The Impact of Homogeneous and Heterogeneous Parceling Strategies on Hierarchical Multidimensional Models

By

Corinne Elizabeth Perkins

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

December, 2014

Nashville, Tennessee

Approved:

David A. Cole, Ph.D.

Kristopher J. Preacher, Ph.D.

Sonya K. Sterba, Ph.D.

Andrew Tomarken, Ph.D.

To my brilliant and adorably strong willed daughter, Aurora Rose

and

To my beloved husband, Perky, who is infinitely supportive. Chi!

ACKNOWLEDGEMENTS

Foremost, I would like to thank my advisor, Professor David Cole, for his infectious enthusiasm and the tremendous knowledge he shares with his students. His extraordinary patience and sense of humor made frustrating obstacles fun learning experiences. I could not have navigated my way through the graduate school milestones without him.

I would also like to thank my committee members: Professor Kristopher Preacher, Professor Andrew Tomarken, and Dr. Sonya Sterba for their encouragement, insightful comments, and tough questions.

On a more personal note, I would like to thank my family for the encouragement and support through this process. I would particularly like to thank my Mom who has always been my favorite confidant. Her support is unconditional and her insight is invaluable.

Thank you to all my Vanderbilt friends who are now a part of my family. Particularly the Quant lunch gang: Dr. Michele Tomlinson, Dr. Laura Williams, Dr. Andriy Koval, Dr. Stijn Smeets, and least of all, Dr. Michael Nelson. You all shared your quantitative wisdom with me every week and offered equally horrifying graduate school stories to let me know that I was not alone. I love you all; and I cannot wait to see all the amazing things you accomplish in the future.

I would also like to thank my oldest and dearest friends. Christina Zona, Jamie Chou, and Roseline Ojo you are constant bright lights in my life. I am so grateful that we have been able to share so much of our lives with each other.

Last but certainly not least, I am truly indebted to my husband, Perky, and my daughter, Aurora. You both inspire me to do better, be better, and go after my goals. You make me laugh and smile at times when it seems impossible. You are the most important things in my life and I owe my successes to your unbelievable support and love. Chi!

Thank you – I love you all.

TABLE OF CONTENTS

Appendix

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

χ^2	Chi-squared
γ	Gamma
λ	Lambda
ω^2	Omega-squared
ANOVA	Analysis of Variance
CES-D	Center for Epidemiological Studies-Depression
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
df	Degrees of Freedom
DG	Data Generative
ESEM	Exploratory Structural Equation Models
Het	Heterogeneous
Hom	Homogeneous
LV	Latent Variable
M	Mean
MB	Mean Bias
MV	Manifest Variable
NNFI	Non-Normed Fit Index
PPEB	Percentage of Parameter Estimate Bias
RMSEA	Root Mean Square Error of Approximation
SD	Standard Deviation
SE	Standard Error
SEM	Structural Equation Modeling
SRMR	Standardized Root Mean Square Residual
TLI	Tucker-Lewis Index

CHAPTER I

Introduction

A *parcel* is “an aggregate-level indicator comprised of the sum (or average) of two or more items, responses, or behaviors” (Little, Cunningham, Shahar, & Widaman, 2002, p. 152), which can then be used as an indicator of a factor in latent variable analyses. This simple-sounding strategy has sparked debate over the validity of models that use a parceling strategy since the first discussion by Cattell in 1956. Since then, empirical and simulation studies have contributed to the debate by reporting (often conflicting) results about the convergence rates, goodness-of-fit indices, and parameter estimates when parceling strategies were compared to total-score models and models using item-level indicators. Other studies have compared specific types of the parceling techniques. Of particular interest are studies that suggest homogeneous parceling methods (combining items that have shared specific variance) and heterogeneous parceling methods (combining items that do not share specific variance) can sometimes produce very different substantive conclusions. Little et al. (2002) contend that some of these discrepancies are due in part to differences in the underlying model that each study used for analysis; specifically, whether the researcher focused on a measurement or structural model. The overall message from the literature has been two-fold: (1) the nature of the researcher’s question is a key factor when deciding to use parcels, and (2) parceling multidimensional constructs can yield different results depending on the parceling strategy used and the true model. The current study explores factors that affect performance of homogeneous versus heterogeneous parceling strategies when examining a multidimensional latent variable (LV) in the context of a simple structural model. In particular, I will examine the effect of the strength of factor loadings (from the higher-order factor to three sub-factors) on model convergence rates, bias in the estimate of the structural parameter, and model fit indices. I will further discuss how the research question ultimately impacts which parceling strategy is more appropriate.

