. The difference of the model structure by using MCFA and MIMIC is presented in Figure 2.2. For second-order factor models, the dichotomous variable representing group membership can be included as the cause indicator on the second-order factor and the group mean difference can be examined by the Wald Z test of the corresponding path coefficient (i.e., the path from the membership variable to the second-order factor). The model is the same as Equation 10 except that all the elements in the model are related to the second-order factor.

As described previously, MIMIC model requires a very strong assumption of the homogeneity of the Σ matrices (i.e., usually assuming a strict MI applies to all groups), which is apparently more rigorous than the assumption of MCFA. This assumption is even stronger when applying MIMIC in a second-order CFA model than a first-order factor because strict MI needs to be fulfilled in both levels of the second-order CFA model. Indeed, among the recent papers that have been reviewed, group comparison in the context of second-order CFA models has been conducted more frequently by using MCFA than MIMIC. Another deficiency of the MIMIC approach is that MIMIC has less flexibility in being able to investigate the sources of heterogeneity between two models than MCFA. That is, MIMIC cannot test if the groups have different model structures and factor loadings[2]. However, because the MIMIC approach results in only one model by combining data from two groups, fewer parameters are estimated and a relatively

[2] MIMIC can test strong MI (i.e., equal intercepts across groups) by adding the path from the grouping variable to the observed variables (see Figure 2.2). The path coefficient is the difference in the intercepts between groups, which can be examined by the Wald z test.



MCFA: chi-square difference test could be conducted to test each level of MI (i.e., configural, weak, strong, and strict MI) by imposing the hierarchical constraints. When comparing the factor means, strong MI should hold. Then the mean of the factor in one group (i.e., Group A in the figure) is constrained as zero and the mean of the factor in other group (i.e., Group B in the figure) is freely estimated.



MIMIC: Data from the two groups are combined to fit one model and a dichotomous variable representing the group information is added. The path coefficient γ is the mean group difference between the two groups, which can be examined by the Wald z test. Strong MI can also be tested by MIMIC model. The path coefficient ν' is the difference in the intercepts between the two groups, which can also be examined by the Wald z test.

Fig. 2.2 Comparing the Model Structures of MCFA and MIMIC smaller sample size (compared with the MCFA approach) is required for the purpose of estimability and convergence.

PARCELING

In this section the theoretical framework for improving model fit by using parceling, the methods of parceling presented by different researchers, residual adjustment for parcels, and the cons of parceling were reviewed.

Theoretical Framework for Improving Model Fit

Bandalos and Finney (2001) reported that among the reasons researchers cited for using/creating item parceling, increased reliability was cited as the most frequently (29%). Little et al. (2002) provided the theoretical framework for using parcels in SEM to increase the reliability of the indicators. They stated that any given indicator of a construct could be represented as:

$$X_i = T_i + S_i + e_i \tag{11}$$

where X_i is the observed score of the indicator, T_i represents the target construct true score, S_i represents a specific component true score unrelated to the construct, and e_i represents the random error that is unreliable. Therefore, the variance of an observed variable can be divided into three components, common variance, specific variance, and unreliable variance, corresponding to the variance of T_i, S_i, and e_i, respectively. If the observed score X is standardized (i.e., the variance equals to 1) and T_i, S_i, and e_i are independent from each other, the relationship of the common variance, the specific variance, and the unreliable variance can be expressed as (Crocker & Algina 1986, p. 295):

$$T_i^2 + S_i^2 + e_i^2 = 1 \tag{12a}$$

In this equation, T_i^2, S_i^2, and e_i^2 have values between 0 and 1. Therefore they can be thought of as proportions, and also can be called communality, specificity, and unreliability of the variable, respectively. Among the three components, T_i^2 and S_i^2 are reliable and the sum of them is called reliability[3] of the variable. When items are parceled, S_i and e_i (consequently S_i^2 and e_i^2) could be reduced, or even canceled out theoretically if there are infinite indicators, because S_i and e_i are assumed to be uncorrelated. Given that the sum of the three components is 1, the reduction in both S_i^2 and e_i^2 increases the communality T_i^2. Due to the larger increment in in T_i^2 than the decrement in S_i^2, the reliability (i.e., $T_i^2 + S_i^2$) will eventually be increased. Note that Little, Lindenberger, and Nesselroade (1999) proposed a different relationship of the three components and it was also applied in Little et al. (2002)'s study. They argued that

[3] Reliability, Communality, specificity, and unreliability are all proportions. They are calculated as the ratio of reliable variance, common variance, specific variance, and random error variance to the total variance, respectively.

the reliability of an item was the square root of the sum of its squared communality and squared specificity, and the sum of the reliability and unreliability is equal to 1:

$$\sqrt{(T_i^2)^2 + (S_i^2)^2} + e_i^2 = 1. \tag{12b}$$

Although the two equations (12a & 12b) express the different relationship of the three components and different mathematical definition of reliability, they consistently display that decreasing S_i^2 and e_i^2 can result in increasing the indicators' reliability in the model.

Previous studies (e.g., Kishton & Widaman, 1994; Little, et al., 2002; Gribbons & Hocevar, 1998; Nasser and Takahashi 2003) have shown that parceling could enhance the model fit by improving the reliability of the indicators. This finding can be supported by examining the sample covariance matrix. For example, in a second-order CFA model, the sample covariance matrix is shown in Equation 4 as $\Sigma = \Lambda(\Gamma\Phi\Gamma' + \Psi)\Lambda' + \Theta_\varepsilon$. As the diagonal elements in the matrix Ψ and Θ_ε decrease by parceling, the sample covariance matrix will approach closer to the population covariance matrix $\Sigma = \Lambda(\Gamma\Phi\Gamma')\Lambda'$ than the sample covariance matrix before parceling does, and then the model fit will be enhanced.

MacCallum, Widaman, Zhang, and Hong (1999) provided a general theoretic framework demonstrating the sources that can potentially threaten model fit. Suppose there is a first-order factor model with common and unique factors in it. The population covariance matrix can be expressed as:

$$\Sigma_p = \Lambda\Sigma_{cc}\Lambda' + \Lambda\Sigma_{cu}\Theta' + \Theta\Sigma_{uc}\Lambda' + \Theta\Sigma_{uu}\Theta' \tag{13}$$

in which Σ_p represents the population covariance, Λ is the matrix for the common factor loadings, Θ is the matrix for the unique factor loadings, Σ_{cc} is a covariance matrix of the common factors, Σ_{cu} is a covariance matrix of the common factors and unique factors, Σ_{uc} is a transpose of Σ_{cu}, and Σ_{uu} is a covariance matrix of the unique factors. Common factors and unique factors, unique factors with each other are usually assumed to be uncorrelated in a population. Therefore the model can be simplified as (p. 87, Equation 10):

$$\Sigma_p = \Lambda\Phi\Lambda' + \Theta_u \tag{14}$$

if all the factors are defined as being standardized. In this equation Φ is the covariance matrix for the common factors, and Θ_u represents the matrix for the unique factors, which is a diagonal matrix and has diagonal entries representing the variance in each variable that is not accounted for by the common factors. However, this model may not hold in empirical samples. MacCallum and his colleagues showed that two sources of sampling error may appear and negatively influence the model fit when the model is built based on the sample covariance matrix. They are the covariances between unique factors and the covarainces between the common and unique factors. Therefore, $\Lambda\Sigma_{cu}\Theta'$ and $\Theta\Sigma_{uc}\Lambda'$ cannot be eliminated from Equation 13, and Σ_{uu} is not an identity matrix. The sample factor analysis model would then be expressed as (p. 88, Equation 13):

$$\Sigma_s = \Lambda S_{cc}\Lambda' + \Lambda S_{cu}\Theta' + \Theta S_{uc}\Lambda' + \Theta S_{uu}\Theta' \qquad (15)$$

where S_{cc}, S_{cu}, S_{uc} and S_{uu} and are the matrices Σ_{cc}, Σ_{cu}, Σ_{uc}, and Σ_{uu} suggested by the sample data, respectively. Θ and Λ are defined as in Equation 13. These non-zero elements in the matrices of S_{cu}, S_{uc}, and S_{uu}, which are due to the random sampling fluctuations, will result in lack of fit in a model to the empirical sample data. Bandalos and Finney (2001) proposed that parceling could enhance model fit through three aspects. Firstly, as Little et al. (2002) stated, parceling could reduce specific variance by eliminating the unique component in scores. Consequently the size of the matrices S_{cu}, S_{uc}, and S_{uu} will be reduced and the model fit will be less penalized for these unmodeled associations in the population matrix. Secondly, as parceling reduces the magnitude of the specific variances, the unique factor loadings will also be decreased. Note that the unique factor loadings in Θ serve as weights for the matrices S_{cu}, S_{uc}, and S_{uu}. As the weights decrease, the three matrices will have less contribution to the solution for Σ_s. Lastly, some parceling strategy in a specific condition could improve model fit in another way. By applying a distributed parceling strategy, the items that share the unique factor (e.g., a method factor) are distributed in each parcel. When these parcels are treated as the indicators of the same common factor, the originally correlated unique variances (i.e., off-diagonal elements in the residual matrix) will turn into the common variances and become part of the variances on the diagonal of S_{cc} instead of the off-diagonal elements in the S_{uu} matrix. As a result, the model can display better fit to the sample.

However, the second way that parceling increases model fit proposed by Bandalos and Finney (2001) is problematic. When parceling decreases the unique variance, the common variance will increases simultaneously. These two mutual changes will result in a decrease of unique factor loadings and an increase of common factor loadings. Just as unique factor loadings serve as weights for the matrices S_{cu}, S_{uc}, and S_{uu}, the common factor loadings in Λ also serve as weights for the matrices S_{cu} and S_{uc}. Therefore, the combined effect of Θ and Λ resulting from parceling on the two matrices S_{cu} and S_{uc} will need further investigation.

Parceling could also improve model fit and provide a more reliable solution by mitigating the problems of nonnormality and noncontinuity. Multivariate normality is one of the major assumptions for using the maximum likelihood (ML) estimator to estimate the coefficients in CFA models. Although there are alternative estimators (e.g., asymptotically distribution-free ADF) when the assumptions of ML are violated, they usually require large sample sizes. For example, West, Finch, and Curran (1995) recommended the sample size of 1000-5000 for using ADF estimator. When sample sizes are small, these estimators become problematic. However, if the items that have skewness and kurtosis in opposite directions are parceled, the skew and kurtosis could be cancelled out and the parceled data may display a more normal distribution, for which the ML is appropriate. This effect has been validated by using empirical data (e.g., Nasser & Takahashi, 2003; Thompson & Melancon, 1996) and simulated data (e.g., Bandalos, 2002; Bandalos, 2008; Hau & Marsh, 2004). Similarly, parceling improves continuity because "scale intervals increase in number and effectively become both smaller and more equal with regard to the distances between points as more items are aggregated" (Little et al. 2002, p. 157). Therefore the models based on parceled categorical data evidence better fit than the ones based on unparceled categorical data (e.g., Bandalos, 2008).

Contrary to the studies listed above, Marsh, Hau, Balla, and Grayson (1998) reached a different conclusion about parceling and model fit. They argued that although the fit indices for the model based on parcels were higher and the model were more likely to yield a proper solution than one based on the same number of individual items (e.g., two parcels vs. two items, six parcels vs. six items), the model based on all the individual items (i.e., twelve items in their study) was modestly better than the those based on parcels. Little et al. (2002) countered that their simulation "was not a fair test" (p. 161). The data in Marsh et al.'s (1998) study was extreme, e.g., all the factor loadings above 0.6. Under such a condition the difference between the models based on items and parcels may be negligible. Bandalos and Finney (2001) also stated the improvement in model fit "is most marked for items with low communalities" (p. 284).

Methods of Parceling

Kishton and Widaman (1994) distinguished two alternative parceling methods. One is called unidimentional, or internal consistency parceling, which parcels the items that share the same facet . The other one is domain representative parceling, which creates parcels by combining items from different facets (Figure 2.3). They compared the two parceling strategies in terms of model fit and coefficient estimates in CFA models and concluded that both methods resulted in appropriate model fit, but domain representative parceling provided better coefficient estimates.

The two methods are similar to the two commonly used parceling methods defined by Hall et al. (1999), namely, isolated and distributed parceling (Figure 2.4). However, isolated and distributed parceling strategies are usually referred to when the items from the same construct are also loaded on a secondary factor in addition to the primary factor. When conducting isolated parceling, the items with secondary loadings as well as primary loadings are combined into the same parcel, whereas in the isolated parceling strategy, the items with secondary loadings as well as primary loadings are put into parcels with items that are not loaded on the same secondary factor. Hence, the influence of the secondary factor is distributed.

Bagozzi and Edwards (1998) proposed three parceling (or aggregating as in their study) methods based on the depth of aggregation. In order to define the three parceling methods, they presented an example involving a CFA model with 16 items and 4 components of a scale, each of which is measured by 4 items. The three parceling methods, in an increasing order according to the depth of aggregation with total

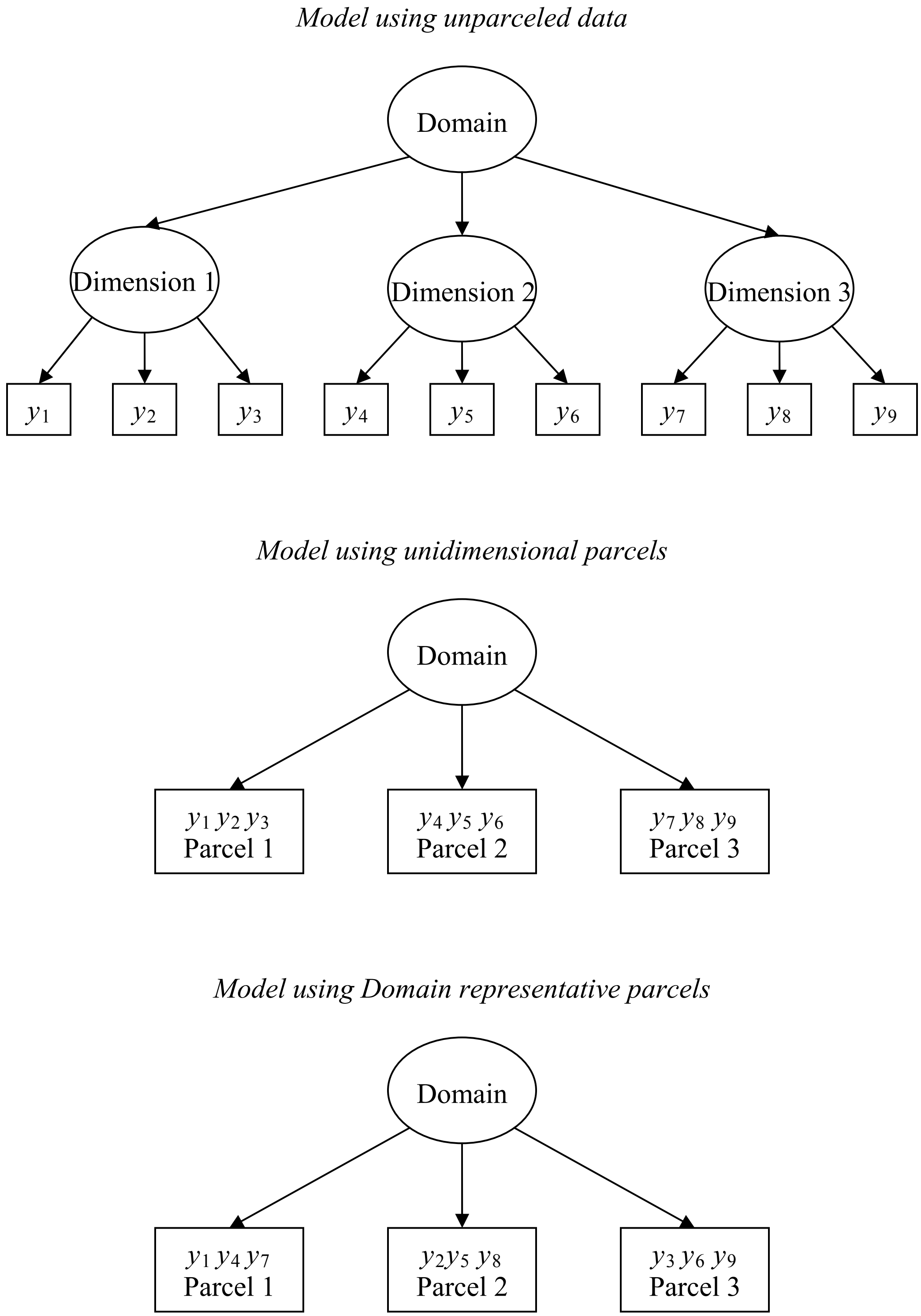


Fig. 2.3 Unidimentional Parceling versus Domain Representative Parceling

Note. The errors are are not shown for simplicity.

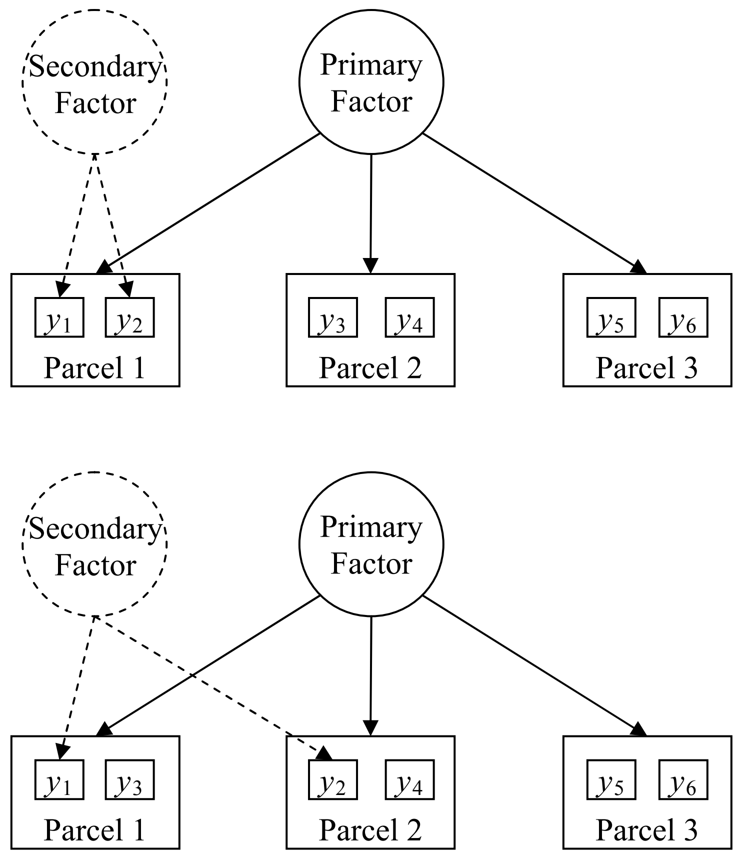


Fig. 2.4 Isolated Parceling (Up) versus Distributed Parceling (Down)

Note. The errors are not shown for simplicity.

disaggregation (i.e., items) as the lowest one, are partial disaggregation, partial aggregation, and total aggregation as shown in Figure 2.5 (Bagozzi & Edwards, 1998, p. 50). In partial disaggregation, parcels are formed within each component and then used as the indicators of the same component; in partial aggregation, all the items in each component are aggregated to form one parcel and then the parcels serve as indicators of the higher-order factors, each of which is defined by two or more components; in the total aggregation, a more general factor replaces these higher-order factors if they are highly correlated with each other. Among the three parceling methods, a second-order factor model is actually involved when conducting partial aggregation and total aggregation, with the components corresponding to the first-order factors and the higher-order factors or the general factor corresponding to the second-order factors. However, neither the partial aggregation model nor the total aggregation model is a true second-order factor model because the first-order factors (i.e. components) are collapsed by aggregating the items that measure these first-order factors.

Bagozzi and his colleague also compared and contrast the four different depth aggregations and concluded that the models are more parsimonious and simpler as the depth of aggregation increases; on the other hand, more information involved in the individual items could be eliminated as the depth of aggregation increases.

The three parceling methods proposed by Bagozzi and Edwards (1998) are not independent from the other parceling methods depicted previously. Unidimentional parceling, domain representative parceling, isolated parceling, or distributed parceling can be conducted with a different level of parceling. Therefore, unidimentional parceling or any of the others could be partial diaggregation, partial aggregation, or total aggregation.

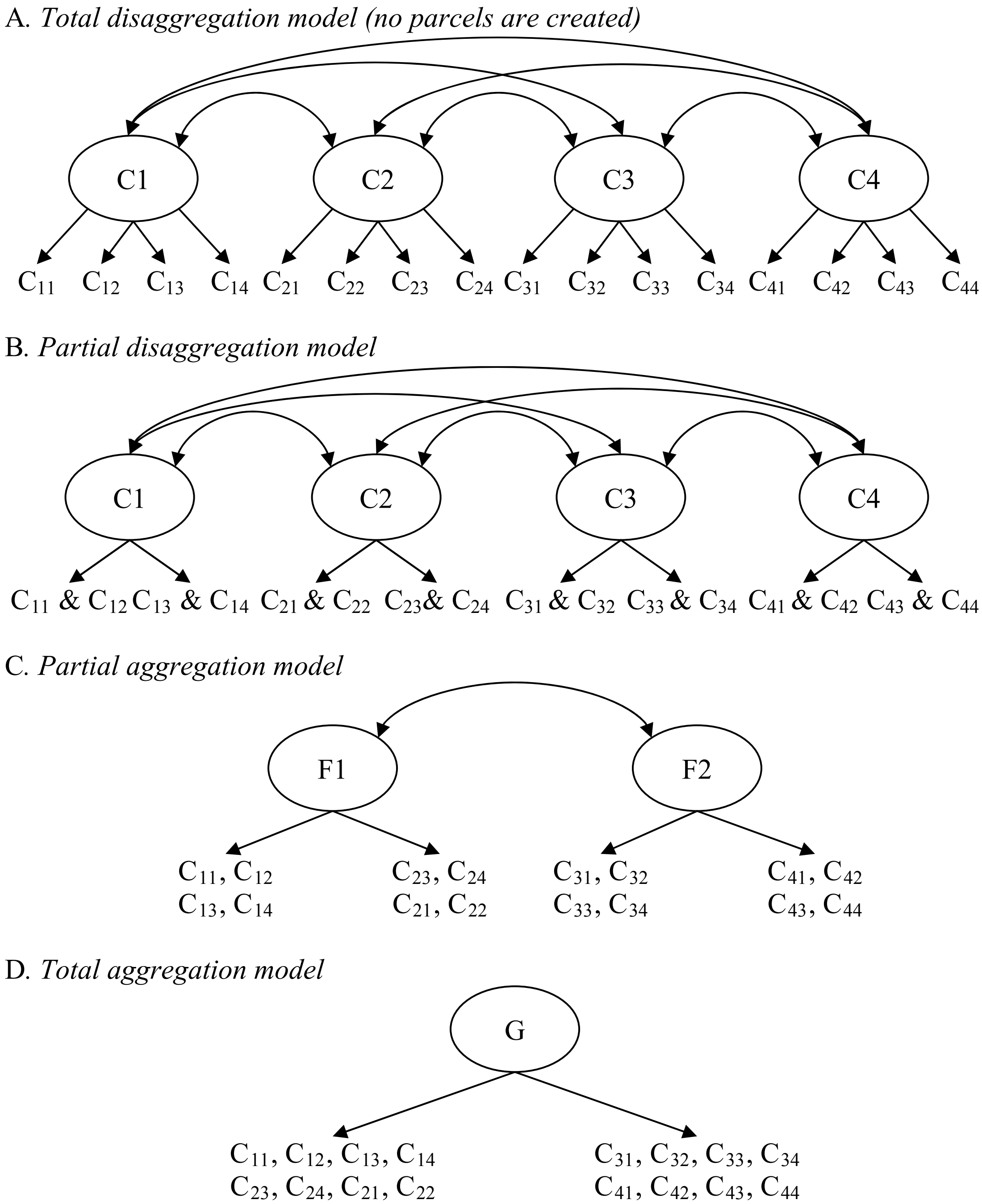


Fig. 2.5 Four Models Using Different Depth of Parceling

Note. C_i = component of scale; C_{ij} = measure j of component I; F_k = facet of scale; G = overal scale. Arrows connecting factors represent intercorrelations among factors.

Residual Adjustment for Parcels

The parcels that collapse the first-order factors (e.g., partial aggregation and total aggregation as defined above) are no longer error-free indicators of the second-order factor. They potentially consist of random error variance besides common variance. In addition, the parcels created by using unidimensional parceling strategy might also contain specific variance, which is originally reflected as the residual variance of the first-order factor. Therefore, the solution of the CFA model based on the parceled data, which is not error-free, might be contaminated. One way to accommodate the random error in this condition is to adjust the variance of the random error as (Sass & Smith, 2006, p. 575, Equation 2):

$$\sigma_c^2 - (\sigma_c^2 \times \alpha) \tag{16}$$

in which σ_c^2 represents the total variance of the parcel and α represents the Cronbach's α (Cronbach, 1951) calculated from the items that are parceled (for more details about measure error adjustment see Bedeian, Day, & Kelloway, 1997). However, it is impossible to identify the specific variance that is represented as the part of the residual variance in a multidimensional parcel (e.g., parcels created by using unidimensional strategy in a second-order CFA model) with the current techniques.