The Importance of the Research Question

The researcher’s question is the first consideration when deciding whether to parcel, and arguably how to parcel. Little et al. (2002) offered an insightful view of the paradoxical findings from parceling studies. They attributed the discrepancies to researchers’ differing fundamental philosophical perspectives about measurement and also to the research questions being asked. Little et al. (2002) described many of the supporters of parcels as taking a “pragmatic-liberal philosophical perspective,” which acknowledges that measurement is a somewhat capricious process that is intrinsically flawed and ultimately defined by the researcher. In contrast, they described critics of parceling as often taking an "empiricist-conservative philosophy of science perspective," which assumes that data should remain in a state as close to the original data collected as possible so that all sources of variance can be analyzed. In addition to these differing philosophical perspectives, Little et al. (2002) suggested that the researcher's question should be the determining factor in whether parcels can be utilized in a study. On the one hand, researchers who are especially interested in assessing the structural paths between latent variables that pertain to a substantive theory should consider using parcels in their analysis. On the other hand, parceling is inadvisable when the researcher is interested in the measurement characteristics of individual items or the structure of the measure itself, especially when the measures are multidimensional.

Simple reasoning dictates that if the researcher is interested in the measurement characteristics of individual items, then parceling is inadvisable, as the integrity of the item is lost when it is combined with additional items. Furthermore, although less obvious, researchers interested in the factor structure of a measure should avoid using parcels due to the fact that parceling can obscure multidimensionality (Little et al., 2002). Commonly, items representing a multidimensional construct will tap into the subordinate constructs (also called *subconstructs* or *lower-order constructs*) that compose the superordinate constructs (also called *higher-order constructs*). Therefore, the factor structure that derives from various parceling strategies will vary depending on which items are combined.

Two parceling strategies are commonly used for multidimensional constructs: *homogeneous parceling* and *heterogeneous parceling*. Homogeneous parcels (also called *facet-representative parcels*) consist of items presumed to represent the same subconstruct. When parcels are homogenous, they constitute manifest representations of the latent subconstruct. Furthermore, a group of homogeneous parcels will share only true score variance with each other (Gibson, 2012; Little, Rhemtulla, Gibson, & Schoemann, 2013). In contrast, heterogeneous parcels (also called *domain-representative parcels*) consist of items that represent each of the subconstructs (ideally, each construct is evenly represented within each parcel). Thus, the variance of the subconstructs is pooled when heterogeneous parceling is utilized, meaning that a group of heterogeneous parcels will not only share true score variance, they will also share specific variance. Conceptually, heterogeneous parcels represent the higher-order construct and mask the existence of sub-factors. Compounding this issue, heterogeneous parceling can not only change the factor structure by eliminating the sub-factors, they can also improve model fit compared to item-level indicators, in part because a more parsimonious model is being estimated (Bandalos & Finney, 2001). Thus, parceling may yield highly compelling support for models that do not accurately represent complex factor structures. In contrast, models with item-level indicators often yield poorer model fits due to the number of parameters, increased model restraints, and the existence of unanticipated local covariances. Such models may be rejected for relatively unimportant reasons. In the end, the literature overwhelmingly advises researchers interested in retaining or exploring the original factor structure to refrain from using parcels (Marsh, Lüdtke, Nagengast, Morin, & Von Davier, 2013).

Many researchers who are uninterested in factor structure but are primarily focused on the structural relations among the higher-order constructs will forsake LV models altogether. They combine items that putatively represent the same construct and test manifest variable (MV) models. Such approaches, of course, retain measurement error in the variables that compose the model. Manifest variable path analyses are still very common (Bollen, 1989; Cole & Preacher, 2014; James, 1982), despite the well-known problems with parameter estimation and model fit that plague this approach. Latent variable approaches can help researchers avoid these problems. Despite the known benefits of analyzing an LV model over an MV path analysis (Coffman & MacCallum, 2005; MacCallum & Austin, 2000; Marsh, Hau, Balla, & Grayson, 1998), logistical problems, financial issues, and even lack of forethought often put researchers in the less-than-ideal position of using a single measure to represent a construct. Parceling represents one approach when researchers have only one measure representing a construct and they wish to ue LV modeling. Surprisingly, however, a literature review by Coffman and MacCallum (2005) noted that over 50% of the articles they reviewed used a MV path analysis, with total-score indicators, rather than an LV structural model. The goal of the current paper is to clarify conditions under which parceling may be an appropriate strategy, even when the underlying variables are multidimensional.

To summarize, researchers have two LV options: use every item as an indicator of the LV or use parcels. Researchers interested in assessing measurement structure should never use parcels; however, for those interested in the structural associations between constructs, parceling may be a viable option. Such situations become complicated when one or more of the underlying LVs is multidimensional. Parceling in the context of multidimensionality requires deciding between homogeneous or heterogeneous parceling strategies. The parceling strategy decision should derive from the researcher's question. On the one hand, if researchers are interested in a relation between a sub-factor and another LV, then then they should use homogeneous parceling to retain the factor structure. On the other hand, if researchers are interested in the relation among higher-order factors (and are not interested in the sub-factors), then heterogeneous parceling might be the better option. The researcher's question should be the primary concern when deciding whether to use parceling in a given study.

The Importance of the Construct Dimensionality

Researchers have recently begun to acknowledge the importance of construct dimensionality in the decision to parcel. Some of the biggest critics of parceling procedures now concede that parceling can be a valid strategy, if and only if the items used to represent the construct are unidimensional (Bagozzi & Heatherton, 1994; Bandalos & Finney, 2001; cf Little et al., 2002). Parcels offer several advantages over item-level indictors. These advantages can be categorized into two areas: *psychometric characteristics of the indictors* and *model fit*. The most commonly cited psychometric advantages of parceling a unidimensional construct, when compared to item-level indicators, are greater reliability, greater communality, and better representation of the construct (see Bagozzi & Edwards, 1998; Coffman & MacCallum, 2005; Kishton & Widaman, 1994; Rushton, Brainerd, & Pressley, 1983). Furthermore, parcels have a smaller likelihood of having distributional violations; that is, they tend to have smaller and more evenly distributed scale intervals (Coffman & MacCallum, 2005; Little et al., 2002). Parcels are also credited with the "advantage" of achieving better model fit and less biased parameter estimates. Improved model fit derives from reduced chances for correlated error terms or dual factor loadings when the covariance matrix shrinks as multiple items combine to make a single parcel (see Bagozzi & Edwards, 1998; Bagozzi & Heatherton, 1994; Gribbons & Hocevar, 1998; Hau & Marsh, 2004), also resulting in simpler model interpretation. Note that simply achieving better model fit is not an advantage, especially if the model ought to be rejected for important misspecifications. Better model fit becomes an advantage only if the model would have otherwise been rejected due to minor, unaccounted-for covariances. Models utilizing parcels have also been shown to have less biased parameter estimates and greater ratios of common-to-unique factor variance (at least when compared to models with item-level indicators; see Bagozzi & Edwards, 1998; Bandalos, 2002; Bandalos & Finney, 2001; Gribbons & Hocevar, 1998).

Researchers have long believed that when items are congeneric, tapping a unidimensional construct, the effects of parceling on parameter estimates and model fit would be relatively small, regardless of which parceling strategy is used (e.g., random assignment, shown in Bandalos, 2002; Hau & Marsh, 2004; Marsh et al., 1998). Recently, however, Sterba and MacCallum (2010) showed that although that belief holds true in the population, item-to-parcel allocation can alter results quite significantly when the sampling error is large (due to small sample size, few items per parcel, or small item communalities). Thus Sterba and MacCallum (2010) suggested that researchers using parcels offer confidence intervals around the parameter estimates and model fit indices when the model indicators are parcels and the sampling error is known to be high.

The idea that item-to-parcel reallocation can change key parameter estimates is troublesome enough, but the problem is magnified when the item pool is multidimensional. Prior to Sterba and MacCallum's (2010) surprising finding about unidimensional items, researchers were well aware of the inconsistent results for model fit and parameter estimates across various parceling techniques of a multidimensional factor. Hence, some researchers (e.g., Bagozzi & Edwards, 1998) warned against parceling any measure that has not consistently shown unidimensionality. Other researchers, however, argued that when the structure of the multidimensional measure is known, there are situations when parceling offers equivalent or less biased parameter estimates and equivalent or better model fit, at least compared to models using total-score and item-level indicators (e.g., Coffman & MacCallum, 2005).