Other ways are available to calculate the reliability such as Spearman-Brown prediction formula and Angoff-Feldt coefficient (Angoff, 1953; Feldt, 1975; Feldt & Brennan, 1989). Feldt and Charter (2003) compared the three diffferent ways of calculating the reliability and concluded that the Spearman-Brown formula might be valid to estimate the reliability of total scores if the parts contributed equally to the total score and the error variances are constant (i.e., the parts are classically parallel). The equation for the Spearman-Brown formula is (Cohen, Cohen, Teresi, Marchi, & Velez, 1990, p. 189, Equation 1):

$$r_{XcXc} = \frac{k\bar{r}_{ij}}{1+(k-1)\bar{r}_{ij}} \tag{17}$$

in which *k* is the number of the parts, and *r* is the mean correlation. If the parts contributed equally to the total score but the error variances varied across the parts, Cronbach's α is appropriate; else if the two parts didn't contribute equally to the total score and the error variances varied across the parts (i.e., the parts are congeneric), the Angoff-Feldt coefficient should be calculated.

The Cons of Parceling

The advantages of parceling to enhancing reliability, normality, continuity and finally model fit are controversial. The area that is of greatest concern involves the dimensionality of the construct in which the items are parceled. When items that have multidimensionality are parceled (e.g., distributed parceling), the appropriate model fit could be obtained even if the latent secondary-factors are not specified (see Hall et al. 1999; Bandalos 1997; Bandalos 2002). Bandalos (2002) explained that when distributing the items that are loaded on a secondary factor as well as the common factor to each parcel (i.e., distributed parceling), the variance from the secondary factor is shared by all the parcels instead of being unique to any of them. Then the influence of the secondary factor will be absorbed by the common factor, resulting in spurious good model fit and biased coefficient estimates. Thus, the model misspecification (i.e., unmodeled secondary factor), which may be detected by using unparceled data, is masked by the good fit indices. Therefore, Bandalos (Bandalos & Finney, 2001; Bandalos 2002) and others stated that parceling strategy could be considered only if unidimentionality is clearly established or the secondary factors that might be masked by applying parceling are neglectable.

In addition, Bandalos and Finney (2001) argued that (1) when using parceling to mitigate the problem of nonnormality resulted from coarsely measured categorical data, the factor structure among the items might be distorted and the parameter estimates might be biased, although parceling does improve the normality of the original items; (2) the low reliability of the scales could be masked by using higher-reliability parcels rather than the original items. Corresponding to the two deficiencies of parceling respectively, Bandalos and the colleague (2001) further suggested that (1) research should include as many scale points as possible before resorting to parceling the coarsely categorical data; and (2) parceling should never be used when the purpose of the study is to develop a scaled instrument because the low communalities (or reliabilities) could serve effectively as an indicator of problems with the items.

Little et al. (2002) commented that the other deficiency of parceling is that although parceling has advantages in statistical aspects, it ruins the meaning of the unstandardized parameter estimates based on the unparceled data. Therefore, researchers should use a parceling strategy with great caution based on applied grounds, particularly in the behavioral and social sciences, in which many scales have established norms based on their means and standard deviations.

CHAPTER III

INVESTIGATIONS IN PARCELING IN SECOND-ORDER CFA MODELS

The present investigation consisted of two parts, (i.e., Study 1 and 2). Study 1 was motivated by the goal of examining the difference of Type I error rates as well as power produced by MCFA and MIMIC in a second-order CFA model. The other factors that were considered included group sizes, magnitude of the factor loadings (i.e., communality), magnitude of the error variance of the observed variables as well as the specific variance of the first-order factors, and the magnitude of the second-order factor mean difference between groups given the fact that these factors might influence the Type I error rate and power for detecting group differences.

Study 2 expanded Study 1 with the goal of investigating how data parceling affected the results achieved in Study 1. Two parceling methods (i.e., unidimentional parceling and domain representative parceling) were evaluated with respect to all the factors that were considered to potentially influence the Type I error rate and power in Study 1. In addition the models based on domain representative parcels were compared with and without measurement error adjustment. The reason that the model with domain representative parcels intead of unidimentional parcels was selected to make the error variance adjusted was that the residual variance of the domain representative parcel was not mixed with the specific vairance, which could not be identified, whereas that of the unidimensional parcel was. The results of the two studies would hopefully provide practical recommendations to researchers in the relative areas.

CHAPTER IV

STUDY 1: TYPE I ERROR RATE AND POWER UNDER MCFA AND MIMIC SECOND-ORDER CFA MODELS

As presented before, the goal of Study 1 was to examine the difference of Type I error rates as well as power produced by MCFA and MIMIC in a second-order CFA model. Unparceled data were used in this study. The methods that were used in Study 1 were introduced first, followed by the results and discussion.

METHODS

Model design and data simulation were introduced in the methods section. Methods section also included the calculation of Type I error rate and power, and how convergence problem was dealt with, which were also applicable for Study 2.

Design of Monte Carlo Simulation

In Study 1 a model with single second-order factor, four first-order factors, and three observed variables for each first-order factor was created for each of the two groups (Figure 2.1). Mplus 5.1 (Muthen & Muthen, 2007) was used to conduct the Monte Carlo study. The mean of the second-order factor in the first group (Group A) was set as 0 whereas the mean value in Group B was set as 0, .2, .3, .4, and .5 to represent the increasing discrepancy of the factor means between the two groups (i.e., $\Delta\kappa$). Also, .2 and .5 represent small and medium effect size respectively (Cohen, 1988).

Indeed, the magnitude of effect size in social sciences generally falls between .20 (small) and .50 (medium). For simplicity, the following steps were performed: The latent intercepts for all observed variables and the first-order factors were set to be 0 in both groups. All the factor loadings including those of the second-order factor and the first-order factors were set to be either .40 (low) or .80 (high) over all conditions, which were as same as the lowest and highest loading respectively in Hancock et al.'s (2000) study. Finally the variance of the second-order factor was set to be 1.0 in both groups. Thus, the equation for the variance of the observed variables:

$$V(y_{ij}) = (\lambda_{ij}\gamma_i)^2 \times V(\xi) + \lambda_{ij}{}^2 \times V(\zeta_i) + V(e_{ij}) \tag{18}$$

in which $V(y_{ij})$ is the variance of the observed variable, $V(\xi)$ is the variance of the second-order factor, $V(\zeta)$ is the specific variance, $V(e_{ij})$ is the variance of the random error, γ_i is the loading of the first-order factor on the second-order factor, and λ_{ij} is the loading of the observed variable on the first-order factor (i = 1 to 4, j = 1 to 3), can be reduced to:

$$V(y_{ij}) = .03 + .16 \times V(\zeta_i) + V(e_{ij}) \tag{19}$$

and

$$V(y_{ij}) = .41 + .64 \times V(\zeta_i) + V(e_{ij}) \tag{20}$$

for factor loadings .40 and .80, respectively.

The different combinations of group size and residual variance resulted in 12 different simulation scenarios as presented in Tables 4.1 and 4.2 for factor loading .8 and .4, respectively. The total sample size was set at 800 with the ratio of 1:1, 1:3, and 3:1, with corresponding group sizes of 400:400, 200:600, and 600:200. This sample size yielded a ratio of sample size to parameter approximately 11 for the MCFA model, and 27 for the MIMIC model, which were both much higher than the ratio 5:1 recommended by Bentler and Chou (1987).

Tables 4.1 and 4.2 also show the specific variances and the random error variances as well as the generalized variance for each group and the condition for each scenario. In Scenario 1-3 the residual variances of the observed variables in Group A were set three times as large as those in Group B, whereas the other parameters were constant across the two groups except the variances of the observed variables. Note that the residual variances of the first order factor (i.e., .36 or .84) were kept constant between groups and calculated with the assumption of the unity variance (i.e., 1.0) of the first-order factors. For Scenarios 4-6, only the specific variances were varied between groups and the random error variances were constrained as same as the specific variances in Scenario 1-3. In Scenario 7- 12, the residual variances of the observed variables and the first-order factors were both varied between groups. The difference is that the two types of residual variances were constant within the group in Scenario 7-9,

Table 4.1 Twelve Simulation Scenarios (Factor Loading = .8 for Both the Second Order-Order Factor and the First-Order Factors)

Scenario	Group size	Residual variances of the first order factors[4] (Ψ)	Residual variances of the observed	Generalized variance	Condition
1	400	[.36, .36, .36, .36]	[.10, .10, …, .10]	2.86E-8	Balanced
	400	[.36, .36, .36, .36]	[.30, .30, …, .30]	3.76E-4	
2	200	[.36, .36, .36, .36]	[.10, .10, …, .10]	2.86E-8	Positive
	600	[.36, .36, .36, .36]	[.30, .30, …, .30]	3.76E-4	
3	600	[.36, .36, .36, .36]	[.10, .10, …, .10]	2.86E-8	Negative
	200	[.36, .36, .36, .36]	[.30, .30, …, .30]	3.76E-4	
4	400	[.10, .10, .10, .10]	[..36, .36, …, .36]	2.60E-4	Balanced
	400	[.30, .30, .30, .30]	[..36, .36, …, .36]	1.36E-3	
5	200	[.10, .10, .10, .10]	[..36, .36, …, .36]	2.60E-4	Positive
	600	[.30, .30, .30, .30]	[..36, .36, …, .36]	1.36E-3	
6	600	[.10, .10, .10, .10]	[..36, .36, …, .36]	2.60E-4	Negative
	200	[.30, .30, .30, .30]	[..36, .36, …, .36]	1.36E-3	
7	400	[.10, .10, .10, .10]	[.10, .10, …, .10]	1.30E-9	Balanced
	400	[.30, .30, .30, .30]	[.30, .30, …, .30]	2.56E-4	
8	200	[.10, .10, .10, .10]	[.10, .10, …, .10]	1.30E-9	Positive
	600	[.30, .30, .30, .30]	[.30, .30, …, .30]	2.56E-4	
9	600	[.10, .10, .10, .10]	[.10, .10, …, .10]	1.30E-9	Negative
	200	[.30, .30, .30, .30]	[.30, .30, …, .30]	2.56E-4	
10	400	[.30, .30, .30, .30]	[.10, .10, …, .10]	1.73E-08	Balanced
	400	[.10, .10, .10, .10]	[.30, .30, …, .30]	4.23E-05	
11	200	[.30, .30, .30, .30]	[.10, .10, …, .10]	1.73E-08	Positive
	600	[.10, .10, .10, .10]	[.30, .30, …, .30]	4.23E-05	
12	600	[.30, .30, .30, .30]	[.10, .10, …, .10]	1.73E-08	Negative
	200	[.10, .10, .10, .10]	[.30, .30, …, .30]	4.23E-05	

[4] All these numbers are the diagonal elements of the corresponding residual variance-covariance matrix.

Table 4.2 Twelve Simulation Scenarios (Factor Loading = .4 for Both the Second Order-Order Factor and the First-Order Factors)

Scenario	Group size	Residual variances of the first order factors[5] (Ψ)	Residual variances of the observed variables[4] (Θ)	Generalized variance	Condition
1	400	[.84, .84, .84, .84]	[.10, .10, …, .10]	1.03E-09	Balanced
	400	[.84, .84, .84, .84]	[.30, .30, …, .30]	2.30E-05	
2	200	[.84, .84, .84, .84]	[.10, .10, …, .10]	1.03E-09	Positive
	600	[.84, .84, .84, .84]	[.30, .30, …, .30]	2.30E-05	
3	600	[.84, .84, .84, .84]	[.10, .10, …, .10]	1.03E-09	Negative
	200	[.84, .84, .84, .84]	[.30, .30, …, .30]	2.30E-05	
4	400	[.10, .10, .10, .10]	[.84, .84, .84, .84]	0.21	Balanced
	400	[.30, .30, .30, .30]	[.84, .84, .84, .84]	0.31	
5	200	[.10, .10, .10, .10]	[.84, .84, .84, .84]	0.21	Positive
	600	[.30, .30, .30, .30]	[.84, .84, .84, .84]	0.31	
6	600	[.10, .10, .10, .10]	[.84, .84, .84, .84]	0.21	Negative
	200	[.30, .30, .30, .30]	[.84, .84, .84, .84]	0.31	
7	400	[.10, .10, .10, .10]	[.10, .10, …, .10]	1.49E-11	Balanced
	400	[.30, .30, .30, .30]	[.30, .30, …, .30]	4.34E-06	
8	200	[.10, .10, .10, .10]	[.10, .10, …, .10]	1.49E-11	Positive
	600	[.30, .30, .30, .30]	[.30, .30, …, .30]	4.34E-06	
9	600	[.10, .10, .10, .10]	[.10, .10, …, .10]	1.49E-11	Negative
	200	[.30, .30, .30, .30]	[.30, .30, …, .30]	4.34E-06	
10	400	[.30, .30, .30, .30]	[.10, .10, …, .10]	8.08E-11	Balanced
	400	[.10, .10, .10, .10]	[.30, .30, …, .30]	1.83E-06	
11	200	[.30, .30, .30, .30]	[.10, .10, …, .10]	8.08E-11	Positive
	600	[.10, .10, .10, .10]	[.30, .30, …, .30]	1.83E-06	
12	600	[.30, .30, .30, .30]	[.10, .10, …, .10]	8.08E-11	Negative
	200	[.10, .10, .10, .10]	[.30, .30, …, .30]	1.83E-06	

[5] All these numbers are the diagonal elements of the corresponding residual variance-covariance matrix.

while they were varied within the group in Scenario 10-12. Given Equation 4, the variance-covariance matrix for each group in the 12 scenarios was obtained and then the generalized variance (i.e., the determinant of variance-covariance matrix) was calculated (Tables 4.1 and 4.2).

Two points were suggested by the generalized variances of the two groups in the 12 scenarios. Firstly, the two groups in each scenario showed extraordinary disparity in generalized variance. According to Kaplan's (1995) definition, Scenario 2, 5, 8 and 11 were in the positive condition (i.e., smaller samples have smaller generalized variances), and Scenario 3, 6, 9, and 12 were in the negative condition (i.e., smaller samples have larger generalized variances). Secondly, the relative discrepancy of the generalized variances between the two groups reached the minimum in Scenarios 4-6, in which the residual variances of the observed variables were identical across the two groups, and reached the maximum in Scenario 7-9, in which the random error variances and the specific variances were constant within the group. It indicates that the residual variance of the observed variables and the first-order factors had the primary and secondary influence respectively on the relative discrepancy of the generalized variance between the two groups.

Five hundred replications were generated for each scenario. Finally a total of 60,000 replications (i.e., $2 \times 12 \times 5 \times 500$) was used for the data analysis in Study 1.

Empirical Type I Error and Empirical Power

As stated previously, theoretically, the meaningful comparison of the means of the second-order factors can only be made when the two models have configural invariance, identical loadings of all the first- and second-order factors, and identical latent intercepts of the observed variables and the first-order factors (i.e., strong MI). Given the truth that only the residual variances were varied systematically with sample sizes across the 12 scenarios, these assumptions were completely fulfilled in the study. The empirical Type I error rate was calculated under the condition that the mean values of the second-order factors were identical between the groups. When conducting MCFA, the mean of the second-order factor in Group A was set as 0, whereas the mean in Group B was freely estimated. Within each scenario the number of the times in which the freely estimated factor mean was statistically significant ($p < .05$) was tallied. When conducting MIMIC analysis, firstly the data for the two groups were concatenated to form a single data file. Secondly, a dummy coded grouping variable representing the group membership (i.e., A and B) was added in the data file. The number of times when the path coefficient from the grouping variable to the second-order latent factor was statistically significant ($p < .05$) was counted. The percentage of significant findings was the empirical Type I error rate given that the mean values of the second-order factors were identical between the two groups.

Power analysis was conducted in the same way as Type I error rate was calculated but under the condition that the mean of the second-order factors was different from the other (i.e., .2, .3, .4, or .5 in this study). The empirical power was the percentage of significant findings ($p < .05$). Finally the general linear model with power or Type I error rate as the dependent variable and the factors that were considered to influence power or Type I error rate as the independent variables was run to examine the statistic significance of the factors. Given the fact that the numbers of the calculated Type I error rates and power are small (e.g., only 48 Type I error rates resulted from Study 1), the general linear model may not have high enough power to detect the significant factors that influence Type I error rate or power. Therefore, the general linear model was considered as a secondary way to analyze the data in this study.

Convergence Problem

MCFA or MIMIC analysis may yield a non-convergent result in either Study 1 or 2 (or both). In this event, the original dataset that was created in Study 1 as well as the datasets with the parcels which were calculated based on the dataset in Study 1 by using the two parceling strategies was deleted, a new dataset was generated and the parcels as well as their variances of the measurement error were recalculated until 500 convergent results were achieved in both Study 1 and 2 for the scenario.

RESULTS

Results are presented in the sequence of convergence rate, empirical Type I error rate (including the high factor loadings and the low factor loadings), empirical power with the high factor loadings, and empirical power with the low factor loadings. The results under the condition of the high factor loadings and the low factor loadings were compared.

Convergence Rate

Both of the MCFA and MIMIC models did not have a convergence problem when the factor loadings were high (i.e., 0.8). However, when the factor loadings were low (i.e., 0.4), convergence problem appeared in some scenarios. The first two columns in Table 4.3 showed the average convergence rates (i.e., the proportion of the convergent results among the 500 replications) across the five different values of $\Delta\kappa$ (i.e., 0, .2, .3, .4, and .5) within each scenario for each of MCFA and MIMIC approaches under the condition of the low factor loadings. These values indicate that MIMIC models generally have a higher convergence rate than MCFA models. This result is consistent with the widely accepted advantage of MIMIC models that MIMIC models require smaller sample size than MCFA models for the purpose of the convergence. Therefore, MIMIC models have higher convergence rates than MCFA models given equal sample size. The other information the first two columns in Table 4.3 provided is that Scenarios 4, 5, and 6 had lower convergence rates than the other scenarios, particularly for the MCFA approach. For example, the convergence rate in Scenario 5 by using MCFA approach was as low as .207, which means that among the 500 replications for each $\Delta\kappa$, only 20.7% of them (i.e., 103 replications) on average were convergent.

New datasets were created to make 500 convergent solutions for each scenario in which the convergent rate is lower than 100%. The values of empirical Type I error rate

Table 4.3 Average Convergence Rate across the Five Mean-Differences When Factor Loadings Were Low

	Unparceled		UD parceling		DR parceling		DR parceling with EA	
Scenario	MCFA	MIMIC	MCFA	MIMIC	MCFA	MIMIC	MCFA	MIMIC
1	0.986	1.000	0.991	1.000	1.000	1.000	1.000	1.000
2	0.970	1.000	0.968	1.000	1.000	1.000	1.000	1.000
3	0.870	1.000	0.944	1.000	1.000	1.000	1.000	1.000
4	0.324	0.934	0.881	0.992	0.962	1.000	0.998	1.000
5	0.207	0.942	0.762	0.992	0.869	1.000	0.718	1.000
6	0.221	0.919	0.834	0.992	0.953	1.000	0.706	1.000
7	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000
8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
9	0.954	1.000	0.990	1.000	1.000	1.000	1.000	1.000
10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
11	0.943	1.000	1.000	1.000	1.000	1.000	1.000	1.000
12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: UD = Unidimensional; DR = Domain representative; EA = Error adjustment.

and power resulted from the originally convergent models and the 500 convergent models for each of MCFA and MIMIC approaches were compared and it was found that two results were fairly close even if the convergence rate was very low.

Empirical Type I Error Rate

The empirical Type I error rates of MCFA and MIMIC for the 12 scenarios in each of the two sets of the factor loadings were shown in Table 4.4. The results of the

Table 4.4 Empirical Type I Error Rates of MCFA and MIMIC (Unparceled Data)

		Factor loading = 0.8			Factor loading = 0.4		
Scenario	Condition	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.
1	=	0.048	0.048	0.000	0.052	0.036	0.016
2	+	0.038	0.040	-0.002	0.064	<u>0.020</u>	0.044
3	-	0.060	0.042	0.018	0.049	0.038	0.011
4	=	0.032	0.056	-0.024	0.046	<u>0.018</u>	0.028
5	+	0.054	0.040	0.014	0.053	<u>0.009</u>	0.044
6	-	0.048	0.058	-0.010	0.029	<u>0.007</u>	0.022
7	=	0.054	0.056	-0.002	0.044	0.032	0.012
8	+	0.040	0.044	-0.004	0.046	<u>0.016</u>	0.030
9	-	0.058	0.062	-0.004	0.043	<u>0.094</u>	-0.051
10	=	0.052	0.050	0.002	0.058	0.046	0.012
11	+	0.064	0.064	0.000	0.052	0.046	0.006
12	-	0.068	0.050	0.018	0.051	<u>0.076</u>	-0.025

Note*:* A equal (=) sign indicates a balanced condition in which two groups have identical sample sizes; a plus (+) sign indicates a positive pairing of sample with with generalized variance; a negative (-) sign indicates a negative pairing. Diff. = difference of MIMIC from MCFA. Any error rate falling beyound Bradley's (1978) liberal criterion of [.5α, 1.5 α] (i.e., [.025, .075]) is underlined.

general linear model including the estimated effect of each factor as well as the standard error, the t-value and the *p*-value were presented in Table 4.5. It was shown in Table 4.5 that none of the factors were statistically significant for Type I error rate. That is, the difference between MCFA and MIMIC in terms of Type I error rate was not statistically significant, holding the other factors constant; the differences between positive and balanced condition, negative and balanced condition were not statistically significant, holding the other factors constant; and the factor loadings did not impact Type I error rate significantly. However, as stated before, the general linear model for Type I error rate was based on 48 calculated Type I error rates (i.e., 2 approaches × 12 scenarios (including 3 conditions) × 2 factor loadings). Therefore the sample size might not be large enough to conduct statistic analysis for detecting the factors that impact Type I error rate.

Table 4.5 Parameter Estimates of the General Linear Model with Type I Error Rate as the Dependent Variable

	Parameter	Effect	Std. Error	t	$p^{\dagger\dagger\dagger}$
Type I error rate	Intercept	.027	.008	3.387	.002
	Approach = MCFA	.006	.004	1.443	.156
	Approach = MIMIC†				
	Condition = negative	.007	.005	1.197	.238
	Condition = Positive	-.002	.005	-.433	.667
	Condition = balanced ††				
	Loadings	.021	.011	1.871	.068

Note: †MIMIC is the reference of the two approaches. ††Balanced condition is the reference of the three conditions. †††All the *p*-values are two-tailed.

More interesting is to examine the plots of Type I error rate in Figure 4.1. In general, when the factor loadings were high, Type I error rates of the two approaches were very close, and showed no obvious pattern as the conditions change (Figure 4.1). The discrepancies of the Type I error rate between the two approaches ranged from .000 to .024. None of the Type I error rates fell outside Bradley's (1978) liberal criterion of [.5α, 1.5 α] for empirical Type I error from simulation studies (i.e., [.025, .075] if we take α as .05). This indicates that both of the methods, MCFA and MIMIC, control Type I error at an acceptable level when the factor loadings are as high as .8, regardless of the condition the two groups are in. When the factor loadings decreased to .4, MCFA showed higher Type I error rates than MIMIC except in Scenario 9 and 12. However, MCFA appeared to maintain Type I error control within Bradley's literal criterion in all of the 12 scenarios, whereas 7 out of 12 Type I error rates of MIMIC (i.e., .020 in Scenario 2, .018 in Scenario 4, .009 in Scenario 5, .007 in Scenario 6, .016 in Scenario 8, .094 in Scenario 9, and .076 in Scenario 12) were out of Bradley's literal criterion. These out-of-criterion Type I error rates indicate that MIMIC tends to be conservative in the positive condition, and liberal in the negative condition (except in Scenario 6), but maintain Type I error rate well in the balanced condition (except in Scenario 4). This behavior is opposite to that in Hancock et al.'s (2000) study, in which MIMIC tended to be conservative in the negative condition and liberal in the positive condition. However, it is consistent with the typically observed for T^2 (e.g., Hakstian, Roed, & Lind, 1979; Holloway, & Dunn, 1967).

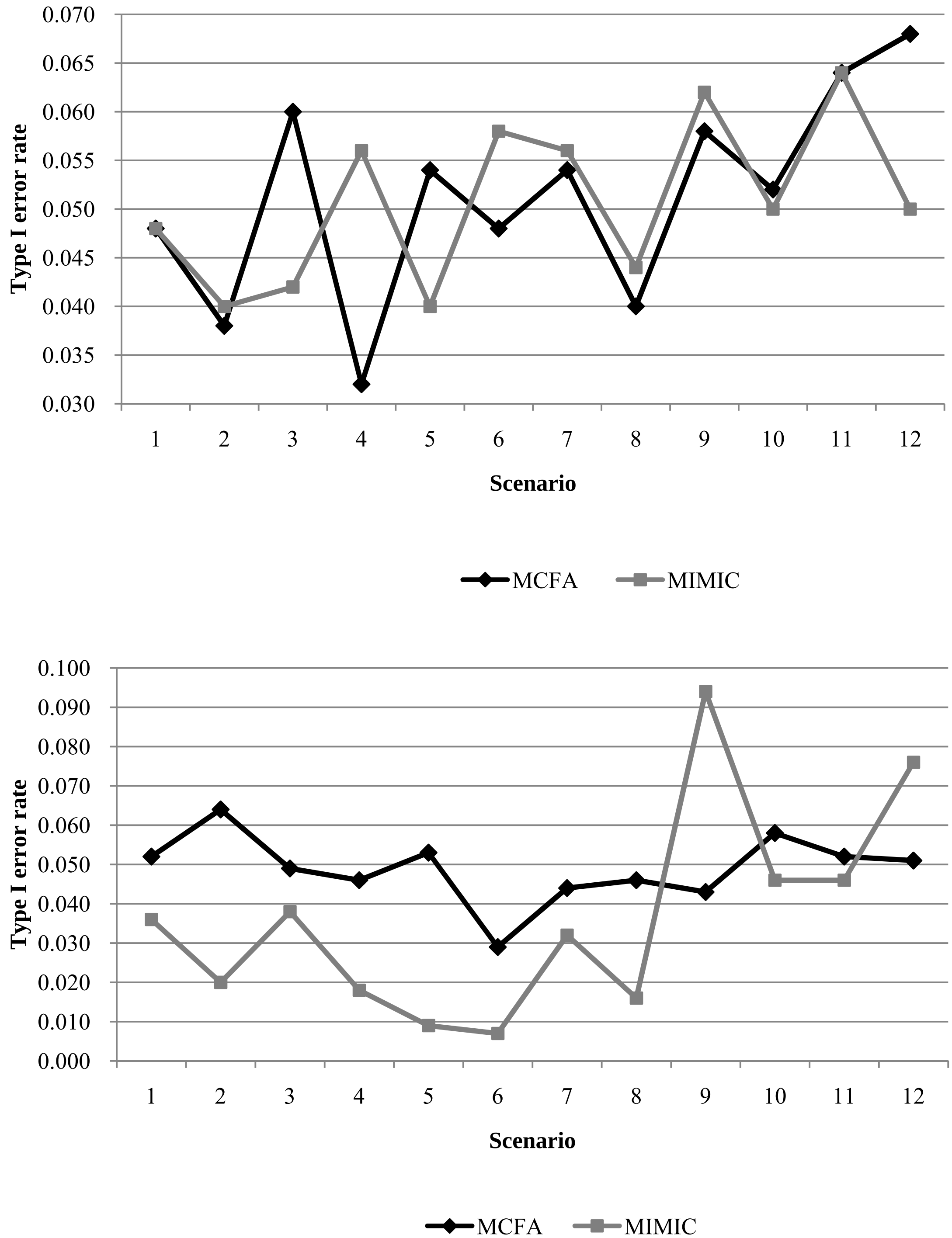


Fig. 4.1 Plots of Empirical Type I Error Rates of MCFA and MIMIC Approaches with the Unparceled Data

Note. Upper: factor loading = .8; Down: factor loading = .4.