Very little research has compared homogeneous and heterogeneous parceling strategies. Kishton and Widaman (1994) were the first researchers to compare homogeneous and heterogeneous parceling strategies (they referred to them as "internally consistent/unidimensional" and "domain-representative" parcels). Their study showed that these parceling strategies can yield different results. Since then, only four simulation studies have (1) analyzed a model that estimated structural paths, (2) used non-unidimensional measures, and (3) used both homogeneous and heterogeneous parceling strategies. Models and results of the four simulation studies are discussed below.

First, Hall et al. (1999) showed that homogeneous and heterogeneous parceling strategies can yield significantly different parameter estimates and fit indices when a multidimensional measure is misspecified as a unidimensional factor. They tested this claim by creating a data generative (DG) model with one exogenous LV (estimated from six items) predicting one endogenous LV. Two of the six items that loaded onto the exogenous LV also loaded onto a secondary factor (see Appendix A Figure A1). They estimated two models: one using homogenous parceling (which put the two items with dual loadings into the same parcel) and a second model using heterogeneous parceling (where the two items with dual loadings were put into separate parcels). Neither model included the secondary factor. The model with homogenous parcels yielded unbiased estimates; however, heterogeneous parceling generated biased factor loadings for all of the parcels and the structural path. Hall et al. (1999) then created a second DG model where the secondary factor predicted the endogenous variable. They estimated models using the same homogeneous and heterogeneous parceling models, which ignored the secondary factor and the additional structural path. Again, heterogeneous parceling led to biased estimates of all factor loadings and structural paths. This time, however, homogeneous parceling also yielded biased estimates of measurement and structural parameters. Despite these problems, heterogeneous parceling yielded better model fits than did the homogenous models, which on average had less biased path estimates.

Second, Bandalos (2002) expanded upon Hall et al.'s (1999) work by having a slightly more complicated structural model where two exogenous LVs (not just one) predicted the endogenous LV. Furthermore, items from both exogenous LVs loaded onto the unaccounted for secondary factor (see Appendix A Figure A2). She also increased the percentage of items with dual factor loadings from 33.3% to 50% (from two of six items to six of twelve items, respectively). Bandalos concluded that there was no difference in parameter bias between homogenous and heterogeneous parceling strategies for the main structural paths; however, the homogeneous parceling technique resulted in less bias for the exogenous factor correlation. Furthermore, when comparing parceled models to models using item-level indicators, she stressed that although there were no differences between the models in parameter estimates, the model fit (measured by RMSEA and CFI) of the parceled model was "deceptively good" and that using parcels "can obfuscate the true factor structure of a set of items" (p. 99).

Third, Marsh et al. (2013) wrote a compelling paper about the problems of parceling when there are cross-loadings. Through multiple empirical and simulation studies, they sought to show that exploratory structural equation models (ESEM) typically yield less biased parameter estimates than a confirmatory factor analysis (CFA) using either items or parcels as manifest variables. Although their goal was not to compare homogeneous and heterogeneous parceling strategies to each other, they did reemphasize the dangers of using a parceling strategy when the construct is multidimensional due to cross-loadings.

Fourth, Coffman and MacCallum (2005) conducted a two-part study where they estimated total-score, reliability-corrected, homogeneous parceling and heterogeneous parceling models. They analyzed artificial data from a simulation and empirical data collected from a national survey. Although they used homogeneous and heterogeneous parceling for each of the higher-order factors their goal was to compare total-score and reliability corrected models to the parceling strategies, not to compare the parceling strategies themselves. Thus, differences between the parceling techniques were not discussed in detail. Close examination of the results, however, reveals an interesting disparity between the simulation results and empirical study results.

Coffman and MacCallum's (2005) DG model included four higher-order factors, each with three sub-factors that, in turn, were represented by three MVs. The model was constructed so that two of the higher-order factors were correlated with each other and predicted a third higher-order factor, which then predicted a fourth higher-order factor (see Appendix A Figure A3). Their empirical model, on the other hand, had only three higher-order factors, each with only two sub-factors. They tested two empirical models: (1) two higher-order factors were correlated and predicted a third higher-order factor and (2) a mediation model where one higher-order factor predicted a second factor, which then predicted the third higher-order factor. Unfortunately, the items in the empirical example were not evenly distributed among the sub-factors. Two of the sub-factors had only two items (i.e. positively worded mastery items and emotion/feeling CES-D items), whereas one of the sub-factors had seven items (i.e., negatively worded self-esteem items; see Appendix A Figure A4 for a path diagram of the empirical model with two exogenous variables). This imbalance made the even distribution of sub-factor items to heterogeneous parcels impossible, thus causing some of the parcels in the "heterogeneous model" to become homogeneous.

The results of the simulation study differed from those of the empirical study. The simulation study, which did not include any dual factor loadings and included all of the LVs in the estimation model, revealed that the homogeneous parceling model resulted in less biased parameter estimates than did the heterogeneous parceling model. Furthermore, although both parceling models provided extremely good fit indices, the fit for the homogeneous parceling strategy was just slightly better. The empirical study, with all its real world idiosyncrasies and imperfect representation of heterogeneous parceling, presented a completely different picture than the simulation study. For the empirical study, neither of the homogeneous parceling models in either the dual exogenous model or the mediation model had a χ^2 statistic or RMSEA that indicated an acceptable fit. Both heterogeneous parceling models, however, fit the data well according to both the χ^2 and RMSEA fit indices. Despite the good fit of the heterogeneous model, it is impossible to compare the parameter estimates from either model to the "true" parameters, leaving open the question as to which approach yielded more accurate results.

There are several possible reasons why the homogeneous parceling strategy worked in the simulation model and the heterogeneous parceling strategy worked better with the empirical data. One thing is certain; the reason homogeneous parceling of the empirical data did not meet the standards for model fit was because paths that needed to be estimated were left out of the model. That is, more paths were needed in the homogeneous model to recreate the variance-covariance matrix than were needed in the heterogeneous parceling model. Furthermore, the heterogeneous parcels were probably more similar to one another, thus increasing parcel communality. In contrast, the homogeneous parcels likely had greater internal consistency; however, they were probably less similar to one another, had lower communalities, and had greater unique variance.

The take-away message from these four simulation studies is that researchers interested in parceling multidimensional constructs must first understand the structure of the construct prior to parceling (e.g., Bagozzi & Edwards, 1998; Bagozzi & Heatherton, 1994; Coffman & MacCallum, 2005; Gribbons & Hocevar, 1998; Hall et al., 1999; Kishton & Widaman, 1994; Marsh et al., 2013). The three studies that failed to account for a secondary factor or dual loadings yielded poor results when parceling; however, the simulated model that accounted for the multidimensional structure of the LVs before parceling yielded more accurate path estimates and better model fit when utilizing a parceling strategy. Still, Coffman and MacCallum's (2005) study showed that even knowing the factor structure of a multidimensional measure does not ensure that all parceling strategies will yield the same results.

Another important point is that although most of these studies acknowledge that the LV extracted from homogeneous and heterogeneous parceling strategies are qualitatively different, they unwisely compare the resultant factor loadings and structural paths of both parceling strategies to each other. Furthermore, they often conclude that the paths in homogeneous parceling models are less biased than paths estimated in heterogeneous parceling models. If the LV is qualitatively different, then it is inappropriate to assume that the factor loadings in homogeneous and heterogeneous parceling models should be equivalent; similarly it is erroneous to believe that the structural paths in both parceling models should be identical. To emphasize this point refer to the hierarchical factor depicted in Figure 1. Depression is a higher-order factor extracted from three subfactors (Affect, Behavior, and Cognition), which in turn are each extracted from three items.

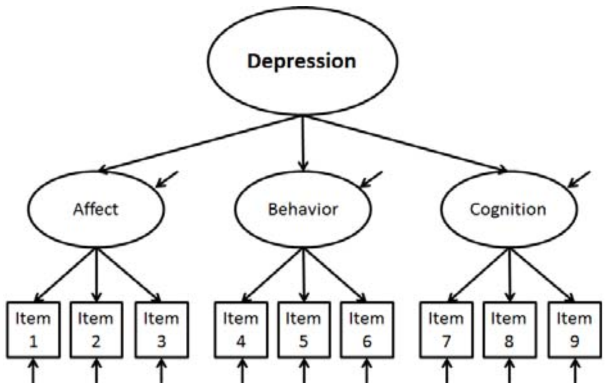


Figure 1. Example of a hierarchical factor.