Empirical Power with the High Factor Loadings

The empirical power in the 12 scenarios for the high factor loadings were presented in the upper part of Table 4.6 and were also plotted as shown in Figure 4.2. In this figure, the power of the two approaches at different values of $\Delta\kappa$ was plotted together (left) and separately (right). As expected, the statistical power increased as $\Delta\kappa$ increased from small to median effect size. The results of the general linear model for power presented in Table 4.7 also showed that $\Delta\kappa$ was a significant factor for power ($p < .001$) even with the small sample size (i.e., 2 approaches × 12 scenarios (including 3 conditions) × 2 factor loadings × 4 values of $\Delta\kappa$ = 192). Meanwhile the discrepancy between MCFA and MIMIC decreased as $\Delta\kappa$ increased. The power of both MCFA and MIMIC reached maximum 1.000 when $\Delta\kappa$ was .5 (i.e., median effect size). When $\Delta\kappa$ was less than median effect size, the two approaches displayed different patterns across the 12 scenarios. Firstly, in the balanced condition (i.e., Scenario 1, 4, 7, and 10), the power of MCFA was always larger than that of MIMIC regardless of the magnitude of the residual variance in each level. This advantage of MCFA disappeared when $\Delta\kappa$ increased to .4. In this situation, the power of both of the two methods reached 1.000. Secondly, in the positive condition, MCFA showed higher power than MIMIC except in Scenario 11, in which MIMIC showed higher power with superiority of .002 at $\Delta\kappa = .4$ to .032 at $\Delta\kappa = .2$. The two methods exchanged their superiority in terms of power (i.e., MIMIC displayed higher power than MCFA) when it came to the negative condition except the one in Scenario 12. In Scenario 12, MCFA was superior to MIMIC by .006 at $\Delta\kappa = .4$ to .312 at $\Delta\kappa = .2$. Note

Table 4.6 Empirical Power of MCFA and MIMIC (Unparceled Data)

			Δκ = .2			Δκ = .3			Δκ = .4			Δκ = .5		
	Scenario	Condition	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.
Factor loading = 0.8	1	=	0.828	0.744	0.084	0.990	0.978	0.012	1.000	1.000	0.000	1.000	1.000	0.000
	2	+	0.920	0.624	0.296	1.000	0.906	0.094	1.000	0.996	0.004	1.000	1.000	0.000
	3	-	0.556	0.626	-0.070	0.886	0.922	-0.036	0.974	0.992	-0.018	1.000	1.000	0.000
	4	=	0.804	0.750	0.054	0.990	0.978	0.012	1.000	1.000	0.000	1.000	1.000	0.000
	5	+	0.906	0.626	0.280	1.000	0.918	0.082	1.000	0.998	0.002	1.000	1.000	0.000
	6	-	0.552	0.640	-0.088	0.886	0.932	-0.046	0.980	0.992	-0.012	1.000	1.000	0.000
	7	=	0.838	0.754	0.084	0.994	0.982	0.012	1.000	1.000	0.000	1.000	1.000	0.000
	8	+	0.930	0.638	0.292	1.000	0.922	0.078	1.000	0.998	0.002	1.000	1.000	0.000
	9	-	0.584	0.654	-0.070	0.908	0.932	-0.024	0.988	0.994	-0.006	1.000	1.000	0.000
	10	=	0.868	0.770	0.098	1.000	0.986	0.014	1.000	1.000	0.000	1.000	1.000	0.000
	11	+	0.620	0.652	-0.032	0.920	0.938	-0.018	0.992	0.994	-0.002	1.000	1.000	0.000
	12	-	0.954	0.642	0.312	1.000	0.932	0.068	1.000	0.994	0.006	1.000	1.000	0.000

Table 4.6 (Continued)

	Scenario	Condition	Δκ= .2			Δκ= .3			Δκ= .4			Δκ= .5		
			MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.
Factor loading = 0.4	1	=	0.214	0.322	-0.108	0.433	0.620	-0.187	0.633	0.852	-0.219	0.832	0.966	-0.134
	2	+	0.316	0.228	0.088	0.511	0.476	0.035	0.763	0.702	0.061	0.896	0.902	-0.006
	3	-	0.144	0.266	-0.122	0.253	0.520	-0.267	0.363	0.780	-0.417	0.544	0.920	-0.376
	4	=	0.105	0.078	0.027	0.218	0.184	0.034	0.269	0.391	-0.122	0.489	0.547	-0.058
	5	+	0.158	0.059	0.099	0.333	0.131	0.202	0.500	0.254	0.246	0.520	0.433	0.087
	6	-	0.117	0.068	0.049	0.140	0.139	0.001	0.205	0.245	-0.040	0.238	0.441	-0.203
	7	=	0.267	0.496	-0.229	0.555	0.830	-0.275	0.770	0.976	-0.206	0.938	0.998	-0.060
	8	+	0.386	0.358	0.028	0.644	0.674	-0.030	0.860	0.902	-0.042	0.980	0.998	-0.018
	9	-	0.152	0.438	-0.286	0.309	0.788	-0.479	0.437	0.934	-0.497	0.689	0.990	-0.301
	10	=	0.408	0.490	-0.082	0.726	0.826	-0.100	0.922	0.974	-0.052	0.986	0.998	-0.012
	11	+	0.258	0.330	-0.072	0.486	0.688	-0.202	0.758	0.920	-0.162	0.884	0.994	-0.110
	12	-	0.514	0.432	0.082	0.830	0.732	0.098	0.981	0.932	0.049	0.996	0.984	0.012

Note. A equal (=) sign indicates a balanced condition in which two groups have identical sample sizes; a plus (+) sign indicates a positive pairing of sample with with generalized variance; a negative (-) sign indicates a negative pairing. Diff. = Difference of MIMIC from MCFA.

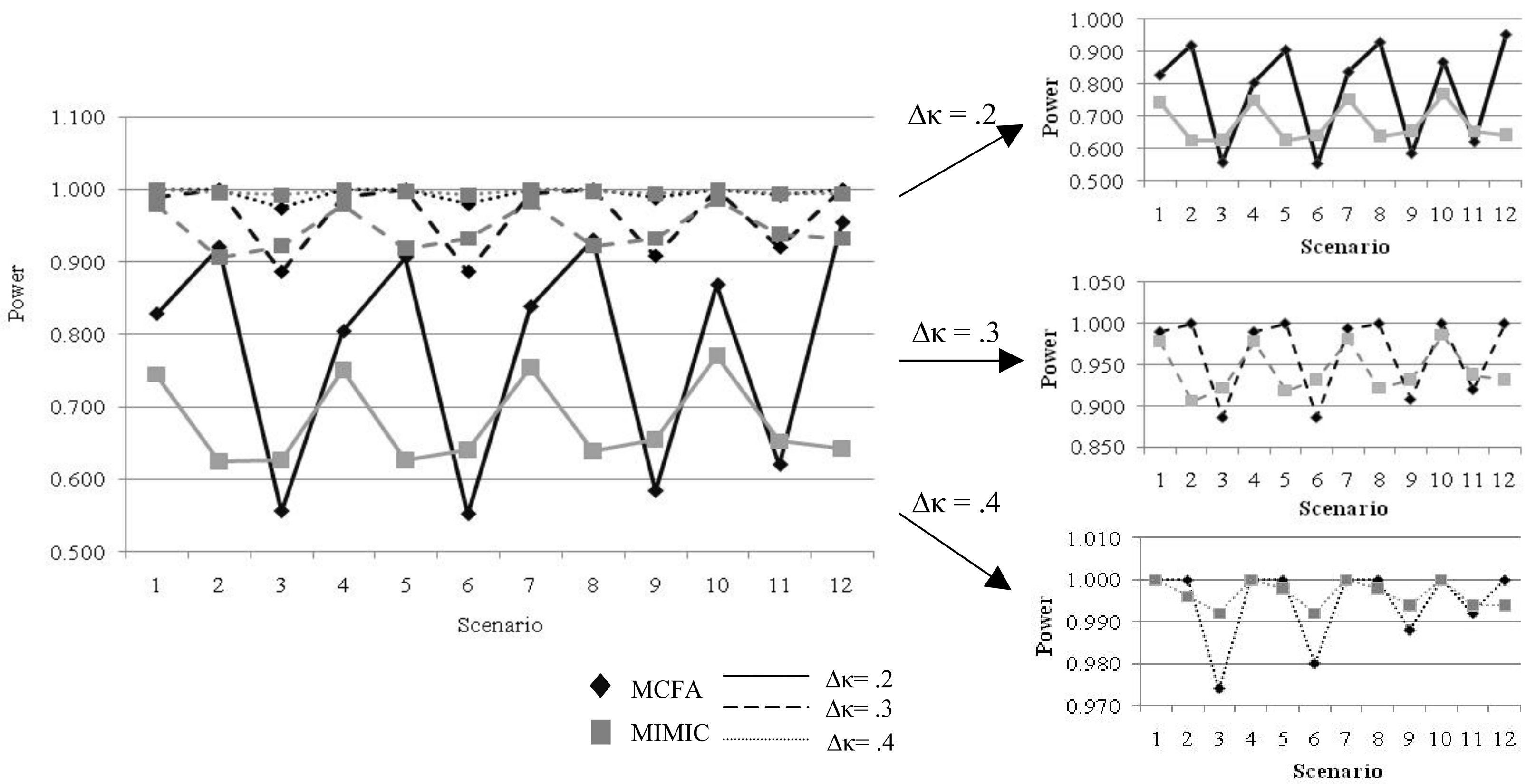


Fig. 4.2 Plots of Empirical Power of MCFA and MIMIC Approaches with the Unparceled Data When the Factor Loadings Were High

Note: The plots for both of the two approaches fall on the straight line at 1.000 when Δκ is .5 (not shown on this figure). The three small figures on the right are the decomposed figures of the one on the left based on the values of Δκ.

Table 4.7 Parameter Estimates of the General Linear Model with Power as the Dependent Variable

	Parameter	Effect	Std. Error	t	$p^{\dagger\dagger\dagger}$
Power	Intercept	-.222	.060	-3.686	.000
	Approach = MCFA	-.038	.025	-1.492	.137
	Approach = MIMIC†				
	Condition = negative	-.071	.031	-2.287	.023
	Condition = Positive	-.030	.031	-.967	.335
	Condition = balanced ††				
	Loadings	.880	.063	13.981	.000
	Δκ	1.313	.113	11.667	.000

Note: †MIMIC is the reference of the two approaches. ††Balanced condition is the reference of the three conditions. †††All the p-values are two-tailed.

that the superiority of the two approaches in the specific scenarios decreased when Δκ increased and completely vanished when the mean difference was .5. Thirdly, MIMIC was more powerful in the balanced condition than in the other conditions. Power decreased moving from the balanced condition to the positive condition and then increased slightly moving from the positive condition to the negative condition except from Scenario 11 to 12. In fact MIMIC's power in the positive condition and negative was very close. On the other hand, MCFA displayed higher power in the positive condition than in the others except in Scenario 11, and then experienced a dramatic drop moving from the positive condition to the negative condition except in Scenario 11 and 12. In fact, the patterns of power for both of the two approaches in the positive condition in Scenario 11 and the negative condition in Scenario 12 would be consistent with those displayed in the other positive and negative conditions if Scenario 11 and 12 were switched.

Empirical Power with the Low Factor Loadings

The results of power analysis with low factor loadings were presented in the lower part of Table 4.6 and the plots were shown in Figure 4.3. Same as Figure 4.2, the empirical power of the two approaches with various values of $\Delta\kappa$ was plotted together and individually for better view in Figure 4.3.

Again, power of both of the two approaches was greater for larger values of $\Delta\kappa$. For example, when $\Delta\kappa$ was 0.2, the highest power for MCFA was .514 (in Scenario 12) and .496 for MIMIC (in Scenario 7). When $\Delta\kappa$ increased to 0.5, the highest power for MCFA was .996 (in Scenario 12) and .998 for MIMIC (in Scenario 7, 8, and 10). Also, as expected, empirical power obtained under the condition of the low factor loadings was lower than that in the high-factor-loading setting by using either of the two approaches in any scenario. This trend was supported by the general linear model (see Table 4.7), in which the factor loadings had strong effect on power ($p < .001$). However, the decreasing discrepancy between the two approaches as $\Delta\kappa$ increased, which was shown with the high factor loadings, was not observed when the factor loadings were as low as .4.

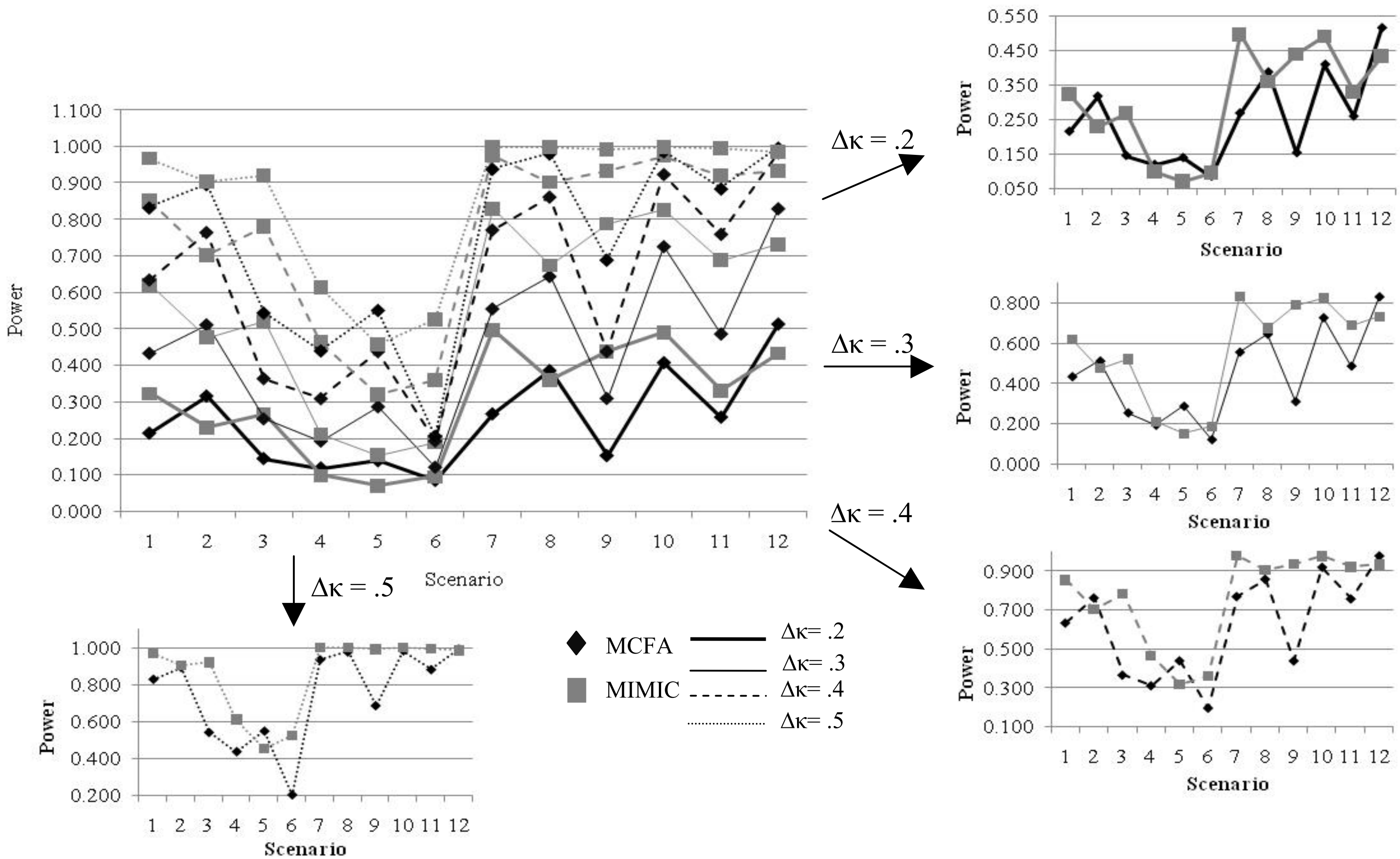


Fig. 4.3 Plots of Empirical Power of MCFA and MIMIC Approaches with the Unparceled Data When the Factor Loadings Were Low

Note: The four small figures are the decomposed figures of the left top one based on the values of Δκ.

The empirical power under the condition of the low factor loadings also showed some patterns across the 12 scenarios. Firstly, the plots for both of the approaches displayed a “U” shape across the 12 scenarios regardless of the values of $\Delta\kappa$. Scenarios 4, 5, and 6 were on the bottom of the “U”, which indicated that the power in the three scenarios was lower than the others. Note that the two approaches, especially MCFA, had the lower convergence rate in the three scenarios than the others. Secondly, similarly as the results when factor loadings were high, MCFA generally had higher power than MIMIC in the positive condition except in Scenario 11, and MIMIC generally had higher power than MCFA in the negative condition except in Scenario 12 at all of the four values of $\Delta\kappa$. However, contrary to the pattern when the factor loadings were high, MIMIC generally displayed higher power than MCFA in the balanced condition except in Scenario 4 at $\Delta\kappa = .2$, in which MCFA’s superiority was a mere .027. Thirdly, MIMIC was generally more powerful in the balanced condition than in the other conditions and MCFA was generally more powerful in the positive condition than in the other conditions except in Scenario 11. These two trends were as same as the ones observed when the factor loadings were high.

In fact, by comparing the plots under the conditions of the high factor loadings and low factor loadings, it was easy to find out that either of the two approaches experienced very similar power changes across the 12 scenarios between the two factor-loading settings. Under the condition of the low factor loadings, the plot of MCFA across the 12 scenarios seems “pushed” down toward the plot of MIMIC compared with their relative position in the plot under the condition of the high-factor loadings. As a result, the high superiority of MCFA in the positive condition under the condition of high factor loadings was generally decreased, the slight superiority of MIMIC in the negative condition was generally increased, and the original superiority of MCFA in the balanced condition turned into inferiority.

DISCUSSION

The results from Study 1 indicate that both of the two approaches control Type I error acceptably in all scenarios under the condition of high fact loadings. It means that the factors considered in the study, including the ratio of group sizes and the ratio of the residual variances of the two groups, have no significant impact on Type I error rate investigated by either of the two approaches if all the factor loadings are as high as .8. In this case, either of them could be conducted in terms of Type I error rate. However, that is not the situation when it comes to the case of low factor loadings. MCFA still maintains appropriate Type I error rates. But MIMIC is more likely to have a liberal Type I error in the negative condition (except in Scenario 6) and conservative Type I error in the positive condition. This behavior can be explained by the plot of the overall model displayed by Equation 10 with grouping variable X on the horizontal axis and the scores of the second-order factor on the vertical axis. Although the second-order factors in both groups are set to have unity variance for the population, it may not be true for the samples. The group that has larger generalized variance tends to produce larger variance of the second-order factor than the group that has smaller generalized variance. In the positive condition, the sample size ratio is positively paired with generalized variance (i.e., the larger sample has larger generalized variance and the smaller sample has smaller generalized variance). Therefore, the large group has the values of the second-order factor distributed more widely around the mean than the small group does. Because of the large sample size, the mean value of the second-order factor is unlikely to deviate from the true value (e.g., 0 in this study) too much. On the other hand, the mean value in the smaller group could not be biased too much from the true value either even if there is one or more outliers because the variance of the second-order factor in the small group is low. This will yield a relatively small slope of the regression line in the plot and hence a deflated Type I error rate. Conversely, if the second-order factor with larger variance has smaller sample size, it is verly likely to have the outliers which are on the leverage position to change the slope of the regression line from 0 and hence an inflated Type I error rate is resulted. When the two groups are in the balanced condition, the two groups will truly reflect the two infinite populations, thereby controlling the Type I error rate appropriately.

However, all the Type I error rates of MIMIC are lower than the bottom limit of Bradley's liberal criterion in Scenarios 4-6, which are in the balanced, positive, and negative condition, respectively. This might be related to the low convergence rate in the three scenarios (Table 4.3). Interestingly, although the convergence rates of MCFA are much lower than that of MIMIC, MCFA is not influenced by the low convergence rate in terms of Type I error rate.

Another question is, why does MIMIC control Type I error rate appropriately under the condition of the high factor loadings? The possible reason is that under the condition of the high factor loadings in this study, the relative difference of the generalized variances (and hence the variances of the second-order factors for the sample) between the two groups is not large enough to influence the Type I error rate when the group sizes are unbalanced.

Study 1 also offers the evidence that MIMIC approach and MCFA approach both have an advantage over each other under some specific situations in terms of power. Under the condition of the high factor loadings, both MCFA and MIMIC approaches reach maximum power 1.0 when the factor mean difference is larger than .4. That means either approach works perfectly in terms of power in any of the balanced condition, negative condition or positive condition if the group difference on the latent mean reaches medium effect size. A more interesting finding is the comparison between the two approaches when effect size lies between small and medium magnitude. An obvious pattern of MCFA and MIMIC is displayed in terms of power from Scenario 1 to 10. Generally MCFA is more powerful than MIMIC in the balanced and the positive condition while MIMIC is favored in the negative condition. However, the two approaches have an opposite pattern in the positive condition specified in Scenario 11 and the negative condition specified in Scenario 12. That is, MCFA displays higher power than MIMIC in the negative condition in Scenario 12, and MIMIC displays higher power than MCFA in the positive condition in Scenario 11. By comparing the specific variances in the two scenarios with the others, it indicates that the specific variances vary across the two groups in a way opposite to that of the random error variances. Note that random error variances are crucial for the generalized variance, which then determines the condition combined with the relative group size. It implies that instead of the condition which is defined by the combination of sample size and generalized variance, the specific variances combined with the sample size may have a more significant impact on the two approaches in terms of power when the specific variances are negatively paired with the generalized variances across the two groups. When the large sample size is associated with the small specific variances, whether it is in the positive or the negative condition, MIMIC is favored. When the larger sample size is associated with the large specific variances whether it is in the positive or the negative condition, MCFA is favored. However, by comparing the power of the two approaches in the balanced condition specified in Scenario 10 in which the specific variances are negatively paired with the random error variances across the two groups and the balanced condition specified in the other scenarios (i.e., Scenario 1, 4, and 7), it is shown that the specific variances have little influence on the relative power magnitudes of the two approaches in the balanced design. In addition, in Scenario 2 and 3 where the specific variances are invariant across the two groups, the relative power of the two approaches is still scarcely impacted by the specific variances. It indicates that the specific variances influence the superiority of one of the two approaches in terms of power only if the magnitudes of specific variances are negatively paired with the specific variances across two groups with disparate group sizes.

Some of the trends the two approaches display under the condition of the high factor loadings still hold when the factor loadings decrease to .4. For example, MCFA is still favored in the positive condition except in Scenario 11 and MIMIC is still favored under the negative condition except in Scenario 12. However, MCFA loses the superiority in the balanced condition it has under the condition of the high factor loadings. In fact, power of both of the two approaches decreases a lot from the condition of high factor loadings to the low factor loadings, but power of MCFA decreases more than that of MIMIC in any of the 12 scenarios. This change is displayed in Figure 4.2 as that MCFA seems being "pushed" downward to those of MIMIC. As a result, MIMIC exhibites the superiority in the balanced condition and increases its superiority in the negative condition. However, MIMIC's superiority in the negative condition under the condition of the low factor loadings sacrifices Type I error control whereas MCFA's superiority in the positive condition under the same condition doesn't.

Thus, it is hard to give a simple answer to the question "which approach should be used in the second-order CFA model, MCFA or MIMIC?" The answer depends on the condition, the magnitude of the factor loadings, and the magnitude of the specific variances. When the factor loadings are as high as .8, MCFA should be favored in the balanced and the positive condition and MIMIC should be favored in the negative condition. However, if the specific variances are negatively paired with the generalized variances across the two groups that have disparate group sizes, MCFA should be favored in the negative condition, while MIMIC should be favored in the positive condition. When the factor loadings are as low as .4 (which is more observed in the empirical studies in educational and social sciences), MIMIC is superior in the balanced condition, and MCFA is favored in both the positive and the negative condition given the fact that it controls Type I error rate better than MIMIC does.