Parceling this hierarchical factor homogeneously would involve averaging (or adding) all of the items that represent the Affect construct (items 1, 2, and 3). Then averaging the items that represent the construct of Behavior (items 4, 5, and 6); and finally, averaging the items representing Cognition (item 7, 8, and 9). Notice that each of these parcels is a manifest representation of the specific sub-factor the items loaded onto. The variance of each of these homogeneous parcels would include the True Score variance, the Specific Variance associated with the particular parcel (Affect, Behavior, or Cognition), and the sum of one-ninth the variance associated with each of the items individual error. For instance the "Behavior" parcel, comprised of items 4, 5 and 6, would have True Score variance, Specific-Behavior variance, and 1/9 error variance associated with item 4, 1/9 the error variance associated item 5, and 1/9 the error variance associated with item 6 (see Gibson, 2012 for the algebraic proof). It follows that the variance of the LV extracted from these three homogeneous parcels (LV_{Hom}) would be comprised of only the True Score variance. On the other hand, heterogeneous parcels would involve averaging items from different subfactors. For example, items 1, 4, and 7 could be averaged to create a heterogeneous parcel. Conceptually, heterogeneous parcels are all manifest representations of the higher-order factor Depression. Furthermore, the variance of each of the heterogeneous parcels includes True Score variance, 1/9 the Specific Variance in each of the items (Affect, Behavior, and Cognition), and 1/9 the error variance of each of the items. A LV extracted from three heterogeneous (LV_{Het}) would therefore include both the True Score variance and the sum of 1/9 the Specific Variance that was in each of the items.

In summary, homogeneous parcels share only True Score variance whereas the heterogeneous parcels share True Score variance and Specific Variance. Therefore, the covariance matrix associated with homogeneous parcels will be smaller than the covariance matrix of heterogeneous parcels. Consequently, the homogeneous parcel factor loadings should be smaller than heterogeneous parcel factor loadings. It also follows that because there is proportionally more variance in LV_{Het} (than LV_{Hom}), structural paths pointing away from LV_{Het} should be smaller. For these reasons it is inadvisable to compare the results of the homogeneous and heterogeneous parceling to each other. Researchers who choose to use parcels, specifically homogeneous or heterogeneous parcels, should choose the parceling strategy based on theoretical grounds and not the results of simulation studies. A researcher who wishes to extract only the True Variance of the factor should use homogeneous parceling. On the other hand, researchers wish to extract True Score variance and Specific variance should use heterogeneous parceling. This study is meant to inform the researcher about the possible consequences of their choice, once a parceling strategy has been chosen a priori.

Overview of the Current Study

The paucity of literature directly comparing homogeneous and heterogeneous parceling strategies is troublesome. The few studies that do compare parceling strategies often use estimation models that fail to include important LVs or structural paths; or in the case of Hall et al. (1999) both. Whereas Marsh et al. (2013) focused on method variance to create homogeneous and heterogeneous parcels, they ignored the cross-loadings of items onto secondary exogenous LVs. Furthermore, they stopped short of examining a multidimensional structural model that exhibits multidimensionality as higher-order LVs rather than bothersome cross-loadings (see Appendix A Figure A5 & Table A1). Only the Coffman and MacCallum (2005) simulation study actually compared homogeneous and heterogeneous parceling under conditions in which the literature suggests that multidimensional parceling strategies should be used: that is (1) when the researcher is interested in the structural paths and (2) when the multidimensional structure is known. Noting the relative lack of attention to this issue, Little et al. (2013) recently concluded that more research is needed to distinguish conditions under which homogeneous parceling is preferred over heterogeneous parceling techniques. The current study will add to the multidimensional parceling literature by exploring whether the strength of higher-order factor loadings (of the sub-factors and higher-order factor of the exogenous LV) influences which parceling strategy is preferable. I hypothesize that the strength of higher-order factor loadings will differentially affect model convergence rates, bias in estimate of the structural parameter, and model goodness-of-fit indices, depending on which parceling strategy was used to estimate the model. Furthermore, I will discuss how the choice of parceling strategy changes the nature of the underlying LV and therefore the interpretation of the surrounding path coefficients.

Data Generative Models. The study will involve nine data generative (DG) models that vary only in terms of the factor loadings from the higher-order factor to the sub-factors, labeled λ_1, λ_2, and λ_3 in Figure 1 (called "higher-order factor loadings"). The remaining structural paths and coefficients of the model will remain constant (only the unique variance associated with sub-factors A, B, and C will change as a function of the higher-order factor loadings). Within each DG model the higher-order factor loadings (λ_1, λ_2, and λ_3) will be equal to each other; but across DG models these standardized loadings will vary from 0.1 to 0.9 (specific values: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9). These values will allow for the easy identification of trends.