I did not investigate in this study the value of factor loading between .4 and .8 that MIMIC starts to lose control of Type I error rate. Therefore, it is possible that MIMIC's superiority in the negative condition sacrifices Type I error control at any factor loadings lower than .8. In fact MIMIC's superiority in the negative condition as well as the positive condition when the specific variances were negatively paired with the generalized variances when the factor loadings were as high as .8 never exceeded .09. Therefore, it would be safe to use MCFA instead of MIMIC in the negative condition at any factor loadings less than .8.

The extremely low Type I error rates for MIMIC in Scenarios 4-6 as well as the "U" shape in the power plots displayed by both of the two approaches with Scenarios 4-6 at the bottom of the "U" when the factor loadings are low indicate that low convergence rate, which might be caused by the high random error variances in this study, could negatively impact Type I error rate for MIMIC and power for both of MCFA and MIMIC. However, MCFA is robust to the low convergence rate with respect to Type I error rate.

CHAPTER V

STUDY 2: EFFECTS OF PARCELING ON GROUP COMPARISON BY USING MCFA AND MIMIC

Study 2 introduced in this chapter was the extension of Study 1 in Chapter IV. Instead of the item-level data, unidimensional parcels and domain representative parcels were used in this study to investigate the difference between MCFA and MIMIC in terms of Type I error rate and power to detect group difference. Same as Chapter IV, this chapter included the sections methods, results, and discussion.

METHODS

No new data were generated in Study 2 since Study 1 data could be applied. This section introduces in detail how the data were parceled by using unidimensional parceling and domain representative parceling and how residual variances were adjusted for the domain representative parcels. The way that Type I error rate and power were calculated and how the convergence problem was dealt with were introduced in Chapter IV.

Parceling

SAS 9.1.3 was used to generate parcels for the data simulated in Study 1 by using a unidimensional parceling strategy and a domain representative parceling strategy respectively. In Figure 2.3, the observed variables were averaged within each first-order construct (i.e., Y_{11}, Y_{12}, and Y_{13}, Y_{21}, Y_{22}, and Y_{23}, and so on) when conducting unidimensional parceling; whereas the observed variables from each first-order construct were averaged to create a parcel (i.e., Y_{11}, Y_{21}, Y_{31}, and Y_{41}, Y_{12}, Y_{22}, Y_{32}, and Y_{42}, and so on) when conducting domain representative parceling. Because the model used to generate the data in Study 1 (Figure 2.1) had four first-order factors, each of which had three observed variables, four parcels were consequently obtained as the indicators for the second-order factor by using unidimensional parceling and three parcels as the indicators for the second-order factor by using domain representative parceling. Both of the two parceling strategies collapsed the first-order factors and the new models based on parcels were actually first-order factor models with the original second-order factor as the first-order factor.

Error Adjustment

The models based on domain representative parcels were run with and without error variance adjustment in Study 2. With error variance adjustment, the error variance was calculated by using Equation 16 for each of the three domain representative parcels in the model. Because all the factor loadings and the residual variances were constant among the observed variables that were parceled, they were classically parallel and the Spearman-Brown formula (Equation 17) was used to calculate the reliability in Equation 16. Note that the reliability coefficient calculated by using this equation might be negative. In this case, it was adjusted to 0 as suggested by Wiersma and Jurs (1990). Then the residual variances of the parcels were constrained to their calculated error variances for the 60,000 datasets when the models were run. These steps were accomplished by using SAS 9.1.3 and Mplus 5.1 (Muthen & Muthen, 2007).

RESULTS

As in Chapter IV, results were reported in the sequence of convergence rate, empirical Type I error rate (including the high factor loadings and the low factor loadings), empirical power with the high factor loadings, and empirical power with the low factor loadings. Instead of comparing MCFA and MIMIC in terms of Type I error rate and power, this section emphasized on comparing the results between Study 1 and 2. In addition, the results of domain representative parceling with error adjustment were presented, and then the results were compared with those of domain representative parceling without error adjustment.

Convergence Rate

As expected, the models using the parceled data did not have a convergence problem when the factor loadings were high. Under the condition of low factor loadings, the convergence problem appeared, especially in Scenarios 4–6 (Table 4.3). Some trends can be observed in Table 4.3. Firstly, the models using the parcels, including unidimensional parcels and domain representative parcels with or without error adjustment, generally showed fewer convergence problems than the models using the unparceled data in the 12 scenarios. Secondly, models using domain representative parcels generally had fewer convergence problems than the models using the unidimensional parcels. Thirdly, as with the models using unparceled data, MIMIC models using parcels never had lower convergence rates than MCFA models. These findings are consistent with the rule of parsimony in SEM that a simpler model with fewer estimated parameters has fewer convergence problems, with other things being equal.

Empirical Type I Error Rate

Empirical Type I error rates of MCFA and MIMIC with high and low factor loadings by using unidimensional and domain representative parceling strategies were shown in Table 5.1 and Figures 5.1 and 5.2. The results of the general linear model were shown in Table 5.2.

In general, the differences between the two approaches were very small regardless of the factor loadings and parceling strategies (p = .546) and most of the differences were lower than .01, especially in the balanced condition. As with the models using unparceled data, none of the models using unidimensional parcels or domain representative parcels had empirical Type I error rates outside Bradley's (1978) liberal criterion when the factor loadings were high.

However, neither of the two approaches, particularly MIMIC, maintained Type I error rates perfectly across the 12 scenarios by using the two parceling strategies when the factor loadings were as low as .4. For example, by using unidimensional parcels, MIMIC had five Type I error rates (i.e., .018 in Scenario 2, .023 in Scenario 5, .022 in Scenario 6, .018 in Scenario 8, and .098 in Scenario 9) out of Bradley's (1978) liberal

Table 5.1 Empirical Type I Error Rates of MCFA and MIMIC (Parceled Data)

			UD parceling			DR parceling		
	Scenario	Condition	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.
Factor loading = 0.8	1	=	0.058	0.056	0.002	0.056	0.058	-0.002
	2	+	0.044	0.040	0.004	0.048	0.040	0.008
	3	-	0.048	0.044	0.004	0.046	0.044	0.002
	4	=	0.060	0.058	0.002	0.056	0.054	0.002
	5	+	0.048	0.042	0.006	0.046	0.040	0.006
	6	-	0.048	0.056	-0.008	0.046	0.058	-0.012
	7	=	0.048	0.054	-0.006	0.054	0.054	0.000
	8	+	0.054	0.044	0.010	0.050	0.044	0.006
	9	-	0.056	0.060	-0.004	0.052	0.062	-0.010
	10	=	0.044	0.048	-0.004	0.050	0.046	0.004
	11	+	0.064	0.062	0.002	0.058	0.062	-0.004
	12	-	0.056	0.052	0.004	0.052	0.052	0.000
Factor loading = 0.4	1	=	0.029	0.034	-0.005	0.036	0.042	-0.006
	2	+	0.044	<u>0.018</u>	0.026	0.050	0.046	0.004
	3	-	<u>0.023</u>	0.036	-0.013	0.046	0.064	-0.018
	4	=	0.028	0.036	-0.008	0.030	0.056	-0.026
	5	+	0.026	<u>0.023</u>	0.003	0.030	0.025	0.005
	6	-	0.032	<u>0.022</u>	0.010	<u>0.015</u>	0.034	-0.019
	7	=	0.030	0.034	-0.004	0.036	0.034	0.002
	8	+	0.042	<u>0.018</u>	0.024	0.042	<u>0.020</u>	0.022
	9	-	0.033	<u>0.098</u>	-0.065	0.054	<u>0.088</u>	-0.034
	10	=	0.052	0.054	-0.002	0.054	0.050	0.004
	11	+	0.056	0.052	0.004	0.060	0.058	0.002
	12	-	0.064	0.070	-0.006	0.064	0.074	-0.010

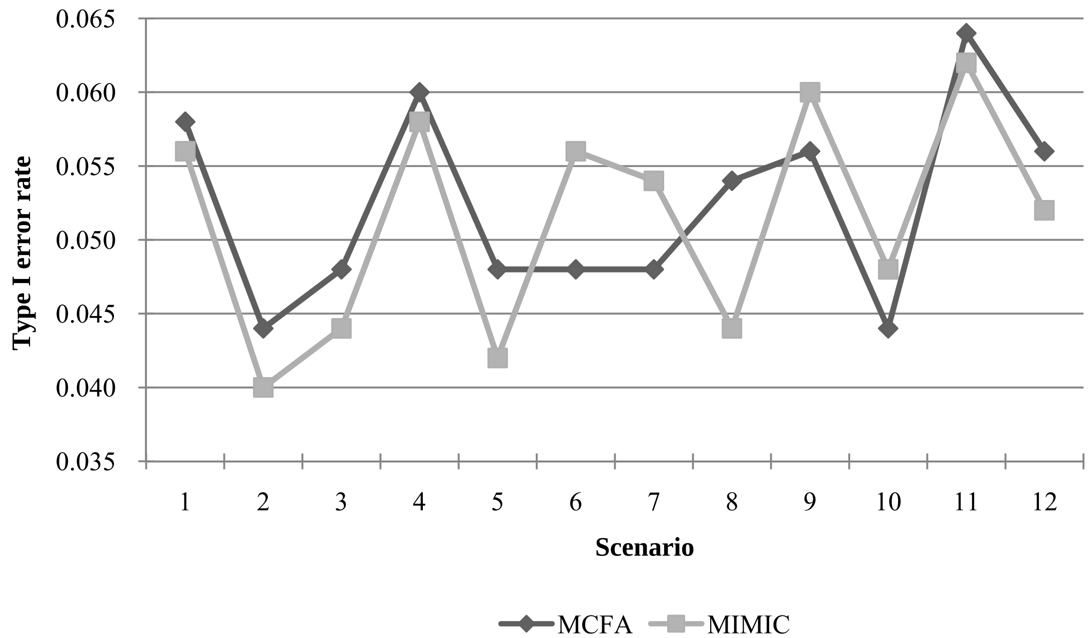
Note: UD = Unidimensional; DR = Domain representative; EA = Error adjustment; Diff = Difference of MIMIC from MCFA. Any error rate falling beyound Bradley’s (1978) liberal criterion of [.5α, 1.5 α] (i.e., [.025, .075]) is underlined.

Table 5.2 Parameter Estimates of the General Linear Model with Type I Error Rate as the Dependent Variable

Parameter	Effect	Std. Error	t	$p^{\dagger\dagger\dagger\dagger}$
Intercept	.031	.005	6.512	.000
Parceling = DR	.002	.003	.635	.527
Parceling = UD	-.002	.004	-.664	.508
Parceling = no†				
Approach = MCFA	.002	.003	.605	.546
Approach = MIMIC††				
Condition = negative	.005	.003	1.645	.103
Condition = Positive	-.001	.003	-.458	.648
Condition = balanced †††				
Loadings	.020	.006	3.274	.001

Note: †Without parceling is the reference of the three methods. ††MIMIC is the reference of the two approaches. †††Balanced condition is the reference of the three conditions. ††††All the *p*-values are two-tailed. DR = Domain representative parceling; UD = Unidimensional parceling.

criterion under the condition of low factor loadings, and MCFA had one (i.e., .023 in Scenario 3). By using domain representative parcels, MCFA had one Type I error rate (i.e., .015 in Scenario 6) and MIMIC had two Type I error rates (i.e., .020 in Scenario 8 and .088 in Scenario 9) out of the criterion. In another words, MCFA still controlled Type I error better than MIMIC did, and MIMIC still showed the tendency of being conservative in the positive condition and liberal in the negative condition when the parceled data were used. However, MIMIC displayed fewer out-of-criterion Type I error rates by using the parceled data than the unparceled original items.



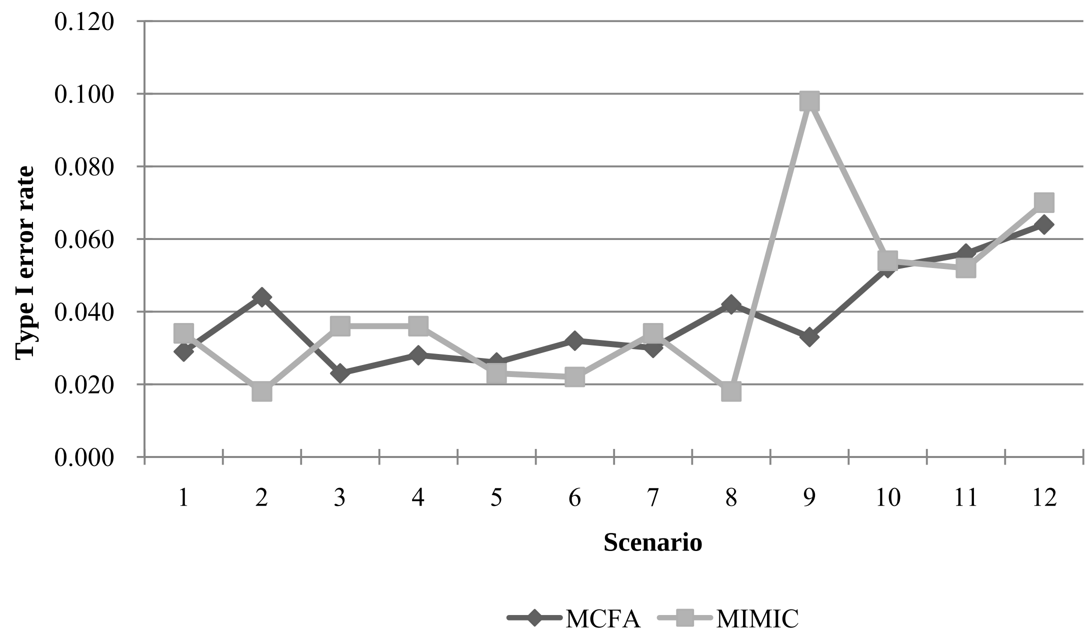
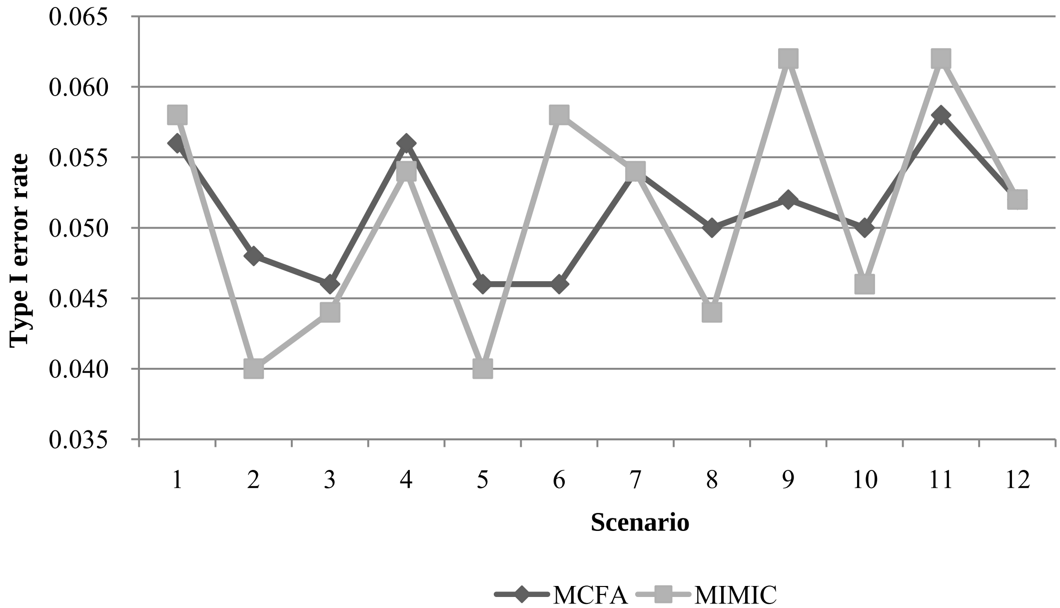


Fig. 5.1 Plots of Empirical Type I Error Rates of MCFA and MIMIC Approaches with the Unidimensional Parcels

Note. Upper: factor loading = .8; Down: factor loading = .4.



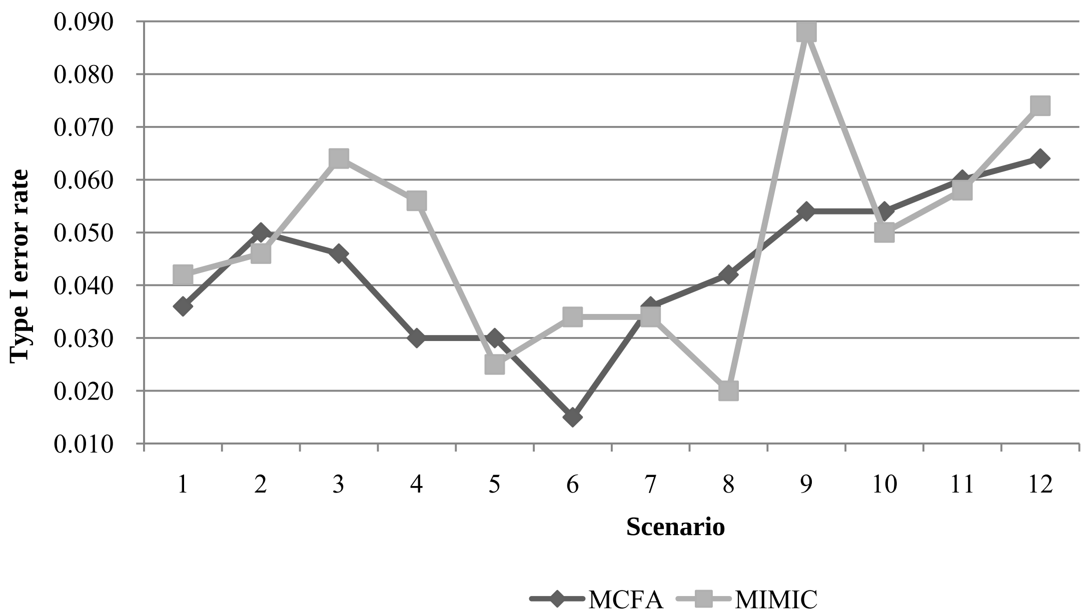


Fig. 5.2 Plots of Empirical Type I Error Rates of MCFA and MIMIC Approaches with the Domain Representative Parcels

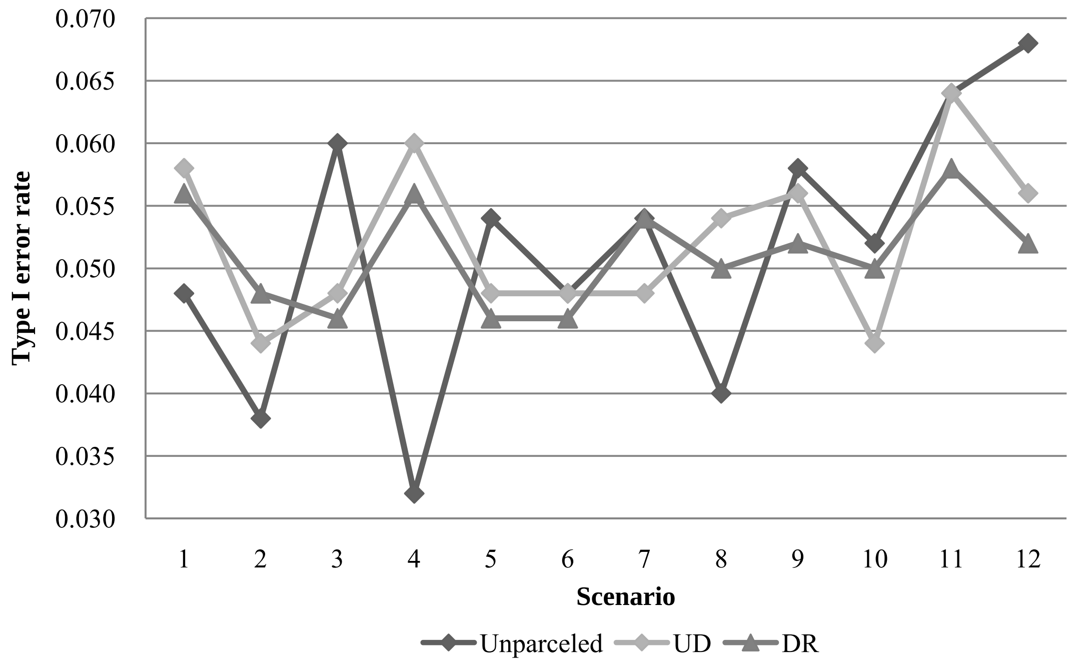
Note. Upper: factor loading = .8; Down: factor loading = .4.

The Type I error rates for unparceled and parceled data were plotted for the purpose of comparison (Figures 5.3 and 5.4). It is indicated that the Type I error rates obtained by using unparceled and parceled data were very close when the factor loadings were high. However, when the factor loadings were low, MCFA generally showed a lower Type I error rate by using the parceled data, especially the unidimensional parcels, while MIMIC generally showed a higher Type I error rate by using the parceled data, especially domain representative parcels, although these differences were not statistically supported when evaluated in a general linear model (see Table 5.2).

Empirical Power with the High Factor Loadings

The empirical power by using the unidimensional parceling strategy and the domain representative parceling strategy under the condition of the high factor loadings (i.e., 0.8) were presented in the upper part of Tables 5.3 and 5.4, respectively. The results of the general linear model were shown in Table 5.5. And the results were also plotted as shown in Figures 5.5 and 5.6. Again, the empirical power increased as $\Delta\kappa$ increased regardless of the parceling strategies ($p < .001$). Both of MCFA and MIMIC reached maximum power 1.000 when $\Delta\kappa$ was .5. The discrepancy between the two approaches decreased as $\Delta\kappa$ increased from .2 to .5. This trend was not obvious on Figures 5.5 and 5.6 because the discrepancy between the two approaches at any level of $\Delta\kappa$ was very minor (most of the discrepancies are below .01), particularly in the balanced condition.

When $\Delta\kappa$ was lower than median effect size, both MCFA and MIMIC displayed similar trends between the two parceling strategies. Firstly, the two approaches achieved



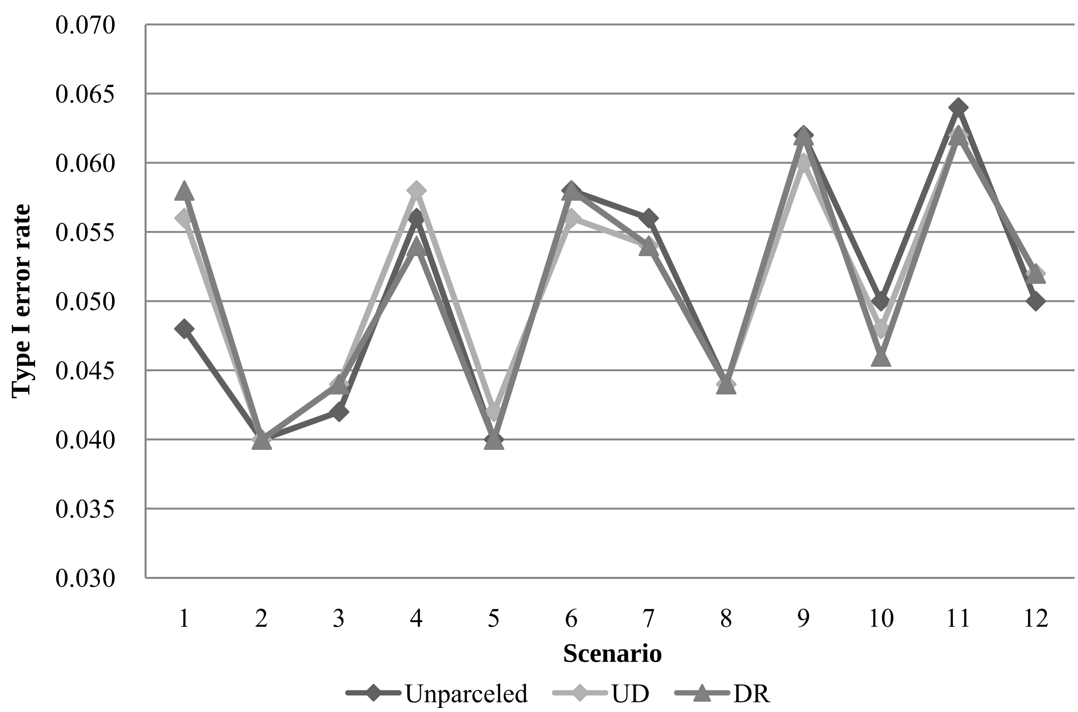
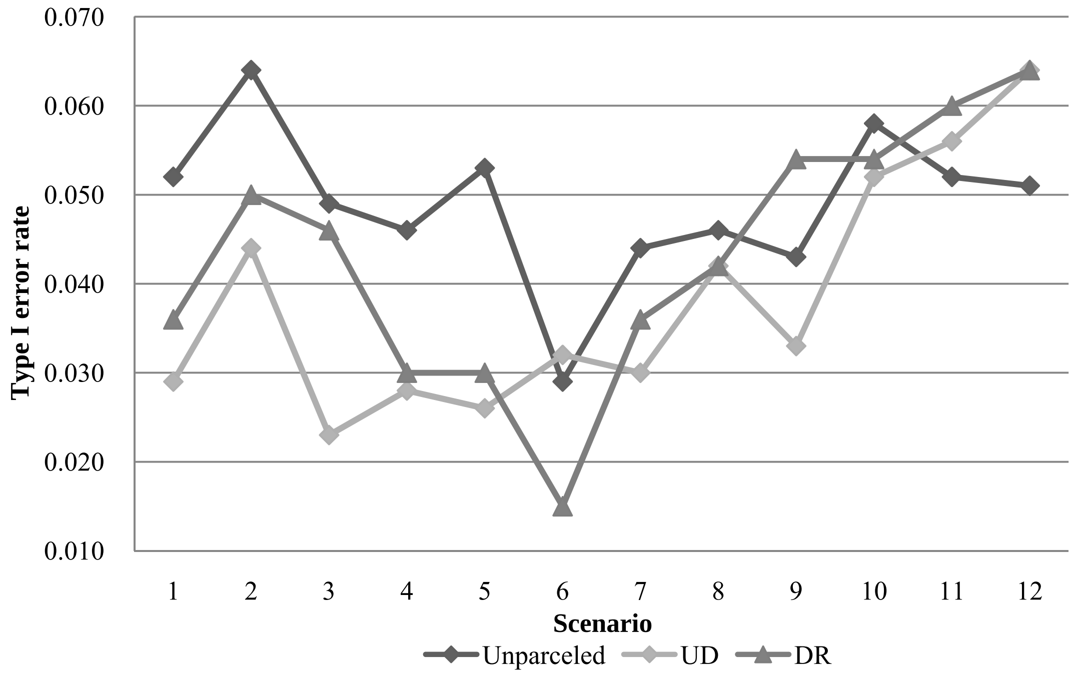


Fig. 5.3 Plots of Type I Error Rates by Using Unparceled Data and Parceled Data under the Condition of the High Factor Loadings

Note: Upper: MCFA; Down: MIMIC; UD = Unidimensional; DR = Domain representative.



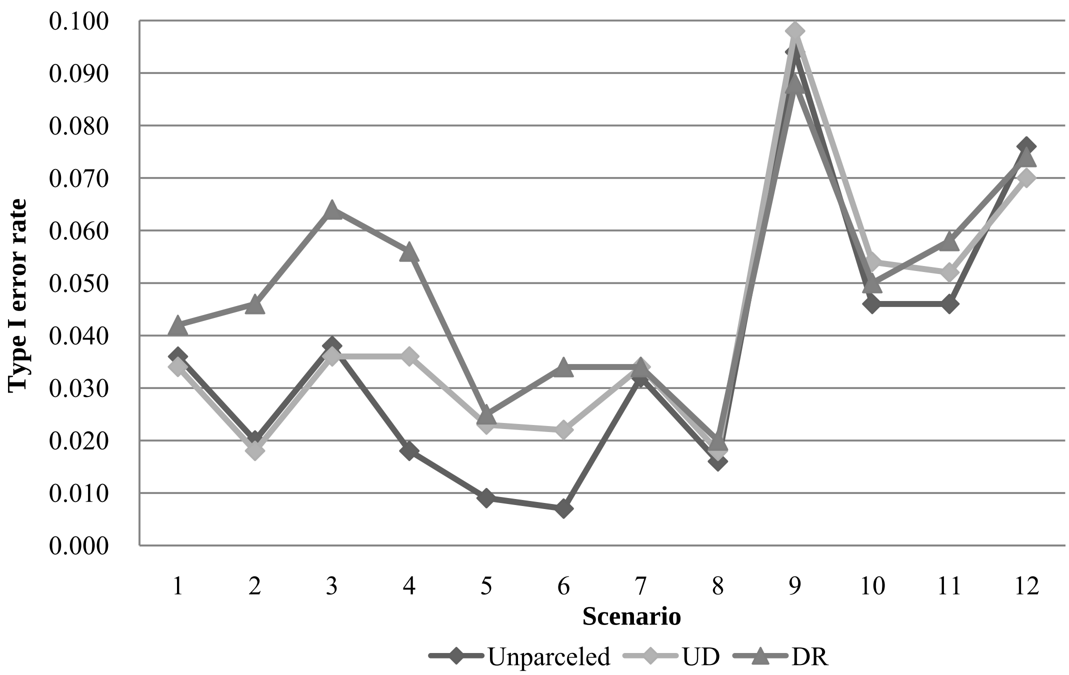


Fig. 5.4 Plots of Type I Error Rates by Using Unparceled Data and Parceled Data under the Condition of the Low Factor Loadings

Note: Upper: MCFA; Down: MIMIC; UD = Unidimensional; DR = Domain representative.

Table 5.3 Power of MCFA and MIMIC (Unidimensional Parcels)

			Δκ= .2			Δκ= .3			Δκ= .4			Δκ= .5		
	Scenario	Condition	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.
Factor loading = 0.8	1	=	0.740	0.746	-0.006	0.976	0.978	-0.002	0.998	0.998	0.000	1.000	1.000	0.000
	2	+	0.648	0.626	0.022	0.916	0.908	0.008	0.994	0.996	-0.002	1.000	1.000	0.000
	3	-	0.626	0.626	0.000	0.916	0.918	-0.002	0.992	0.992	0.000	1.000	1.000	0.000
	4	=	0.752	0.756	-0.004	0.980	0.978	0.002	1.000	1.000	0.000	1.000	1.000	0.000
	5	+	0.648	0.628	0.020	0.926	0.918	0.008	1.000	0.998	0.002	1.000	1.000	0.000
	6	-	0.622	0.644	-0.022	0.930	0.934	-0.004	0.992	0.992	0.000	1.000	1.000	0.000
	7	=	0.766	0.754	0.012	0.980	0.982	-0.002	1.000	1.000	0.000	1.000	1.000	0.000
	8	+	0.670	0.642	0.028	0.932	0.920	0.012	0.998	0.998	0.000	1.000	1.000	0.000
	9	-	0.630	0.652	-0.022	0.924	0.932	-0.008	0.992	0.994	-0.002	1.000	1.000	0.000
	10	=	0.774	0.770	0.004	0.984	0.986	-0.002	1.000	1.000	0.000	1.000	1.000	0.000
	11	+	0.638	0.654	-0.016	0.930	0.938	-0.008	0.992	0.994	-0.002	1.000	1.000	0.000
	12	-	0.650	0.634	0.016	0.938	0.934	0.004	0.994	0.994	0.000	1.000	1.000	0.000

Table 5.3 (Continued)

	Scenario	Condition	Δκ= .2 MCFA	Δκ= .2 MIMIC	Δκ= .2 Diff.	Δκ= .3 MCFA	Δκ= .3 MIMIC	Δκ= .3 Diff.	Δκ= .4 MCFA	Δκ= .4 MIMIC	Δκ= .4 Diff.	Δκ= .5 MCFA	Δκ= .5 MIMIC	Δκ= .5 Diff.
Factor loading = 0.4	1	=	0.236	0.332	-0.096	0.490	0.630	-0.140	0.715	0.858	-0.143	0.896	0.972	-0.076
	2	+	0.226	0.232	-0.006	0.497	0.474	0.023	0.674	0.710	-0.036	0.878	0.906	-0.028
	3	-	0.124	0.278	-0.154	0.273	0.522	-0.249	0.468	0.774	-0.306	0.654	0.920	-0.266
	4	=	0.129	0.166	-0.037	0.282	0.417	-0.135	0.514	0.704	-0.190	0.721	0.876	-0.155
	5	+	0.132	0.142	-0.010	0.274	0.334	-0.060	0.448	0.540	-0.092	0.574	0.687	-0.113
	6	-	0.103	0.178	-0.075	0.204	0.355	-0.151	0.388	0.602	-0.214	0.539	0.812	-0.273
	7	=	0.442	0.504	-0.062	0.798	0.846	-0.048	0.964	0.976	-0.012	0.992	0.998	-0.006
	8	+	0.440	0.370	0.070	0.752	0.692	0.060	0.940	0.908	0.032	0.996	0.998	-0.002
	9	-	0.240	0.450	-0.210	0.538	0.800	-0.262	0.752	0.934	-0.182	0.922	0.990	-0.068
	10	=	0.502	0.504	-0.002	0.828	0.840	-0.012	0.972	0.980	-0.008	1.000	1.000	0.000
	11	+	0.374	0.360	0.014	0.728	0.724	0.004	0.924	0.930	-0.006	0.992	0.996	-0.004
	12	-	0.388	0.440	-0.052	0.718	0.740	-0.022	0.922	0.928	-0.006	0.980	0.984	-0.004

Note. A equal (=) sign indicates a balanced condition in which two groups have identical sample sizes; a plus (+) sign indicates a positive pairing of sample with with generalized variance; a negative (-) sign indicates a negative pairing. Diff. = Difference of MIMIC from MCFA.

Table 5.4 Power of MCFA and MIMIC (Domain Representative Parcels)

			Δκ= .2			Δκ= .3			Δκ= .4			Δκ= .5		
	Scenario	Condition	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.
Factor loading = 0.8	1	=	0.744	0.748	-0.004	0.976	0.976	0.000	1.000	1.000	0.000	1.000	1.000	0.000
	2	+	0.624	0.630	-0.006	0.910	0.910	0.000	0.996	0.996	0.000	1.000	1.000	0.000
	3	-	0.630	0.632	-0.002	0.926	0.922	0.004	0.992	0.992	0.000	1.000	1.000	0.000
	4	=	0.750	0.756	-0.006	0.978	0.978	0.000	1.000	1.000	0.000	1.000	1.000	0.000
	5	+	0.650	0.632	0.018	0.926	0.916	0.010	0.998	0.998	0.000	1.000	1.000	0.000
	6	-	0.630	0.640	-0.010	0.928	0.940	-0.012	0.996	0.994	0.002	1.000	1.000	0.000
	7	=	0.764	0.760	0.004	0.982	0.982	0.000	1.000	1.000	0.000	1.000	1.000	0.000
	8	+	0.654	0.644	0.010	0.934	0.920	0.014	0.998	0.998	0.000	1.000	1.000	0.000
	9	-	0.640	0.662	-0.022	0.932	0.936	-0.004	0.992	0.994	-0.002	1.000	1.000	0.000
	10	=	0.770	0.770	0.000	0.980	0.986	-0.006	1.000	1.000	0.000	1.000	1.000	0.000
	11	+	0.650	0.660	-0.010	0.932	0.936	-0.004	0.992	0.994	-0.002	1.000	1.000	0.000
	12	-	0.656	0.638	0.018	0.934	0.932	0.002	0.994	0.994	0.000	1.000	1.000	0.000

Table 5.4 (Continued)

			Δκ= .2			Δκ= .3			Δκ= .4			Δκ= .5		
	Scenario	Condition	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.
Factor loading = 0.4	1	=	0.374	0.382	-0.008	0.684	0.690	-0.006	0.900	0.908	-0.008	0.982	0.988	-0.006
	2	+	0.352	0.302	0.050	0.600	0.574	0.026	0.796	0.756	0.040	0.944	0.938	0.006
	3	-	0.272	0.314	-0.042	0.520	0.590	-0.070	0.774	0.826	-0.052	0.914	0.934	-0.020
	4	=	0.200	0.234	-0.034	0.463	0.514	-0.051	0.718	0.792	-0.074	0.881	0.920	-0.039
	5	+	0.162	0.194	-0.032	0.354	0.377	-0.023	0.592	0.621	-0.029	0.748	0.770	-0.022
	6	-	0.141	0.198	-0.057	0.316	0.426	-0.110	0.528	0.658	-0.130	0.726	0.842	-0.116
	7	=	0.520	0.530	-0.010	0.854	0.858	-0.004	0.980	0.980	0.000	0.998	0.998	0.000
	8	+	0.488	0.390	0.098	0.790	0.712	0.078	0.958	0.926	0.032	1.000	0.998	0.002
	9	-	0.334	0.442	-0.108	0.682	0.790	-0.108	0.870	0.936	-0.066	0.984	0.992	-0.008
	10	=	0.528	0.522	0.006	0.844	0.846	-0.002	0.984	0.978	0.006	1.000	1.000	0.000
	11	+	0.404	0.362	0.042	0.774	0.730	0.044	0.954	0.932	0.022	0.998	0.996	0.002
	12	-	0.422	0.458	-0.036	0.724	0.744	-0.020	0.922	0.936	-0.014	0.984	0.984	0.000

Note. A equal (=) sign indicates a balanced condition in which two groups have identical sample sizes; a plus (+) sign indicates a positive pairing of sample with with generalized variance; a negative (-) sign indicates a negative pairing. Diff. = Difference of MIMIC from MCFA.

Table 5.5 Parameter Estimates of the General Linear Model with Power as the Dependent Variable

	Parameter	Effect	Std. Error	t	*p*†††††
Power	Intercept	-.110	.029	-3.789	.000
	Parceling = DR	.057	.015	3.665	.000
	Parceling = UD	.043	.015	2.847	.005
	Parceling = no†				
	Approach = MCFA	-.040	.012	-3.213	.001
	Approach = MIMIC††				
	Condition = negative	-.070	.015	-4.616	.000
	Condition = Positive	-.052	.015	-3.419	.001
	Condition = balanced †††				
	Loadings	.701	.031	22.669	.000
	$\Delta\kappa$	1.263	.049	25.970	.000
	DR×loadings	-.354	.076	-4.648	.000
	UD×loadings	-.181	.074	-2.449	.015
	Unparceled×loadings††††				

Note: †Without parceling is the reference of the three methods. ††MIMIC is the reference of the two approaches. †††Balanced condition is the reference of the three conditions. ††††Unparceled×loadings is the reference. †††††All the *p*-values are two-tailed. DR = Domain representative parceling; UD = Unidimensional parceling.

higher power in the balanced condition than the negative or positive condition by using either of the two parceling strategies ($p < .001$ and $p = .001$, respectively). These trends were similar to the trend that MIMIC showed by using the unparceled data but inconsistent with the trend that MCFA displayed by using the unparceled data. Secondly, power for MCFA generally had superiority over MIMIC in the positive condition except in Scenario 11, and MIMIC generally had superiority over MCFA in the negative condition except in Scenario 12. This trend was as same as the trend the two approaches

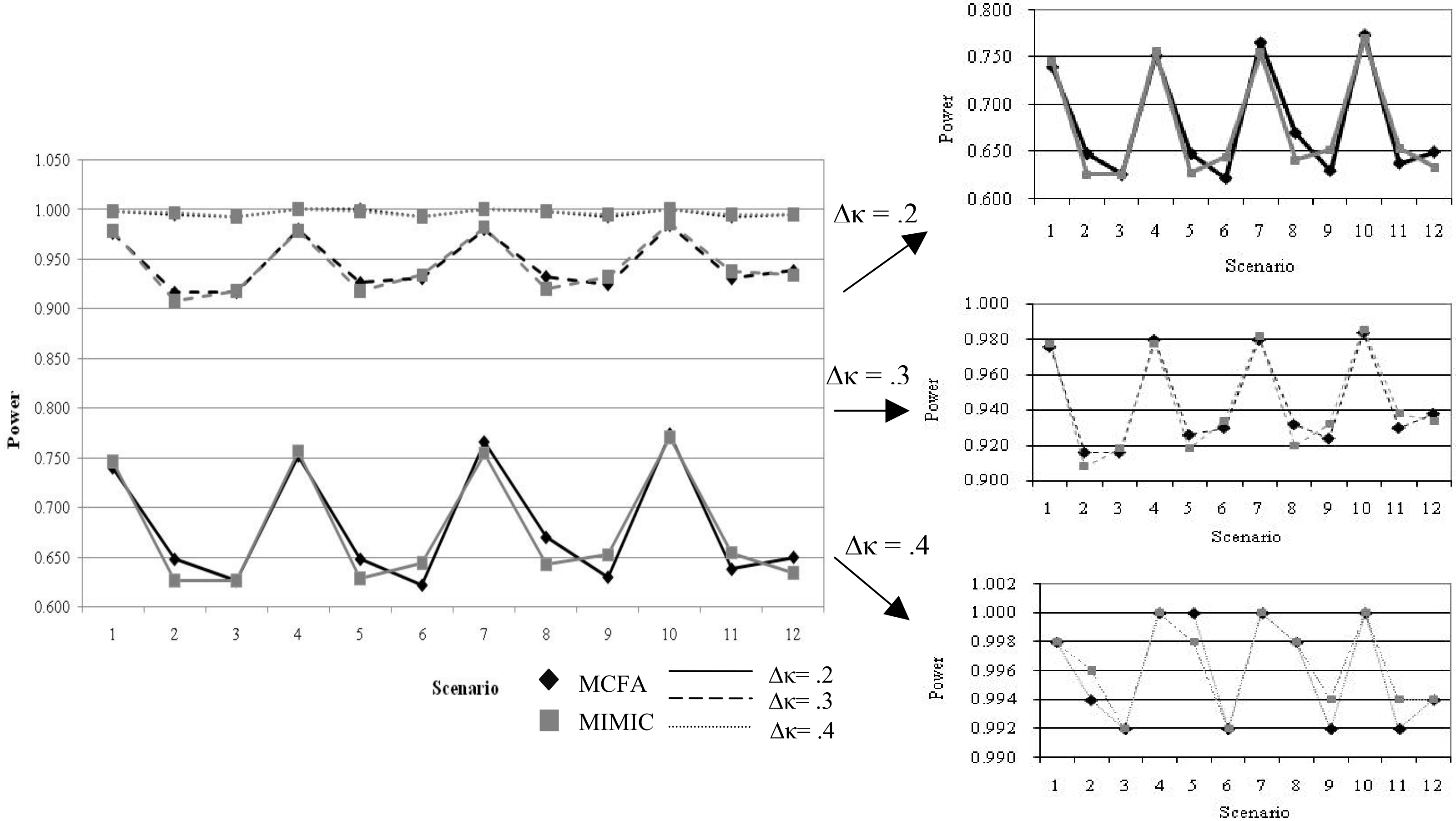


Fig. 5.5 Plots of Empirical Power of MCFA and MIMIC Approaches with the Unidimensional Parcels When the Factor Loadings Were High

Note: The plots for both of the two approaches fall on the straight line at 1.000 when Δκ is .5 (not shown on this figure). The three small figures on the right are the decomposed figures of the one on the left based on the values of Δκ.

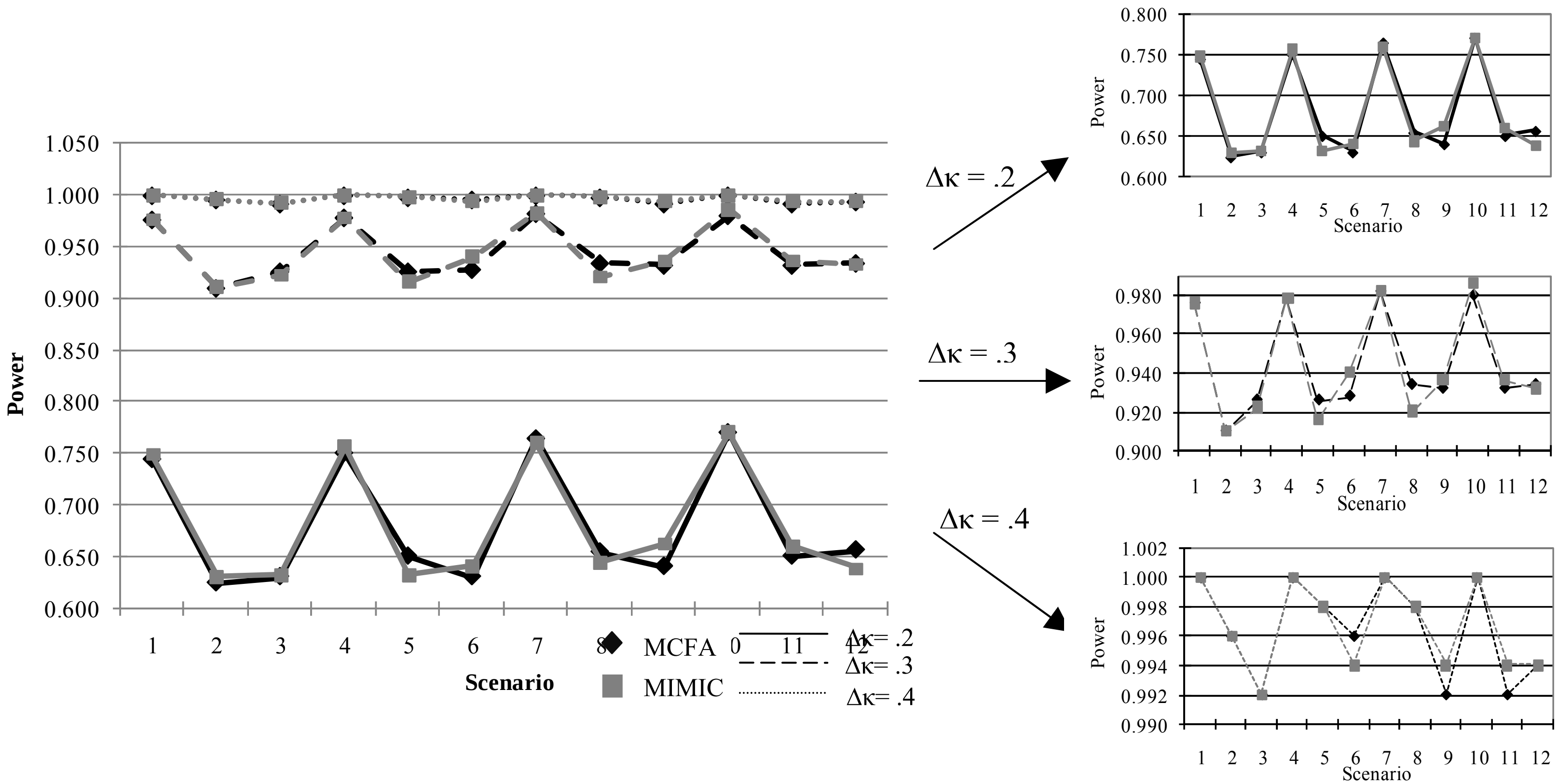
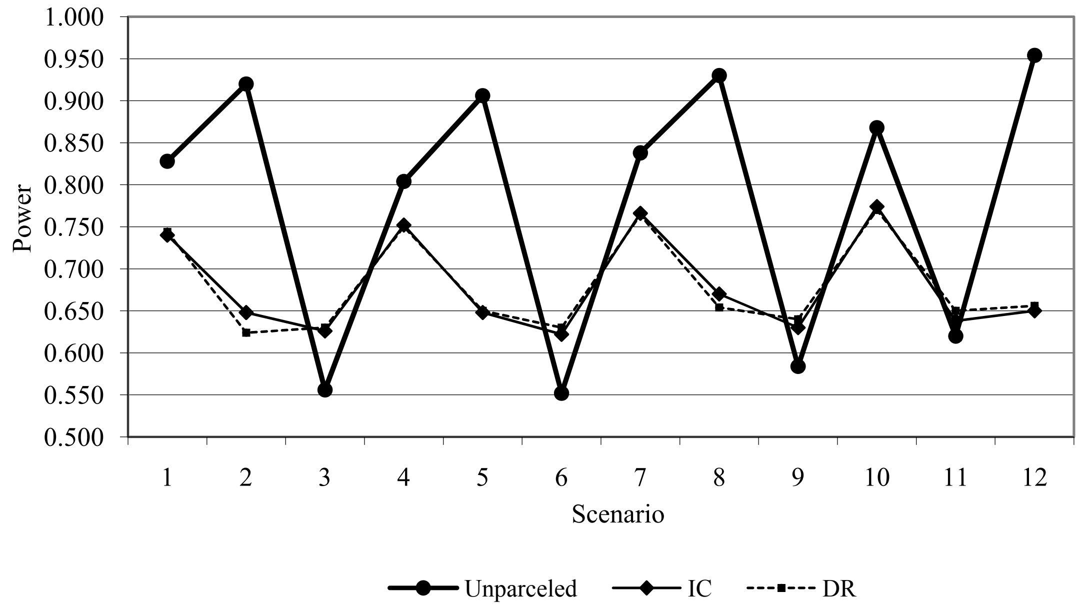


Fig. 5.6 Plots of Empirical Power of MCFA and MIMIC Approaches with the Domain Representative Parcels When the Factor Loadings Were High

Note: The plots for both of the two approaches fall on the straight line at 1.000 when Δκ is .5 (not shown on this figure). The three small figures on the right are the decomposed figures of the one on the left based on the values of Δκ displayed for unparceled data under the condition of the high factor loadings, except that the superiority of either of the two approaches displayed over the other one by using the unparceled data was higher than that by using the unidimensional or domain representative parcels.

The empirical power achieved by using unparceled data, unidimensional parcels, and domain representative parcels under the condition of high factor loadings was plotted at $\Delta\kappa = .2$ for MCFA and MIMIC respectively (Figure 5.7). Small effect size for $\Delta\kappa$ was chosen when making this comparison because either MCFA or MIMIC models using the two parceling strategies and using unparceled data showed the largest discrepancy in terms of power at $\Delta\kappa = .2$, and this discrepancy decreased as $\Delta\kappa$ increased. Figure 5.7 shows that parceling strategies parceling strategies improved power for MCFA in the negative condition except in Scenario 12. However, this improvement was very limited compared with the decrease of power in the balanced or positive condition by applying the parceling strategies. On the other hand, parceling strategies, especially domain representative parceling, did improve the empirical power for MIMIC in most of the scenarios, although this improvement never exceeded .010. For example, in Scenario 1, the power of MIMIC by using unparceled data was .744. It increased to .746 by using unidimensional parceling and .748 by using domain representative parceling.