The structures of the nine DG models will be identical to each other. This structure is best understood by referring to Figure 1; however it will also be described below. Nine items load onto 3 sub-factors (3 items per sub-factor, no dual loadings) and the three sub-factors load onto a higher-order factor. The lower-order factor loadings (of the items onto the sub-factor) will be 0.4. The multidimensional higher-order factor will have a single predictive structural path (γ, gamma) to an endogenous manifest variable. Gamma will be 0.25 (standardized) for all of the DG models (models generated to determine Type 1 error had gammas of 0).

Estimated Models. Each of the nine DG models will be estimated twice per generated data set, once using homogeneous parcels and once using heterogeneous parcels. The homogeneous parceling models used for estimation will be structured in the following manner: three parcels (P_{123}, P_{456}, and P_{789}) will load onto a factor (D_{Hom}), which in turn will predict Y, as shown in

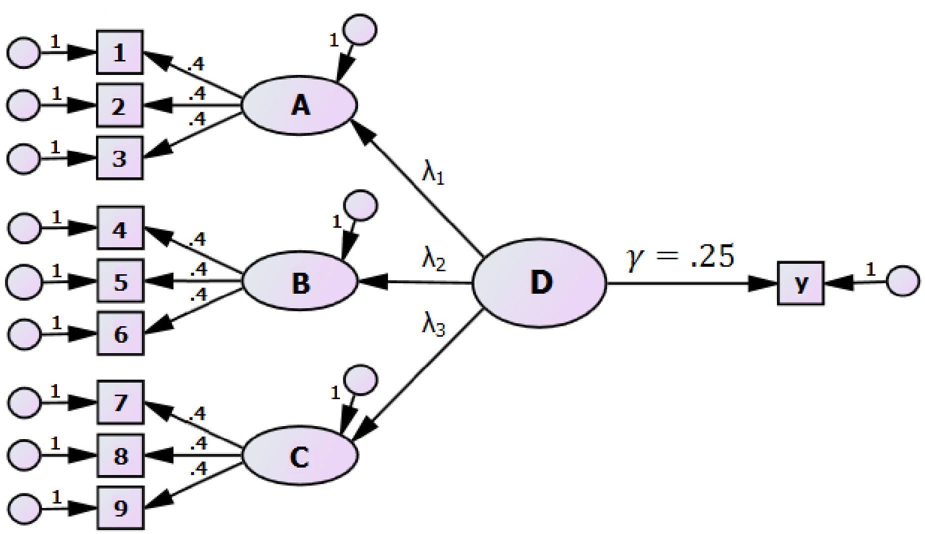


Figure 2. Data generative model. The values shown in this figure will remain the same for all nine DG models. Factor loadings $\boldsymbol{\lambda_1}$, $\boldsymbol{\lambda_2}$, and $\boldsymbol{\lambda_3}$ will be equal in every DG model, but will vary between 0.1 and 0.9 across DG models.

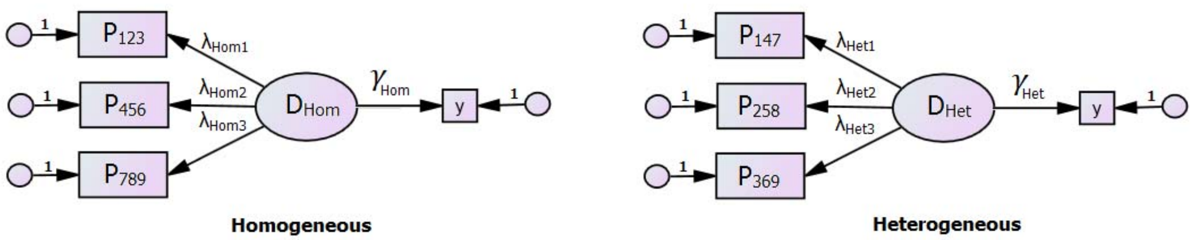


Figure 3. Estimated models. The manifest variables that load onto the exogenous LVs represent parcels that were created by adding the items denoted in the subscript.

Figure 2. Each homogeneous parcel will be the summation of the three items under each sub-factor. For heterogeneous estimated models, one item from each sub-factor will be combined to create a parcel (e.g., P_{147} will comprise items 1, 4, and 7 from Figure 1). In homogeneous parceling, each homogeneous parcel is a representation of a particular sub-factor; in heterogeneous parceling, however, each parcel is a representation of the higher-order factor.