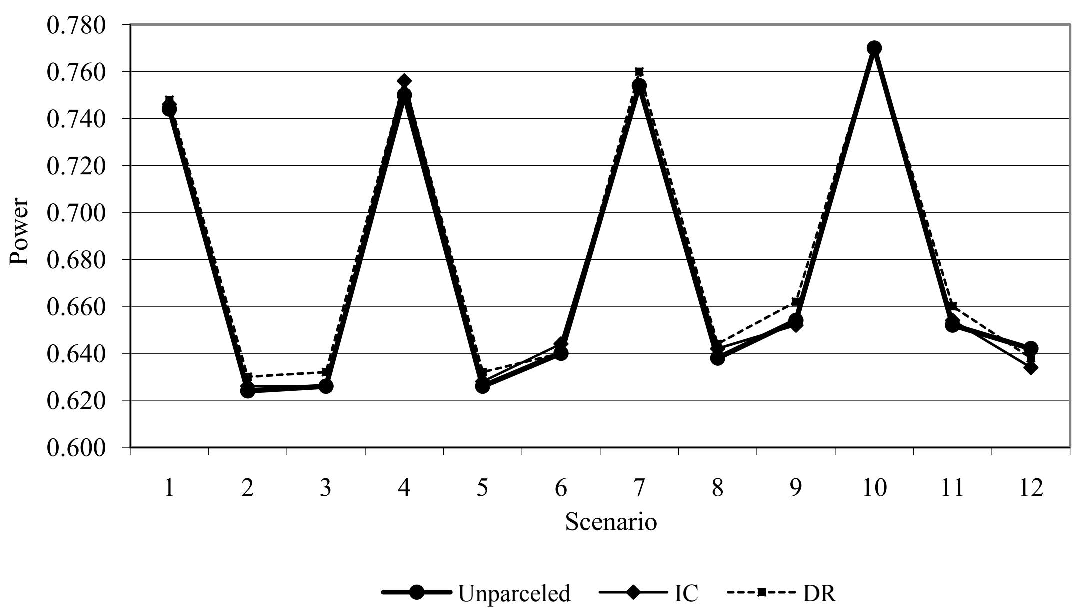


Fig. 5.7 Plots of Empirical Power by Using Unparceled Data and Parceled Data When $\Delta K = .2$ under the Condition of High Factor Loadings

Note: Upper: MCFA; Down: MIMIC; UD = Unidimensional; DR = Domain representative.

Empirical Power with the Low Factor Loadings

The results of power analysis with low factor loadings by applying unidimensional parceling strategy and domain representative parceling are presented in the lower part of Tables 5.3 and 5.4, and plotted in Figures 5.8 and 5.9 respectively. As expected, the values of power with the low factor loadings were lower than those with high factor loadings at each level of $\Delta\kappa$ ($p < .001$), and power increased as $\Delta\kappa$ increased from a small effect size to median effect size, regardless of the parceling strategies ($p < .001$). Compared with the differences of MCFA and MIMIC under the condition of high factor loadings, the differences between the two approaches under the condition of low factor loadings were generally larger. However, the decreasing discrepancy between the two approaches as $\Delta\kappa$ increased, which was displayed under the condition of high factor loadings, was not observed when the factor loadings were low.

Some trends can be observed in Figures 5.8 and 5.9. Firstly, as with the plots of power by for the unparceled data under the condition of low factor loadings, the plots for both of the two approaches by using either unidimensional parcels or domain representative parcels displayed a “U” shape across the 12 scenarios regardless of the values of $\Delta\kappa$. Scenarios 4, 5, and 6 were still at the bottom of the “U”. However, the “U” shape was not as deep as the one obtained by using the unparceled data. Secondly, MIMIC showed superiority in the negative condition at each level of $\Delta\kappa$ by using either of the parceling strategies. MIMIC also shows superiority in the balanced condition in most of the scenarios at each level of $\Delta\kappa$ by using either of the parceling strategies. Actually, in the few scenarios in the balanced condition when MIMIC was not favored,

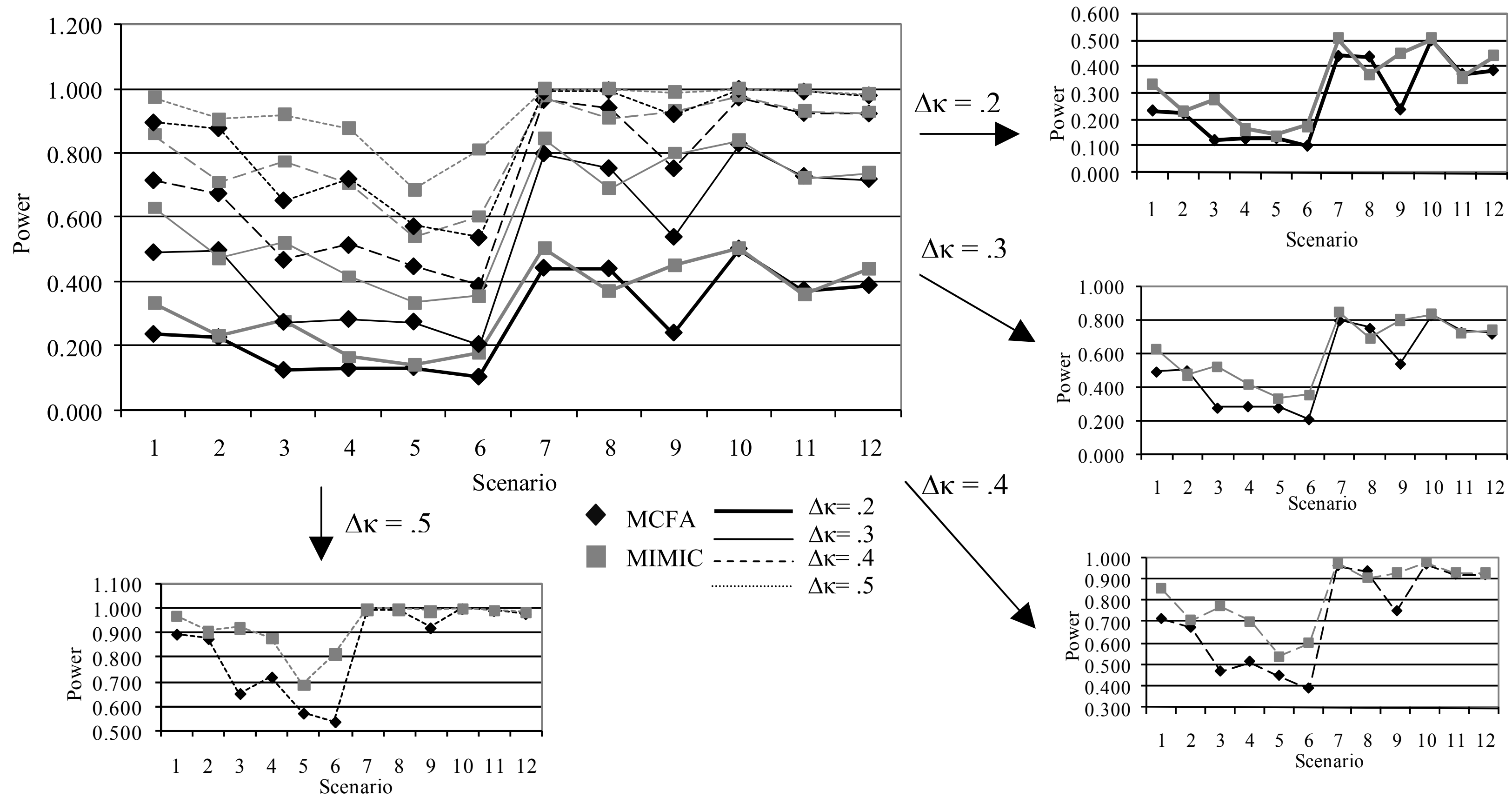


Fig. 5.8 Plots of Empirical Power of MCFA and MIMIC Approaches with the Unidimensional Parcels When the Factor Loadings Were Low

Note: The four small figures are the decomposed figures of the left top one based on the mean differences.

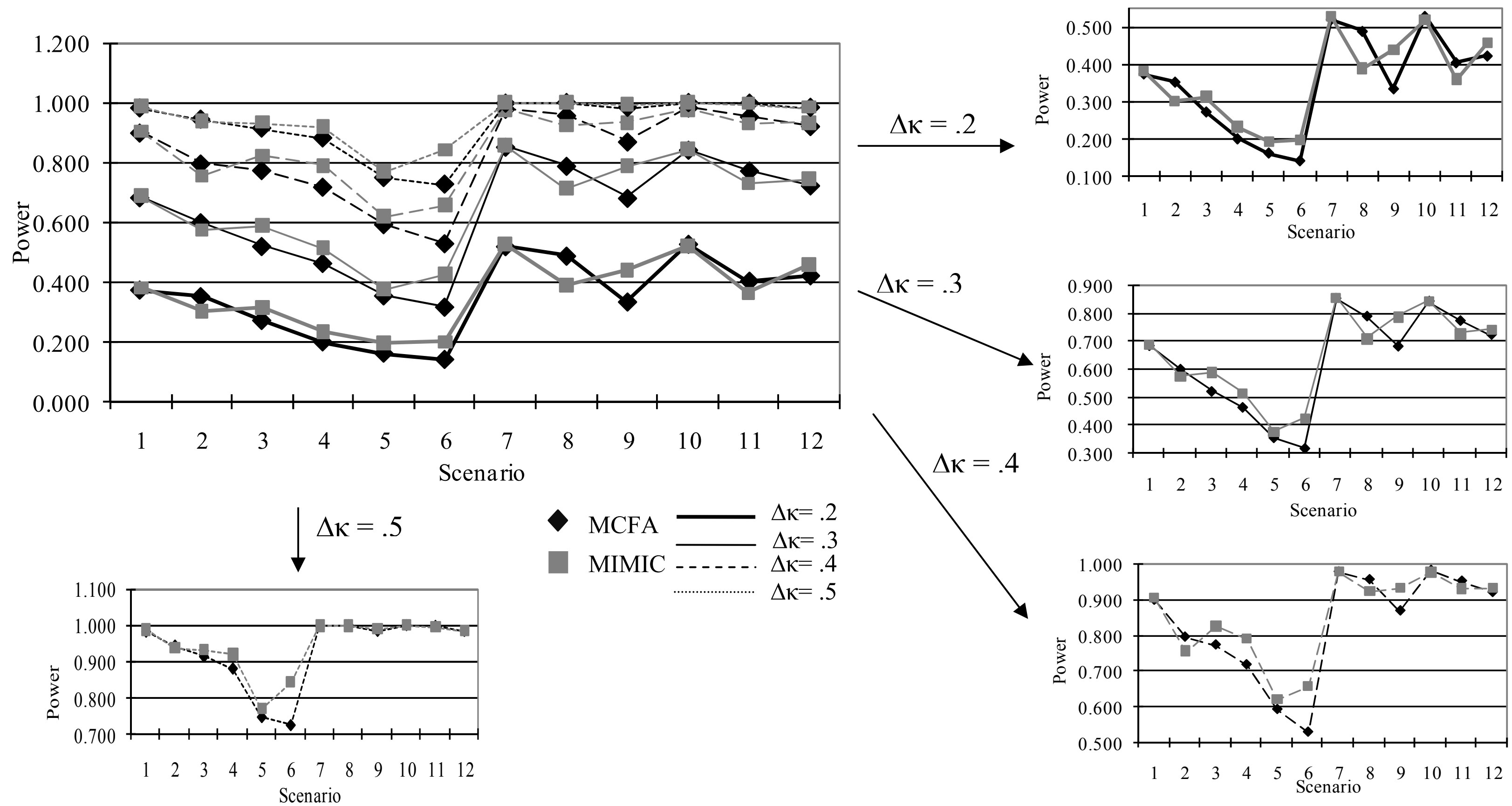


Fig. 5.9 Plots of Empirical Power of MCFA and MIMIC Approaches with the Domain Representative Parcels When the Factor Loadings Were Low

Note: The four small figures are the decomposed figures of the left top one based on the mean differences.

the two approaches had either identical power (e.g., in Scenario 10 when $\Delta\kappa$ was .5 and the parceling strategy was unidimensional) or MCFA's superiority was very minor, which never exceeded .010 (e.g., in Scenario 10, when $\Delta\kappa$ was .2 and the parceling strategy was domain representative). Thirdly, MCFA generally had higher power than MIMIC in the positive condition at each level of $\Delta\kappa$ except in Scenario 5. Note that Scenario 5 had the lowest convergence rate among the 12 scenarios by using either of the parceling strategies.

Again, the comparison of the empirical power achieved by using unparceled data, unidimensional parcels, and domain representative parcels under the condition of the low factor loadings was made at $\Delta\kappa = .2$ for MCFA and MIMIC separately (Figure 5.10). Figure 5.10 shows that for either of the two approaches, the models with parceling strategies applied, especially domain representative strategy, generally had higher power than the models with unparceled data. For MCFA, the greatest superiority by using unidimensional parceling strategy and domain representative parceling was .175 and .253, respectively (Scenario 7) and the lowest superiority was .011 (Scenario 4) and .024 (Scenario 5), respectively. MIMIC models also benefited from using either of the parceling strategies in terms of power. The greatest superiority in power by using unidimensional parceling and domain representative parceling for MIMIC was .084 (Scenario 6) and .137 (Scenario 4), respectively, and the lowest superiority was .004 (Scenario 2) and .004 (Scenario 9), respectively. Compared with the improvement the parceling strategies make under the condition of high factor loadings, the parceling

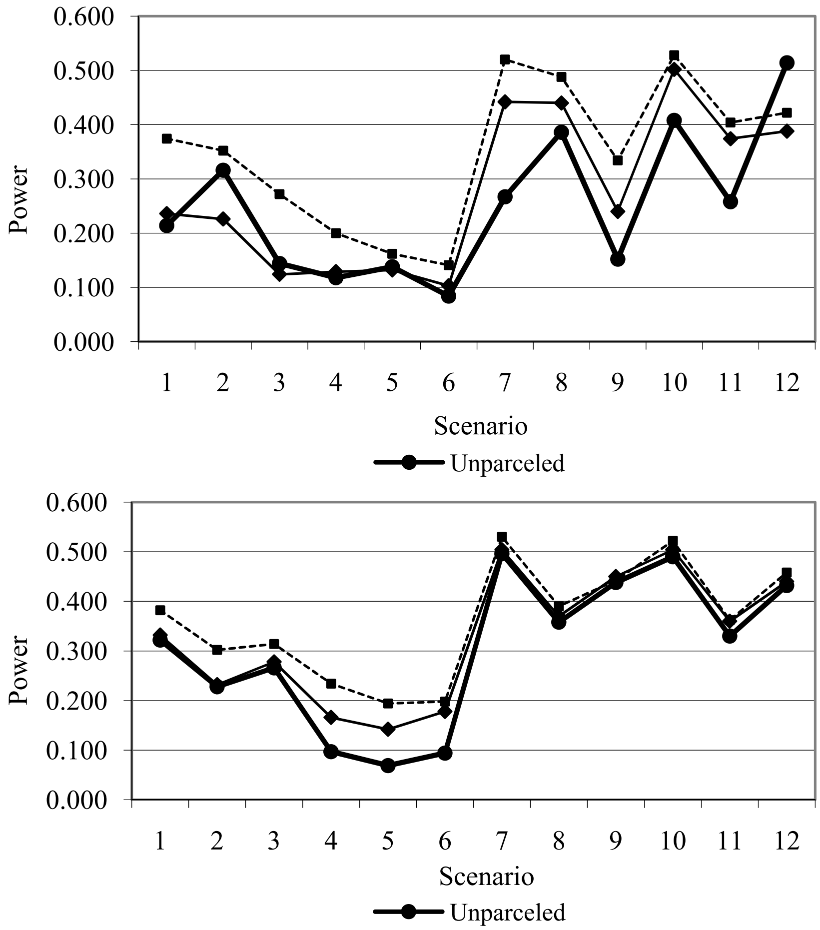


Fig. 5.10 Plots of Empirical Power by Using Unparceled Data and Parceled Data When ΔK = .2 under the Condition of Low Factor Loadings

Note: Upper: MCFA; Down: MIMIC; UD = Unidimensional; DR = Domain representative.

strategies obviously did better in improving power of MCFA and MIMIC models when the factor loadings were low.

The different behaviors parceling strategies displayed under the condition of high factor loadings and low factor loadings were also statistically supported by the general linear model. It is shown in Table 5.5 that the interactions of parceling strategies and the factor loadings were statistically significant. The general linear model was also run for the high factor loadings and the low factor separately. The results (Table 5.6) indicate that parceling did not significantly improve power under the condition of the high factor loadings, whereas power was significantly improved by parceling under the condition of the low factor loadings. The coefficients of domain representative parceling and unidimensional parceling under the condition of the low factor loadings were .127 and .092, respectively, which indicated that when holding other factors constant, the power obtained by using domain representative parceling and unidimensional parceling was on average .127 and .092 higher than that obtained by using unparceled data respectively, with both *p*-values less than .001.

Error Adjustment

The empirical power of MCFA and MIMIC models were also examined by using domain representative parceling strategy with the error adjustment. As introduced before, when the first-order factor was collapsed by aggregating the items, the second-order factor was derived from the parcels instead of the error-free first-order factors. As a result, the empirical power and the empirical Type I error rate when examining the mean value of the second-order factor across different groups might be negatively influenced. By doing the error adjustment, the residual variance of the parcel was constraint as the calculated measurement error variance so that each parcel can be treated as an error-free indicator.

Table 5.6 Parameter Estimates of the General Linear Model under the Condition of the High Factor Loadings and the Low Factor Loadings with Power as the Dependent Variable

	Parameter	Effect	Std. Error	t	p†††††
High factor loadings	Intercept	.641	.016	38.993	.000
	Parceling = DR	-.014	.011	-1.240	.216
	Parceling = UD	-.006	.011	-.507	.613
	Parceling = no†				
	Approach = MCFA	.004	.009	.472	.638
	Approach = MIMIC††				
	Condition = negative	-.050	.011	-4.527	.000
	Condition = Positive	-.040	.011	-3.622	.000
	Condition = balanced †††				
	$\Delta\kappa$	.876	.036	24.606	.000
Low factor loadings	Intercept	.084	.038	2.200	.029
	Parceling = DR	.127	.026	4.867	.000
	Parceling = UD	.092	.026	3.584	.000
	Parceling = no†				
	Approach = MCFA	-.084	.021	-4.001	.000
	Approach = MIMIC††				
	Condition = negative	-.090	.026	-3.496	.001
	Condition = Positive	-.063	.026	-2.472	.014
	Condition = balanced †††				
	$\Delta\kappa$	1.649	.082	20.041	.000

Note: †Without parceling is the reference of the three methods. ††MIMIC is the reference of the two approaches. †††Balanced condition is the reference of the three conditions. ††††Unparceled×loadings is the reference. †††††All the *p*-values are two-tailed. DR = Domain representative parceling; UD = Unidimensional parceling.

The results of the empirical Type I error rate and the empirical power with the error adjustment were presented in Tables 5.7 and 5.8 respectively. Because the goal of adjusting the residual variance in this study was to examine whether parceling induced bias in terms of Type I error rate and power when the first-order factors were collapsed, the empirical Type I error rate and power achieved by using domain representative parceling with error adjustment vs. without error adjustment were plotted shown in Figures 5.11 and 5.12, respectively. For the same reason stated before, power with and without error adjustment was only compared when $\Delta\kappa$ was .2. It is shown in Figure 5.11 that Type I error rates achieved from the two methods were very close when the factor loadings were high, particularly for MIMIC, and the difference for either of the two approaches never exceeded .005. When the factor loadings were low, the difference was still generally lower than .005 except those in Scenarios 4, 5, and 6 for both of the two approaches. For example, the differences between with and without error adjustment for MCFA in Scenarios 4, 5, and 6 were .028, .030, and .017, and for MIMIC were .014, .013, and .006, respectively, and for both of the two approaches, the Type I error rates achieved without error adjustment in the three scenarios were more conservative. Again, note that the two approaches, especially MCFA, had severe convergence problem in the three scenarios when the factor loadings were low.

The comparison of the empirical power at small effect size (i.e., $\Delta\kappa = .2$) between domain representative parceling without error adjustment and with error adjustment showed very similar patterns as those of the Type I error rate. That is, when the factor loadings were high, adjusting the residual variance did not yield much difference of

Table 5.7 Type I Error Rates of MCFA and MIMIC (Domain Representative Parcels with Error Adjustment)

		Factor loading = 0.8			Factor loading = 0.4		
Scenario	Condition	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.
1	=	0.054	0.058	-0.004	0.034	0.038	-0.004
2	+	0.046	0.038	0.008	0.034	0.046	-0.012
3	-	0.046	0.040	0.006	0.044	0.060	-0.016
4	=	0.054	0.056	-0.002	0.058	0.070	-0.012
5	+	0.044	0.040	0.004	0.060	0.038	0.022
6	-	0.050	0.056	-0.006	0.032	0.040	-0.008
7	=	0.050	0.052	-0.002	0.040	0.036	0.004
8	+	0.050	0.044	0.006	0.040	0.020	0.020
9	-	0.052	0.062	-0.010	0.052	<u>0.092</u>	-0.040
10	=	0.048	0.046	0.002	0.054	0.052	0.002
11	+	0.054	0.058	-0.004	0.066	0.056	0.010
12	-	0.054	0.054	0.000	0.066	0.074	-0.008

Note*:* A equal (=) sign indicates a balanced condition in which two groups have identical sample sizes; a plus (+) sign indicates a positive pairing of sample with with generalized variance; a negative (-) sign indicates a negative pairing. Diff. = difference of MIMIC from MCFA. Any error rate falling beyound Bradley's (1978) liberal criterion of [.5α, 1.5 α] (i.e., [.025, .075]) is underlined.

Table 5.8 Power of MCFA and MIMIC (Domain Representative Parcels with Error Adjustment)

			Δκ= .2			Δκ= .3			Δκ= .4			Δκ= .5		
	Scenario	Condition	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.
Factor loading = 0.8	1	=	0.742	0.746	-0.004	0.978	0.978	0.000	1.000	1.000	0.000	1.000	1.000	0.000
	2	+	0.634	0.630	0.004	0.910	0.910	0.000	0.996	0.996	0.000	1.000	1.000	0.000
	3	-	0.624	0.628	-0.004	0.924	0.922	0.002	0.992	0.992	0.000	1.000	1.000	0.000
	4	=	0.746	0.752	-0.006	0.978	0.978	0.000	1.000	1.000	0.000	1.000	1.000	0.000
	5	+	0.650	0.628	0.022	0.924	0.916	0.008	0.998	0.998	0.000	1.000	1.000	0.000
	6	-	0.620	0.636	-0.016	0.928	0.936	-0.008	0.994	0.994	0.000	1.000	1.000	0.000
	7	=	0.758	0.760	-0.002	0.982	0.982	0.000	1.000	1.000	0.000	1.000	1.000	0.000
	8	+	0.666	0.642	0.024	0.936	0.920	0.016	0.998	0.998	0.000	1.000	1.000	0.000
	9	-	0.626	0.662	-0.036	0.930	0.930	0.000	0.994	0.994	0.000	1.000	1.000	0.000
	10	=	0.766	0.766	0.000	0.982	0.986	-0.004	1.000	1.000	0.000	1.000	1.000	0.000
	11	+	0.650	0.654	-0.004	0.932	0.936	-0.004	0.992	0.994	-0.002	1.000	1.000	0.000
	12	-	0.660	0.632	0.028	0.936	0.934	0.002	0.994	0.994	0.000	1.000	1.000	0.000

Table 5.8 (Continued)

	Scenario	Condition	Δκ= .2			Δκ= .3			Δκ= .4			Δκ= .5		
			MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.	MCFA	MIMIC	Diff.
Factor loading = 0.4	1	=	0.378	0.384	-0.006	0.688	0.690	-0.002	0.906	0.906	0.000	0.984	0.988	-0.004
	2	+	0.344	0.300	0.044	0.600	0.572	0.028	0.792	0.758	0.034	0.948	0.940	0.008
	3	-	0.266	0.314	-0.048	0.522	0.570	-0.048	0.790	0.820	-0.030	0.922	0.932	-0.010
	4	=	0.246	0.252	-0.006	0.484	0.528	-0.044	0.740	0.772	-0.032	0.910	0.920	-0.010
	5	+	0.104	0.200	-0.096	0.118	0.420	-0.302	0.200	0.648	-0.448	0.348	0.796	-0.448
	6	-	0.176	0.206	-0.030	0.374	0.382	-0.008	0.610	0.640	-0.030	0.840	0.850	-0.010
	7	=	0.514	0.532	-0.018	0.850	0.862	-0.012	0.980	0.980	0.000	0.998	0.998	0.000
	8	+	0.498	0.396	0.102	0.774	0.710	0.064	0.954	0.928	0.026	1.000	0.998	0.002
	9	-	0.358	0.440	-0.082	0.718	0.788	-0.070	0.890	0.944	-0.054	0.984	0.992	-0.008
	10	=	0.528	0.526	0.002	0.842	0.840	0.002	0.978	0.978	0.000	1.000	1.000	0.000
	11	+	0.402	0.368	0.034	0.774	0.734	0.040	0.952	0.940	0.012	1.000	0.996	0.004
	12	-	0.424	0.454	-0.030	0.738	0.740	-0.002	0.924	0.932	-0.008	0.984	0.986	-0.002

Note. A equal (=) sign indicates a balanced condition in which two groups have identical sample sizes; a plus (+) sign indicates a positive pairing of sample with with generalized variance; a negative (-) sign indicates a negative pairing. Diff. = Difference of MIMIC from MCFA.

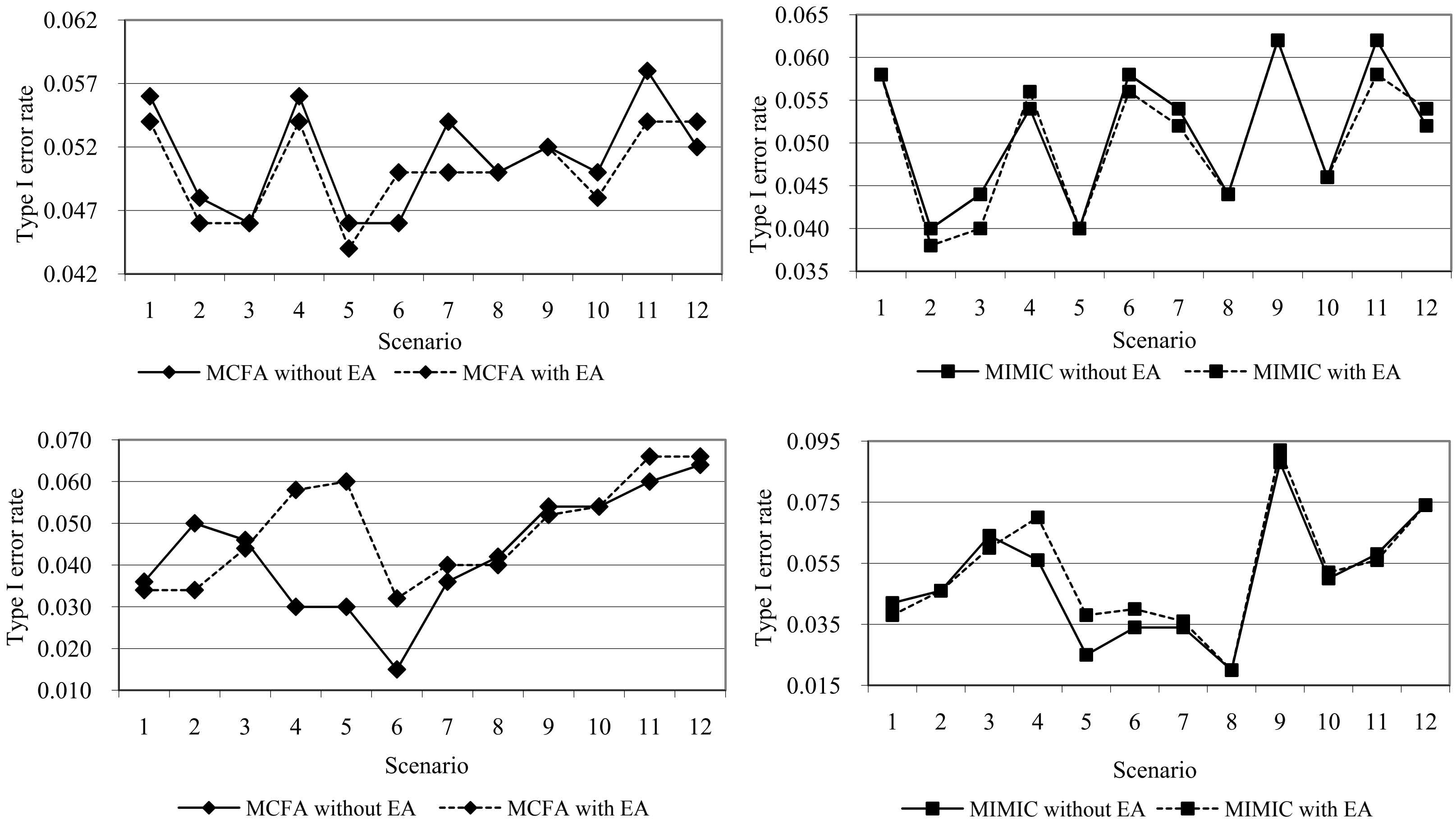


Fig. 5.11 Plots of Type I Error Rates by Using Domain Representative Parceling Strategy with and without Error Adjustment

Note. Up left: MCFA with high factor loadings; Up right: MIMIC with high factor loadings; Down left: MCFA with low factor loadings; Down right: MIMIC with low factor loadings

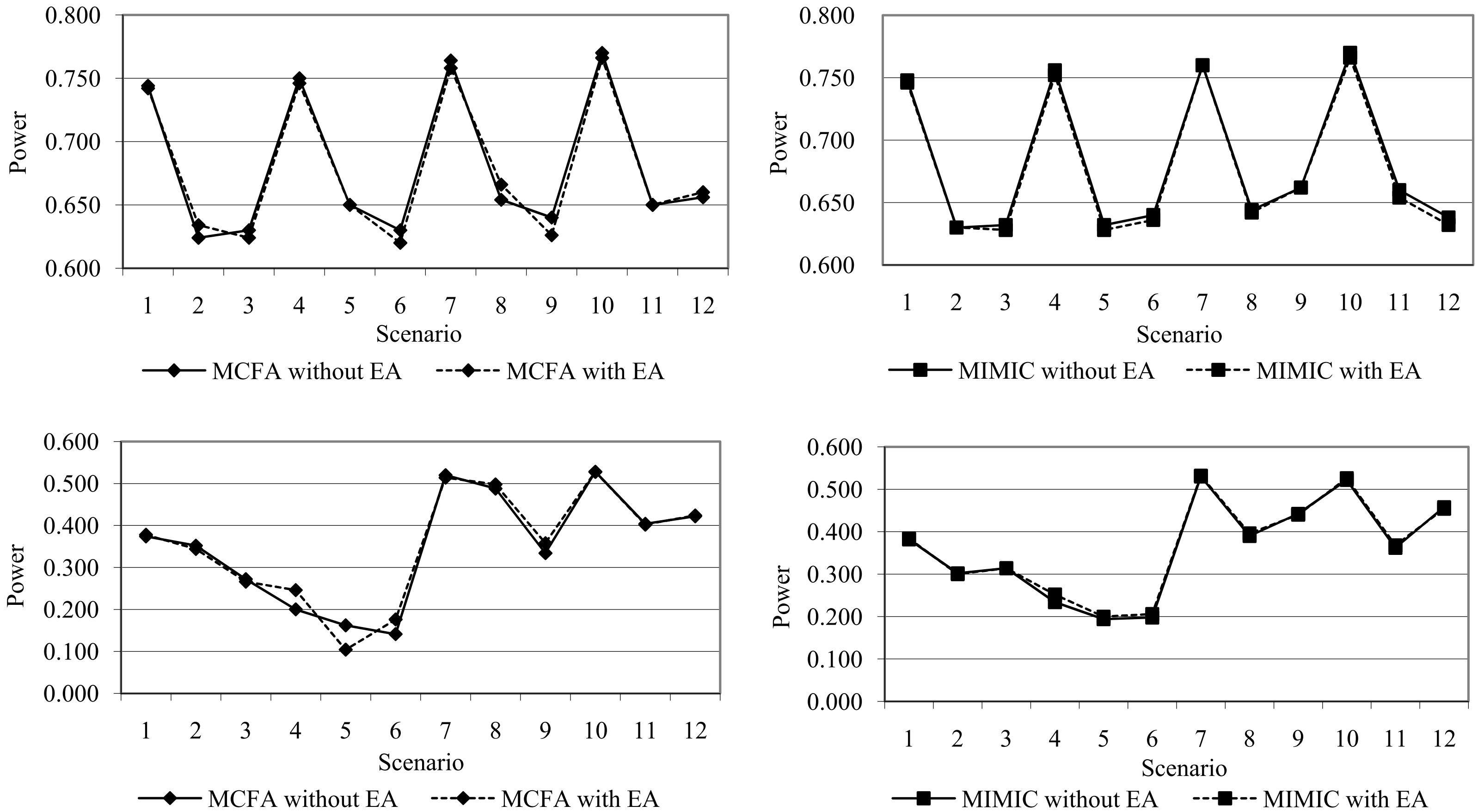


Fig. 5.12 Plots of Power by Using Domain Representative Parceling Strategy with and without Error Adjustment When ΔK = .2

Note. Up left: MCFA with high factor loadings; Up right: MIMIC with high factor loadings; Down left: MCFA with low factor loadings; Down right: MIMIC with low factor loadings power for either of the two approaches (e.g., most of the differences are around .005); when the factor loadings were low, the differences still remained lower than .01, except those of MCFA in the scenarios 4, 5, and 6. In these three scenarios, the differences between with and without error adjustment for MCFA were .046, .058, and .035, respectively, which were much higher than the differences in the other scenarios. The differences for MIMIC in the three scenarios were .018, .006, and .008, respectively, which were slightly higher than the differences in the other scenarios. The results of general linear model analysis (Table 5.9) also indicate that neither Type I error rate (p = .553) nor power (p = .853) was significantly different between using domain representative parceling without error adjustment and with error adjustment.

DISCUSSION

The results in study 2 show that MCFA and MIMIC both control Type I error rate well and the type I error rates are generally very close between the two approaches under the condition of high factor loadings by using the parceling strategies. These two approaches also show very similar values of power across the 12 scenarios when the factor loadings are high. It indicates that both of the two approaches with parceled data are equally favored when the factor loadings are high. When the factor loadings are low, MCFA with parceled data controls Type I error rate more appropriately than MIMIC given the MIMIC's tendency of being more conservative in the positive condition and liberal in the negative condition. And MCFA tends to have superiority in terms of power in the positive condition, and MIMIC tends to be more powerful than MCFA in the

Table 5.9 Parameter Estimates of the General Linear Model (Domain Representative Parcels)

	Parameter	Effect	Std. Error	t	$p^{\dagger\dagger\dagger\dagger}$
Type I error	Intercept	.044	.005	9.235	.000
	Parceling = DR without EA	-.002	.003	-.596	.553
	Parceling = DR with EA†				
	Approach = MCFA	-.002	.003	-.976	.332
	Approach = MIMIC††				
	Condition = negative	.005	.003	1.611	.111
	Condition = Positive	-.004	.003	-1.145	.255
	Condition = balanced †††				
	Loadings	.006	.006	1.026	.308
Power	Intercept	-.054	.031	-1.748	.081
	Parceling = DR without EA	.002	.013	.185	.853
	Parceling = DR with EA†				
	Approach = MCFA	-.012	.013	-.904	.367
	Approach = MIMIC††				
	Condition = negative	-.064	.016	-4.075	.000
	Condition = Positive	-.072	.016	-4.592	.000
	Condition = balanced †††				
	Loadings	.530	.032	16.646	.000
	$\Delta\kappa$	1.466	.057	25.738	.000

Note: †Domain representative parceling with error adjustment is the reference. †† MIMIC is the reference of the two approaches. †††Balanced condition is the reference of the three conditions. ††††All the *p*-values are two-tailed. DR = Domain representative parceling; EA = error adjustment.

balanced and negative condition (although the superiorities of the two approaches are very minor when using domain representative parceling). These two patterns are same as those observed in Study 1 when the original items were used under the condition of the low factor loadings.

More interesting is the comparison of the results between Study 2 and Study 1. It indicates that parceling does have effects on group comparison by using MCFA and MIMIC under the Condition of either the high factor loadings or the low factor loadings. Firstly, when the factor loadings are high, parceling decreases the power of MCFA in the balanced and positive condition but increases the power of MCFA in the negative condition, although the improvement never exceeds .1. However, parceling does not influence MIMIC in terms of power too much (Figure 5.7). Therefore, parceling strategies might not be necessary for the purpose of group comparison under the Condition of the high factor loadings.

Secondly, parceling, especially domain representative parceling, improves MIMIC's ability in controlling Type I error rate when the factor loadings are low. This improvement might be caused by the decreasing relative discrepancy of the generalized variances between the two groups after the parceling strategies were used. Indeed, domain parceling strategy results in smaller relative discrepancy than unidimensional parceling does in this study. As the relative discrepancy of the generalized variances decreases, the coefficient of the regression line (i.e., γ) tends to reflect the true difference of the means between the two groups. Parceling also improves MIMIC's ability of controlling Type I error rate in Scenarios 4-6 by improving the convergence rates in the three scenarios. Indeed, the three scenarios have higher convergence rates by using unidimensional parcels (two out of three under-criterion Type I error rates) than using the original data (three out of three under-criterion Type I error rates), and even higher convergence rates by using domain representative parcels (zero out of three under-criterion Type I error rates) than using unidimensional parcels.

Thirdly, parceling generally decreases MCFA's Type I error rate but increases MIMIC's Type I error rate under the condition of the low factor loadings (Figure 5.4). As a result, the two approaches have more consistent Type I error rates. On the other hand, parceling, especially domain representative parceling, increases the power of both of the two approaches over all the conditions under the circumstance of the low factor loadings (Figure 5.10).

It leads to the conclusion that parceling could positively impact the group comparison under the condition of the low factor loadings by improving power, mitigating the convergence problem, and controlling Type I error rate within reasonable range for MIMIC. Domain representative parceling obviously does better in all the three aspects than unidimensional parceling. Therefore MCFA models with domain representative parcels are preferred to MCFA and MIMIC models with unidimensional parcels and original items. On the other hand if the few out-of-criterion Type I error rates for MIMIC when domain representative parceling is applied could be ignored, MIMIC would be a good choice too over all the conditions. Given the fact that the two approaches show very minor differences across the 12 scenarios under the condition of the low factor loadings when domain representative parceling is used, the two approaches are equally favored in this case.

Study 2 also offer the evidence that collapsing the first-order factors by using domain parceling strategies does not bring much bias to the results when the convergence rate is high. However, the discrepancies of either Type I error rate or power between with error adjustment and without adjustment for both of the two approaches are large in Scenarios 4-6 which have severe convergence problem under the condition of the low factor loadings. It means that neither the Type I error rate nor the power is reliable when the convergence rate is low.

CHAPTER VI

CONCLUSIONS

In educational and psychological studies, factor loadings as high as .8 are not typically observed. Also the differences in the mean values of the latent factors across different groups usually have small effect sizes. This study shows that MCFA should be favored when the groups are not in the balanced condition (i.e., different group sizes are paired with the different generalized variances) if researchers in social sciences and other related disciplines investigate the mean differences of the second-order factor. On the other hand, when the groups are in the balanced condition, MIMIC outperforms MCFA with respect to both Type I error rate and power.

Convergence problems uncovered here could be caused by the large random error variances in second-order CFA models. Both MCFA and MIMIC display extremely low power when the convergence rate is low. In this case reevaluation of the study design or/and the method of data collection should have higher priority than selecting the approach for analyzing the data.

This study also offers evidence that parceling has some advantages over using the original unparceled data. Firstly, parceling can improve MIMIC's ability of controlling Type I error rate and significantly improve power under the condition of the low factor loadings, especially domain representative parceling. When applying domain representative parceling under the condition of the low factor loadings, the two approaches, i.e., MCFA and MIMIC, have very minor differences with respect to Type I error rate and power, and hence are equally favored across all the conditions.

Secondly, parceling can increase the convergence rate. As a result, the problems of out-of-criterion Type I error rate for MIMIC and the extremely low power for the two approaches resulted from the low convergence rate could be mitigated. However, in future studies researchers should attempt to determine possible reasons for the low convergence rate (e.g., high random error variance and low reliability in this study) and then solve these problems before they resort to parceling. Lastly, parceling does not induce bias to Type I error rate and power of the second-order CFA models when collapsing the first-order factors.

The limitations of the study begin with the condition that the factor loadings in both levels were restricted to a specific fixed value, which may not adequately represent the range of realistic models. In addition, the domain representative parceling strategy resulted in a more parsimonious first-order CFA model than the unidimensional parceling did in this study. Therefore the parsimony of the models with domain representative parcels might be the cause of the higher power and better control of Type I error rate. These factors will be considered in further research comparing the two methods in the context of second-order CFA models.

REFERENCES

Agrawal, A., & Lynskey, M. T. (2007). Does gender contribute to heterogeneity in criteria for cannabis abuse and dependence? Results from the national epidemiological survey on alcohol and related conditions. *Drug and Alcohol Dependence, 88*, 300–307.

Anderson, K. G., Smith, G. T., McCarthy, D. M., Fischer, S. F., Fister, S., Grodin, D., Boerner, L. M., & Hill K. K. (2005). Elementary school drinking: The role of temperament and learning. *Psychology of Addictive Behaviors, 19*, 21-27.

Angoff, W. H. (1953). Test reliability and effective test length. *Psychometrika, 18,* 1–14.

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

Bandalos, D. L. (1997). Assessing sources of error in structural equation models: The effects of sample size, reliability, and model misspecifications. *Structural Equation Modeling, 4,* 177–192.

Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 269–296). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling, 9,* 78–102.

Bandalos, D. L. (2008). Is parceling really necessary? A comparison of results from item parceling and categorical variable methodology. *Structural Equation Modeling, 15,* 211-240.

Bedeian, A. G., Day, D. V., & Kelloway, E. K. (1997). Correcting for measurement error attenuation in structural equation models: Some important reminders. *Educational and Psychological Measurement, 57*, 785-799.

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

Bradley, J. C. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, 144-152.

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

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling, 12*, 471-492.

Cohen, P., Cohen, J., Teresi, J., Marchi, M., & Velez, C. N. (1990). Problems in the measurement of latent variables in structural equations causal models. *Applied Psychological Measurement, 14*, 183-196.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Belmont, CA: Wadsworth.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Ps*ychometrika, 16,* 297–334.

Feldt, L. S. (1975). Estimation of the reliability of a test divided into two parts of unequal length. *Psychometrika, 40,* 557–561.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.) *Educational measurement* (3rd ed., pp. 105–146). New York: American Council on Education and Macmillan.

Feldt, L. S., & Charter, R. A. (2003). Estimating the reliability of a test split into two parts of equal or unequal length. *Psychological methods, 8*, 102-109.

Gorsuch, R. L. (1983). *Factor Analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gribbons, B. C., & Hocevar, D. (1998). Levels of aggregation in higher level confirmatory factor analysis: Application for academic self-concept. *Structural Equation Modeling, 5*, 377-390.

Hakstian, A. R., Roed, J. C., & Lind, J. C. (1979). Two-sample T^2 procedure and the assumption of homogeneous covariance matrices. *Psychological Bulletin*, *86*, 1255–1263.

Hall, R. J., Snell, A. F., & Singer-Foust, M. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods*, 2, 233–256.

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, 7(4),* 534-556.

Hau, K., & Marsh, H.W. (2004). The use of item parcels in structural equation modeling: Non-normal data and small sample sizes. *British Psychological Society, 57,* 327–351.

Holloway, L. N.,& Dunn, O. J. (1967). The robustness of Hotelling's T^2. *Journal of the American Statistical Association*, *62*, 124–136.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data*. Beverly Hills, CA: Sage.

Kishton, J. M., & Widaman, K. F. (1994). Unidimensional versus domain representative parceling of questionnaire items: An empirical example. *Educational and Psychological Measurement, 54,*757–765.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling (2nd ed.).* New York, NY: The Guilford Press.

Little, T. D., Lindenberger U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods, 4*, 192-211.

Little, T. D., Cunningham,W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9, 151–173.

MacCallum, R. C.,Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4,* 84–99.

Marsh, H.W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33,* 181–220.

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

Meuleners, L. B., Lee, A. H., Binns, C. W., & Lower A. (2003). Quality of life for adolescents: Assessing measurement properties using structural equation modeling. *Quality of Life Research, 12*, 283-290.

Muthen, L. K., & Muthen, B. O. (2007). *Mplus user's guide (version 5)*. Los Angeles, CA: Author.

Nasser, F., & Takahashi, T. (2003). The effect of using item parcels on ad hoc goodness-of-fit indexes in confirmatory factor analysis: An example using Sarason's Reactions to Tests. *Applied Measurement in Education*, *16*, 75–97.

Nasser, F., & Wisenbaker, J. (2006). A Monte Carlo study investigating the impact of item parceling strategies on parameter estimates and their standard errors in CFA. *Structural Equation Modeling*, *13*, 204–228.

Prats, D. C. (1990, April). *The effects of forming miniscales on the construct validity of the test anxiety inventory.* Paper presented at the National Council of Measurement of Education, Boston, MA.

Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research, 23,* 51–67.

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

Spearman, C. (1927). *The abilities of man.* London: Macmillan.

Thompson, B. (1990). SECONDOR: A program that computes a second-order principal components analysis and various interpretation aids. *Educational and Psychological Measurement*, *50*, 575-580.

Thompson, B., & Melancon, J. (November, 1996). *Using item "testlets/parcels" in confirmatory factor analysis: An example using the PPDP-78.* Paper presented at the annual meeting of the Mid-South Educational Research Association, Tuscaloosa, AL.

Tinsley, H. E. A., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology*, *34*, 414-424.

Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, *24,* 185–201.

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

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention:*

Methodological advances from alcohol and substance abuse research (pp. 281–324). Washington, DC: American Psychological Association.

Wiersma , W. and Jurs, S. O. (1990). *Educational Measurement and Testing.* (2nd Ed.) Boston: Allyn and Bacon.

Windle, M., & Mason, A. W. (2004). General and specific predictors of behavioral and emotional problems among adolescents. *Journal of Emotional and Behavioral Disorders*, *12*, 49-61.

APPENDIX A

SAS CODE FOR UNIDIMENSIONAL PARCELING

```
%macro files;
%do i= 1 %to 12;
%do j = 1 %to 500;
%let n =&j;
%let m=&i;
%let file1 = G:\second_order_factor\monte carlo\unparceled\loading=.4\mean
difference=.5\s;
%let file2 = G:\second_order_factor\monte carlo\IC-parcel\loading=.4\mean
        difference=.5\s;
%let file3 = \s;
%let file4 = .dat;
%let file_in = &file1&m&file3&n&file4;
%let file_out =&file2&m&file3&n&file4;
data temp3;
infile "&file_in";
input x1  1-12 x2 13-25 x3 26-38 x4 39-51 x5 52-64 x6 65-77 x7 78-90 x8 91-103 x9
104-116  x10 117-129 x11 130-142 x12 143-155 group $ 158;
c1=(x1+x2+x3)/3;
c2=(x4+x5+x6)/3;
c3=(x7+x8+x9)/3;
c4=(x10+x11+x12)/3;
drop x1-x12;
run;
data parcel;
set temp3;
FILE "&file_out";
put @1 c1 12.6
   @13 c2 12.6
   @25 c3 12.6
   @37 c4 12.6
   @51 group 12.0;
run ;
%end;
%end;
%Mend;

%files;
```

APPENDIX B

SAS CODE FOR DOMAIN REPRESENTATIVE PARCELING

```
%macro files;
%do i= 1 %to 12;
%do j = 1 %to 500;
%let n =&j;
%let m=&i;
%let file1 = G:\second_order_factor\monte carlo\unparceled\loading=.4\mean
        difference=0\s;
%let file2 = G:\second_order_factor\monte carlo\DR-parcel\loading=.4\mean
        difference=.5\s;
%let file3 = \s;
%let file4 = .dat;
%let file_in = &file1&m&file3&n&file4;
%let file_out =&file2&m&file3&n&file4;
data temp3;
infile "&file_in";
input x1  1-12 x2 13-25 x3 26-38 x4 39-51 x5 52-64 x6 65-77 x7 78-90 x8 91-103 x9
        104-116
    x10 117-129 x11 130-142 x12 143-155 group $ 158;
c1=(x1+x4+x7+x10)/4;
c2=(x2+x5+x8+x11)/4;
c3=(x3+x6+x9+x12)/4;
drop x1-x12;
run;
data parcel;
set temp3;
FILE "&file_out";
put @1 c1 12.6
   @13 c2 12.6
   @25 c3 12.6
   @51 group 12.0;
run ;
%end;
%end;
%Mend;

%files;
```

APPENDIX C

SAS CODE FOR RUNNING MCFA MODELS WITH DOMAIN REPRESENTATIVE PARCELS WITH ERROR ADJUSTMENT

```
%macro mplus;


X mkdir "G:\DR addjustment\MCFA";
%do i= 1 %to 12;
%do j = 1 %to 500;
%let m=&i;
%let n =&j;
%let file1=G:\second_order_factor\monte carlo\unparceled\loading=.8\mean
difference=.5\s;
%let file2 = \s;
%let file3 = .dat;
/*Import unparceled data*/
data temp;
infile "&file1&m&file2&m&n&file3";
input x1  1-12 x2 13-25 x3 26-38 x4 39-51 x5 52-64 x6 65-77 x7 78-90 x8 91-103 x9
104-116 x10 117-129 x11 130-142 x12 143-155 group $ 158;;
run;
data group1;
set temp;
if group=1;
c11=(x1+x4+x7+x10)/4;
c12=(x2+x5+x8+x11)/4;
c13=(x3+x6+x9+x12)/4;
run;

data group2;
set temp;
if group=2;
c21=(x1+x4+x7+x10)/4;
c22=(x2+x5+x8+x11)/4;
c23=(x3+x6+x9+x12)/4;
run;

proc iml;
RESET NOLOG;
```

```
/*---------------------Calcualte error variance of composite 1 in group 1-----------------------
---*/
USE group1;
READ all var {x1 x4 x7 x10} INTO X_cons11;
read all var{c11} into X_comp11; /*cons means construct, comp means composite*/
*PRINT "The Data Matrix is" X_cons1;
*print "the composite score matrix is" X_comp1;

 /*Below is to calculate the reliability of the composite score 1*/
/*# of rows is 400, # of columns for the construct is 4, fo the composite is 1*/
 XBAR_cons11 = X_cons11(|+,|)`/400;
*PRINT, "XBAR_cons1 = " XBAR_cons1;
SUMSQ_cons11=X_cons11`*X_cons11-(XBAR_cons11*XBAR_cons11`)#400;
S_cons11=SUMSQ_cons11/(400-1);
*PRINT , "The Variance-Covariance Matrix is " S_cons1;
gdiag_cons11=diag(s_cons11);*diagnoal matrix with variance on the diagnoal;
gg_cons11=sqrt(inv(gdiag_cons11));*Now 1/Sqrt(Var(X_i)) on the diagonal;
corr11=gg_cons11*s_cons11*gg_cons11;
*print "THE CORRELATION MATRIX IS" corr11;
corr11_1_3=corr11[1,3];
corr11_2_3=corr11[2,3];
corr11_1_2=corr11[1,2];
corr11_1_4=corr11[1,4];
corr11_2_4=corr11[2,4];
corr11_3_4=corr11[3,4];
corr11_ave=(corr11_1_3+corr11_2_3+corr11_1_2+corr11_1_4+corr11_2_4+corr11_3_
4)/6;
comp11_relia=4*corr11_ave/(1+3*corr11_ave);
*print corr11_ave;
print comp11_relia;

/*Below is to calculate the variance of the composite score 1 in group 1*/
 XBAR_comp11 = X_comp11(|+,|)`/400;
*PRINT, "XBAR_comp11 = " XBAR_comp11;
SUMSQ_comp11=X_comp11`*X_comp11-(XBAR_comp11*XBAR_comp11`)#400;
S_comp11=SUMSQ_comp11/(400-1);
*PRINT , "The Variance-Covariance Matrix is " S_comp11;
variance_comp11=s_comp11[1,1];
print variance_comp11;

/*Below is to calculate the measurement error variance of composite 1 in group 1*/
error_comp11=variance_comp11*(1-comp11_relia);
print, "error variance of component 1 is"  error_comp11;
```

```
/*----------------------Calcualte error variance of composite 1 in group 2----------------------
---*/
USE group2;
READ all var {x1 x4 x7 x10} INTO X_cons21;
read all var{c21} into X_comp21; /*cons means construct, comp means composite*/
*PRINT "The Data Matrix is" X_cons1;
*print "the composite score matrix is" X_comp1;

 /*Below is to calculate the reliability of the composite score 1*/
/*# of rows is 400, # of columns for the construct is 4, fo the composite is 1*/
 XBAR_cons21 = X_cons21(|+,|)`/400;
*PRINT, "XBAR_cons21 = " XBAR_cons21;
SUMSQ_cons21=X_cons21`*X_cons21-(XBAR_cons21*XBAR_cons21`)#400;
S_cons21=SUMSQ_cons21/(400-1);
*PRINT , "The Variance-Covariance Matrix is " S_cons21;
gdiag_cons21=diag(s_cons21);*diagnoal matrix with variance on the diagnoal;
gg_cons21=sqrt(inv(gdiag_cons21));*Now 1/Sqrt(Var(X_i)) on the diagonal;
corr21=gg_cons21*s_cons21*gg_cons21;
*print "THE CORRELATION MATRIX IS" corr21;
corr21_1_3=corr21[1,3];
corr21_2_3=corr21[2,3];
corr21_1_2=corr21[1,2];
corr21_1_4=corr21[1,4];
corr21_2_4=corr21[2,4];
corr21_3_4=corr21[3,4];
corr21_ave=(corr21_1_3+corr21_2_3+corr21_1_2+corr21_1_4+corr21_2_4+corr21_3_
4)/6;
comp21_relia=4*corr21_ave/(1+3*corr21_ave);
*print corr21_ave;
print comp21_relia;

/*Below is to calculate the variance of the composite score 1 in group 2*/
 XBAR_comp21 = X_comp21(|+,|)`/400;
*PRINT, "XBAR_comp21 = " XBAR_comp21;
SUMSQ_comp21=X_comp21`*X_comp21-(XBAR_comp21*XBAR_comp21`)#400;
S_comp21=SUMSQ_comp21/(400-1);
*PRINT , "The Variance-Covariance Matrix is " S_comp21;
variance_comp21=s_comp21[1,1];
print variance_comp21;

/*Below is to calculate the measurement error variance of composite 1 in group 2*/
error_comp21=variance_comp21*(1-comp21_relia);
print, "error variance of component 1 is"  error_comp21;
```

```
/*--------------------Calcualte error variance of composite 2 in group 1------------------------
---*/

USE group1;
READ all var {x2 x5 x8 x11} INTO X_cons12;
read all var{c12} into X_comp12; /*cons means construct, comp means composite*/
* PRINT "The Data Matrix is" X_cons12;
* print "the composite score matrix is" X_comp12;

 /*Below is to calculate the reliability of the composite score 2*/
/*# of rows is 400, # of columns for the construct is 4, fo the composite is 2*/
 XBAR_cons12 = X_cons12(|+,|)`/400;
*PRINT, "XBAR_cons12 = " XBAR_cons12;
SUMSQ_cons12=X_cons12`*X_cons12-(XBAR_cons12*XBAR_cons12`)#400;
S_cons12=SUMSQ_cons12/(400-1);
*PRINT , "The Variance-Covariance Matrix is " S_cons12;
gdiag_cons12=diag(s_cons12);*diagnoal matrix with variance on the diagnoal;
gg_cons12=sqrt(inv(gdiag_cons12));*Now 1/Sqrt(Var(X_i)) on the diagonal;
corr12=gg_cons12*s_cons12*gg_cons12;
*print "THE CORRELATION MATRIX IS" corr12;
corr12_1_3=corr12[1,3];
corr12_2_3=corr12[2,3];
corr12_1_2=corr12[1,2];
corr12_1_4=corr12[1,4];
corr12_2_4=corr12[2,4];
corr12_3_4=corr12[3,4];
corr12_ave=(corr12_1_3+corr12_2_3+corr12_1_2+corr12_1_4+corr12_2_4+corr12_3_
4)/6;
comp12_relia=4*corr12_ave/(1+3*corr12_ave);
*print corr12_ave;
print comp12_relia;

/*Below is to calculate the variance of the composite score*/
 XBAR_comp12 = X_comp12(|+,|)`/400;
*PRINT, "XBAR_comp12 = " XBAR_comp12;
SUMSQ_comp12=X_comp12`*X_comp12-(XBAR_comp12*XBAR_comp12`)#400;
S_comp12=SUMSQ_comp12/(400-1);
*PRINT , "The Variance-Covariance Matrix is " S_comp12;
variance_comp12=s_comp12[1,1];
print variance_comp12;

/*Below is to calculate the measurement error variance*/
error_comp12=variance_comp12*(1-comp12_relia);
print, "error variance of component 2 is"  error_comp12;
```

```
/*------------------------Calcualte error variance of composite 2 in group 2-------------------
---*/

USE group2;
READ all var {x2 x5 x8 x11} INTO X_cons22;
read all var{c22} into X_comp22; /*cons means construct, comp means composite*/
* PRINT "The Data Matrix is" X_cons22;
* print "the composite score matrix is" X_comp22;

 /*Below is to calculate the reliability of the composite score 2*/
/*# of rows is 400, # of columns for the construct is 4, fo the composite is 2*/
 XBAR_cons22 = X_cons22(|+,|)`/400;
*PRINT, "XBAR_cons22 = " XBAR_cons22;
SUMSQ_cons22=X_cons22`*X_cons22-(XBAR_cons22*XBAR_cons22`)#400;
S_cons22=SUMSQ_cons22/(400-1);
*PRINT , "The Variance-Covariance Matrix is " S_cons22;
gdiag_cons22=diag(s_cons22);*diagnoal matrix with variance on the diagnoal;
gg_cons22=sqrt(inv(gdiag_cons22));*Now 1/Sqrt(Var(X_i)) on the diagonal;
corr22=gg_cons22*s_cons22*gg_cons22;
*print "THE CORRELATION MATRIX IS" corr22;
corr22_1_3=corr22[1,3];
corr22_2_3=corr22[2,3];
corr22_1_2=corr22[1,2];
corr22_1_4=corr22[1,4];
corr22_2_4=corr22[2,4];
corr22_3_4=corr22[3,4];
corr22_ave=(corr22_1_3+corr22_2_3+corr22_1_2+corr22_1_4+corr22_2_4+corr22_3_
4)/6;
comp22_relia=4*corr22_ave/(1+3*corr22_ave);
*print corr22_ave;
print comp22_relia;

/*Below is to calculate the variance of the composite score*/
 XBAR_comp22 = X_comp22(|+,|)`/400;
*PRINT, "XBAR_comp22 = " XBAR_comp22;
SUMSQ_comp22=X_comp22`*X_comp22-(XBAR_comp22*XBAR_comp22`)#400;
S_comp22=SUMSQ_comp22/(400-1);
*PRINT , "The Variance-Covariance Matrix is " S_comp22;
variance_comp22=s_comp22[1,1];
print variance_comp22;

/*Below is to calculate the measurement error variance*/
error_comp22=variance_comp22*(1-comp22_relia);
```

```
print, "error variance of component 2 is"  error_comp22;


/*--------------------------Calcualte error variance of composite 3 in group 1-------------------
---*/
USE group1;
READ all var {x3 x6 x9 x12} INTO X_cons13;
read all var{c13} into X_comp13; /*cons means construct, comp means composite*/
*PRINT "The Data Matrix is" X_cons13;
*print "the composite score matrix is" X_comp13;

 /*Below is to calculate the reliability of the composite score 3*/
/*# of rows is 400, # of columns for the construct is 4, for the composite is 3*/
 XBAR_cons13 = X_cons13(|+,|)`/400;
*PRINT, "XBAR_cons13 = " XBAR_cons13;
SUMSQ_cons13=X_cons13`*X_cons13-(XBAR_cons13*XBAR_cons13`)#400;
S_cons13=SUMSQ_cons13/(400-1);
*PRINT , "The Variance-Covariance Matrix is " S_cons13;
gdiag_cons13=diag(s_cons13);*diagnoal matrix with variance on the diagnoal;
gg_cons13=sqrt(inv(gdiag_cons13));*Now 1/Sqrt(Var(X_i)) on the diagonal;
corr13=gg_cons13*s_cons13*gg_cons13;
*print "THE CORRELATION MATRIX IS" corr13;
corr13_1_3=corr13[1,3];
corr13_2_3=corr13[2,3];
corr13_1_2=corr13[1,2];
corr13_1_4=corr13[1,4];
corr13_2_4=corr13[2,4];
corr13_3_4=corr13[3,4];
corr13_ave=(corr13_1_3+corr13_2_3+corr13_1_2+corr13_1_4+corr13_2_4+corr13_3_
4)/6;
comp13_relia=4*corr13_ave/(1+3*corr13_ave);
*print corr13_ave;
print comp13_relia;

/*Below is to calculate the variance of the composite score*/
 XBAR_comp13 = X_comp13(|+,|)`/400;
*PRINT, "XBAR_comp13 = " XBAR_comp13;
SUMSQ_comp13=X_comp13`*X_comp13-(XBAR_comp13*XBAR_comp13`)#400;
S_comp13=SUMSQ_comp13/(400-1);
*PRINT , "The Variance-Covariance Matrix is " S_comp13;
variance_comp13=s_comp13[1,1];
print variance_comp13;

/*Below is to calculate the measurement error variance*/
```

```
error_comp13=variance_comp13*(1-comp13_relia);
print, "error variance of component 3 is"  error_comp13;

/*--------------------Calcualte error variance of composite 3 in group 2------------------------
---*/
USE group2;
READ all var {x3 x6 x9 x12} INTO X_cons23;
read all var{c23} into X_comp23; /*cons means construct, comp means composite*/
*PRINT "The Data Matrix is" X_cons23;
*print "the composite score matrix is" X_comp23;

 /*Below is to calculate the reliability of the composite score 3*/
/*# of rows is 400, # of columns for the construct is 4, for the composite is 3*/
 XBAR_cons23 = X_cons23(|+,|)`/400;
*PRINT, "XBAR_cons23 = " XBAR_cons23;
SUMSQ_cons23=X_cons23`*X_cons23-(XBAR_cons23*XBAR_cons23`)#400;
S_cons23=SUMSQ_cons23/(400-1);
*PRINT , "The Variance-Covariance Matrix is " S_cons23;
gdiag_cons23=diag(s_cons23);*diagnoal matrix with variance on the diagnoal;
gg_cons23=sqrt(inv(gdiag_cons23));*Now 1/Sqrt(Var(X_i)) on the diagonal;
corr23=gg_cons23*s_cons23*gg_cons23;
*print "THE CORRELATION MATRIX IS" corr23;
corr23_1_3=corr23[1,3];
corr23_2_3=corr23[2,3];
corr23_1_2=corr23[1,2];
corr23_1_4=corr23[1,4];
corr23_2_4=corr23[2,4];
corr23_3_4=corr23[3,4];
corr23_ave=(corr23_1_3+corr23_2_3+corr23_1_2+corr23_1_4+corr23_2_4+corr23_3_
4)/6;
comp23_relia=4*corr23_ave/(1+3*corr23_ave);
*print corr23_ave;
print comp23_relia;

/*Below is to calculate the variance of the composite score*/
 XBAR_comp23 = X_comp23(|+,|)`/400;
*PRINT, "XBAR_comp23 = " XBAR_comp23;
SUMSQ_comp23=X_comp23`*X_comp23-(XBAR_comp23*XBAR_comp23`)#400;
S_comp23=SUMSQ_comp23/(400-1);
*PRINT , "The Variance-Covariance Matrix is " S_comp23;
variance_comp23=s_comp23[1,1];
print variance_comp23;

/*Below is to calculate the measurement error variance*/
```

```
error_comp23=variance_comp23*(1-comp23_relia);
print, "error variance of component 3 is"  error_comp23;

run;


/*--------------------------GENERATE MPLUS SYNTAX FILES-------------------------------
---*/

file "G:\DR addjustment\MCFA\S&m&n..inp";
put ("title: scenario ")@;put ("&m ")@;put ("sample ")@; put ("&n ")@;put ("MCFA
model");put;
put ("data: file=G:\DR\loading=.8\mean difference=.5\s")@; put("&m")@;
put("\s")@;put("&m")@; put ("&n")@;put (".dat;"); put;
put ("variable: ")@; put("names=x1-x3 ")@; put ("g;");
put ("usevariables=x1-x3;"); put;
put ("grouping is g (1=g1 2=g2);");put;
put ("analysis: type=mgroup meanstructure;"); put;
put ("model:");
put ("f ")@; put ("by ")@; put("x2@1 x1 x3;"); put;
put ("[f];"); put;
put ("model g1:"); put;
put ("f by x2@1 x1 x3;"); put;
put ("x1@")@; put (error_comp11)@; put (";");put;
put ("x2@")@; put (error_comp12)@; put (";");put;
put ("x3@")@; put (error_comp13)@; put (";");put;
put ("model g2:"); put;
put ("f by x2@1 x1 x3;"); put;
put ("x1@")@; put (error_comp21)@; put (";");put;
put ("x2@")@; put (error_comp22)@; put (";");put;
put ("x3@")@; put (error_comp23)@; put (";");put;


closefile "G:\DR addjustment\MCFA\S&m&n..inp";

/*------------------------------Call mplus to run the model-------------------------------------------
--*/

X call "C:\Program Files\Mplus\mplus.exe" "G:\DR addjustment\MCFA\S&m&n..inp"
"G:\DR addjustment\MCFA\S&m&n..out";

/*-------------------- STRIP p VALUES FROM MPLUS PRINTOUT ------------------------
--*/
```

```
data one;
infile "G:\DR addjustment\MCFA\S&m&n..out" truncover scanover flowover;
input @'Means' var1 $ var2 - var4 p;
keep p;
scenario = &m;
sample = &n;

/* WRITE POWER VALUES TO A TEXT FILE */
file "G:\DR addjustment\MCFA\p-value.dat" mod;
if p<999.000;
put @1 (scenario) (5.0) @29 (sample) (5.0) @20 (p) (5.3);

%end;
%end;
%mend;
%mplus;

run;
```

APPENDIX D

SAS CODE FOR RUNNING MIMIC MODELS WITH DOMAIN REPRESENTATIVE PARCELS WITH ERROR ADJUSTMENT

```
%macro mplus;

X mkdir "G:\DR addjustment\mimic";

%do i= 1 %to 12;
%do j = 1 %to 500;
%let m=&i;
%let n =&j;
%let file1=G:\second_order_factor\monte carlo\unparceled\loading=.8\mean
difference=.5\s;
%let file2 = \s;
%let file3 = .dat;

/*Import unparceled data*/
data temp;
infile "&file1&m&file2&m&n&file3";
input x1  1-12 x2 13-25 x3 26-38 x4 39-51 x5 52-64 x6 65-77 x7 78-90 x8 91-103 x9
104-116 x10 117-129 x11 130-142 x12 143-155 group $ 158;
c1=(x1+x4+x7+x10)/4;
c2=(x2+x5+x8+x11)/4;
c3=(x3+x6+x9+x12)/4;
run;

proc iml;
RESET NOLOG;
/*-------------------------------Calcualte error variance of composite 1--------------------------
---*/
USE temp;
READ all var {x1 x4 x7 x10} INTO X_cons1;
read all var{c1} into X_comp1; /*cons means construct, comp means composite*/
*PRINT "The Data Matrix is" X_cons1;
*print "the composite score matrix is" X_comp1;

 /*Below is to calculate the reliability of the composite score 1*/
/*# of rows is 800, # of columns for the construct is 4, fo the composite is 1*/
 XBAR_cons1 = X_cons1(|+,|)`/800;
```

```
*PRINT, "XBAR_cons1 = " XBAR_cons1;
SUMSQ_cons1=X_cons1`*X_cons1-(XBAR_cons1*XBAR_cons1`)#800;
S_cons1=SUMSQ_cons1/(800-1);
*PRINT , "The Variance-Covariance Matrix is " S_cons1;
gdiag_cons1=diag(s_cons1);*diagnoal matrix with variance on the diagnoal;
gg_cons1=sqrt(inv(gdiag_cons1));*Now 1/Sqrt(Var(X_i)) on the diagonal;
corr1=gg_cons1*s_cons1*gg_cons1;
*print "THE CORRELATION MATRIX IS" corr1;
corr1_1_3=corr1[1,3];
corr1_2_3=corr1[2,3];
corr1_1_2=corr1[1,2];
corr1_1_4=corr1[1,4];
corr1_2_4=corr1[2,4];
corr1_3_4=corr1[3,4];
corr1_ave=(corr1_1_3+corr1_2_3+corr1_1_2+corr1_1_4+corr1_2_4+corr1_3_4)/6;
comp1_relia=4*corr1_ave/(1+3*corr1_ave);
if comp1_relia < 0 then comp1_relia = 0;
*print corr1_ave;
print, "the reliability of composite 1 for scenario &m and sample &n is" comp1_relia;

/*Below is to calculate the variance of the composite score*/
 XBAR_comp1 = X_comp1(|+,|)`/800;
*PRINT, "XBAR_comp1 = " XBAR_comp1;
SUMSQ_comp1=X_comp1`*X_comp1-(XBAR_comp1*XBAR_comp1`)#800;
S_comp1=SUMSQ_comp1/(800-1);
*PRINT , "The Variance-Covariance Matrix is " S_comp1;
variance_comp1=s_comp1[1,1];
print variance_comp1;

/*Below is to calculate the measurement error variance*/
error_comp1=variance_comp1*(1-comp1_relia);
print, "error variance of component 1 is"  error_comp1;

/*------------------------------Calcualte error variance of composite 2--------------------------
---*/

USE temp;
READ all var {x2 x5 x8 x11} INTO X_cons2;
read all var{c2} into X_comp2; /*cons means construct, comp means composite*/
* PRINT "The Data Matrix is" X_cons2;
* print "the composite score matrix is" X_comp2;

 /*Below is to calculate the reliability of the composite score 2*/
/*# of rows is 800, # of columns for the construct is 4, fo the composite is 2*/
```

```
 XBAR_cons2 = X_cons2(|+,|)`/800;
*PRINT, "XBAR_cons2 = " XBAR_cons2;
SUMSQ_cons2=X_cons2`*X_cons2-(XBAR_cons2*XBAR_cons2`)#800;
S_cons2=SUMSQ_cons2/(800-1);
*PRINT , "The Variance-Covariance Matrix is " S_cons2;
gdiag_cons2=diag(s_cons2);*diagnoal matrix with variance on the diagnoal;
gg_cons2=sqrt(inv(gdiag_cons2));*Now 1/Sqrt(Var(X_i)) on the diagonal;
corr2=gg_cons2*s_cons2*gg_cons2;
*print "THE CORRELATION MATRIX IS" corr2;
corr2_1_3=corr2[1,3];
corr2_2_3=corr2[2,3];
corr2_1_2=corr2[1,2];
corr2_1_4=corr2[1,4];
corr2_2_4=corr2[2,4];
corr2_3_4=corr2[3,4];
corr2_ave=(corr2_1_3+corr2_2_3+corr2_1_2+corr2_1_4+corr2_2_4+corr2_3_4)/6;
comp2_relia=4*corr2_ave/(1+3*corr2_ave);
if comp2_relia < 0 then comp2_relia = 0;
*print corr2_ave;
print, "the reliability of composite 2 for scenario &m and sample &n is" comp2_relia;

/*Below is to calculate the variance of the composite score*/
 XBAR_comp2 = X_comp2(|+,|)`/800;
*PRINT, "XBAR_comp2 = " XBAR_comp2;
SUMSQ_comp2=X_comp2`*X_comp2-(XBAR_comp2*XBAR_comp2`)#800;
S_comp2=SUMSQ_comp2/(800-1);
*PRINT , "The Variance-Covariance Matrix is " S_comp2;
variance_comp2=s_comp2[1,1];
print variance_comp2;

/*Below is to calculate the measurement error variance*/
error_comp2=variance_comp2*(1-comp2_relia);
print, "error variance of component 2 is"  error_comp2;

/*-------------------------------Calcualte error variance of composite 3--------------------------
---*/
USE temp;
READ all var {x3 x6 x9 x12} INTO X_cons3;
read all var{c3} into X_comp3; /*cons means construct, comp means composite*/
*PRINT "The Data Matrix is" X_cons3;
*print "the composite score matrix is" X_comp3;

 /*Below is to calculate the reliability of the composite score 3*/
/*# of rows is 800, # of columns for the construct is 4, for the composite is 3*/
```

```
 XBAR_cons3 = X_cons3(|+,|)`/800;
*PRINT, "XBAR_cons3 = " XBAR_cons3;
SUMSQ_cons3=X_cons3`*X_cons3-(XBAR_cons3*XBAR_cons3`)#800;
S_cons3=SUMSQ_cons3/(800-1);
*PRINT , "The Variance-Covariance Matrix is " S_cons3;
gdiag_cons3=diag(s_cons3);*diagnoal matrix with variance on the diagnoal;
gg_cons3=sqrt(inv(gdiag_cons3));*Now 1/Sqrt(Var(X_i)) on the diagonal;
corr3=gg_cons3*s_cons3*gg_cons3;
*print "THE CORRELATION MATRIX IS" corr3;
corr3_1_3=corr3[1,3];
corr3_2_3=corr3[2,3];
corr3_1_2=corr3[1,2];
corr3_1_4=corr3[1,4];
corr3_2_4=corr3[2,4];
corr3_3_4=corr3[3,4];
corr3_ave=(corr3_1_3+corr3_2_3+corr3_1_2+corr3_1_4+corr3_2_4+corr3_3_4)/6;
comp3_relia=4*corr3_ave/(1+3*corr3_ave);
if comp3_relia < 0 then comp3_relia = 0;
*print corr3_ave;
print, "the reliability of composite 3 for scenario &m and sample &n is" comp3_relia;

/*Below is to calculate the variance of the composite score*/
 XBAR_comp3 = X_comp3(|+,|)`/800;
*PRINT, "XBAR_comp3 = " XBAR_comp3;
SUMSQ_comp3=X_comp3`*X_comp3-(XBAR_comp3*XBAR_comp3`)#800;
S_comp3=SUMSQ_comp3/(800-1);
*PRINT , "The Variance-Covariance Matrix is " S_comp3;
variance_comp3=s_comp3[1,1];
print variance_comp3;

/*Below is to calculate the measurement error variance*/
error_comp3=variance_comp3*(1-comp3_relia);
print, "error variance of component 3 is"  error_comp3;
run;

/*--------------------------------GENERATE MPLUS SYNTAX FILES--------------------------
---*/

file "G:\DR addjustment\mimic\S&m&n..inp";
put ("title: scenario ")@;put ("&m ")@;put ("sample ")@; put ("&n ")@;put ("mimic
model");put;
put ("data: file=G:\DR\loading=.8\mean difference=.5\s")@; put("&m")@;
put("\s")@;put("&m")@; put ("&n")@;put (".dat;"); put;
put ("variable:")@; put("names=x1-x3 ")@; put ("g;");
```

```
put ("usevariables=x1-x3 ")@; put ("g;");put;
put ("model:");
put ("f ")@; put ("by ")@; put("x1-x3;");
put ("x1@")@; put (error_comp1)@; put (";");put;
put ("x2@")@; put (error_comp2)@; put (";");put;
put ("x3@")@; put (error_comp3)@;put (";");put;
put ("f ")@; put ("on ")@; put("g;");
closefile "G:\DR addjustment\mimic\S&m&n..inp";

/*------------------------------------Call mplus to run the model----------------------------------
---*/

X call "C:\Program Files\Mplus\mplus.exe" "G:\DR addjustment\mimic\S&m&n..inp"
"G:\DR addjustment\mimic\S&m&n..out";

/*----------------------- STRIP p VALUES FROM MPLUS PRINTOUT ---------------------
----*/
data one;
infile "G:\DR addjustment\mimic\S&m&n..out" truncover scanover flowover;
input @'F       ON' var1 $ var2 - var4 p;
keep p;
scenario = &m;
sample = &n;

/* WRITE POWER VALUES TO A TEXT FILE */
file "G:\DR addjustment\mimic\p-value.dat" mod;
put @1 (scenario) (5.0) @9 (sample) (5.0) @20 (p) (5.3);

%end;
%end;

%mend;
%mplus;
run;
```

VITA

Name:	Yuanyuan Zou
Address:	Department of Educational Psychology, Texas A&M University, MS 4225 c/o Dr. Oi-man Kwok College Station, TX 77843-4225
Email Address:	malenazou@tamu.com
Education:	B.A., Earth Sciences, Nanjing University, China, 2000 M.S., Geology, Baylor University, 2005 Ph.D., Educational Psychology, Texas A&M University, 2009