Model convergence rates. Identifying conditions that contribute to poor model convergence rates is important because it prevents researchers from investing money and time in studies that are likely to generate results that are not interpretable. Out of the four simulations that compared homogeneous and heterogeneous parceling strategies only Bandalos' (2002) paper discussed convergence rates. In her study only one out of 750 models failed to converge. The model that failed to converge utilized heterogeneous parcels comprised of three items and had a sample of only 100 people. Unfortunately, Bandalos did not represent all of the factors in the DG model in the estimation models. Because parceling should be used only when the structure of the LV is known, Bandalos' (2002) results do not directly pertain to the current study.

If we take a step back from parceling studies and look more broadly at the structural equation modeling (SEM) literature, a study by Gagne and Hancock (2006) offers clues as to how the factor loading magnitudes and sample size affect model convergence rates. They found that for factors extracted from 3 items with equal factor loadings, they needed sample sizes of 50, 100, and 400 when factor loadings were 0.8, 0.6, and 0.4, respectively, to achieve "satisfactory" convergence rates (at least 1,000 successful solutions out of 1,100). Even with sample sizes of 1,000, models with loadings of 0.2 had serious convergence problems. These results are important; however, Gagne and Hancock's study differs from the proposed research in the following key ways. First, the manifest variables in their study were items, not parcels, a fact that affects interpretation of the results.

Second, theirs was only a CFA model, which contained no predictive element (no downstream elements). Based on the Gagne and Hancock (2006) findings as well as theoretical reasoning, the following are hypotheses relating to convergence rates:

1) Because heterogeneous parcels will share both true score variance and specific variance, whereas homogeneous parcels will share only true score variance (Gibson, 2012; Little et al., 2013), the standardized factor loadings are likely to be smaller in the models using homogeneous parcels. If the finding that very small factor loadings decrease the likelihood of convergence (Gagne & Hancock, 2006) applies to structural models, the use of homogeneous parcels will decrease convergence rates, compared to the use of heterogeneous parcels. Thus, I hypothesize that *when the standardized higher-order loadings in the DG model are small, models using the homogeneous parceling strategy will have poorer convergence rates than the models using heterogeneous parceling.*
2) As the standardized higher-order loadings in the DG model increase to medium or large strength, the shared specific variance, found only in the heterogeneous parcels, will no longer be instrumental for retaining factor loadings large enough for successful convergence rates. Thus, I hypothesize that *when the standardized higher-order loadings in the DG model are not small, both parceling strategies will have near perfect convergence rates.*

Bias in estimates of gamma. Once the model converges, the next important question to ask is: Are the structural paths predictably over- or under-estimated? In other words, in this study I seek to identify which parceling strategy is associated with the least biased parameter estimates (preferably this coincides with standard errors that best estimate the true SD of gamma estimates). Since the literature dictates that parcels should never be used by those interested in the structure of a LV and that parcels are defensible only when the researcher is interested in the relation between the LVs, I will not discuss the estimates of factor loadings in this paper. For this reason, the structural path between the exogenous LV and the endogenous manifest variable, commonly referred to as gamma (γ), is the only parameter I will discuss.

Previous simulation studies indicate that when the structural paths in the DG model are included in estimation models utilizing parceling strategies, the structural paths tend to be underestimated (Bandalos, 2002; Coffman & MacCallum, 2005; Hall et al., 1999). Only one of these studies used a DG model with a higher-order factor structure similar to the one proposed here. Coffman and MacCallum (2005) reported that the bias in estimates of the structural paths was less when homogeneous parcels were employed. In fact, the structural paths for the homogeneous parceling model had percentages of parameter estimate biases (PPEB)[1] ranging from 0 to 5; compared to the PPEB from the heterogeneous parceling model that ranged from 25 to 67. It is

[1] Percentage of parameter estimate bias calculated as:

important to note, however, that the intent of their study was not to compare the two parceling strategies. Furthermore, they did not use homogeneous factor loadings in their DG model. For example, the factor loadings from the sub-factor to the items ranged from 0.3 to 0.5 (the current study uses 0.4 consistently) and the factor loadings from the higher-order factor to the sub-factors ranged from 0.6 to 0.8. Based on theoretical reasoning and results from the aforementioned studies, the following is my hypothesis relating to bias in estimates of γ:

> Although none of the studies attempted to explain why homogeneous parcels had less bias, I believe that it is a result of the LV extracted from the parcels. Gibson (2012) illustrates beautifully how the variance of an LV represented by homogeneous parcels (LV_{Hom}) will include only true variance, whereas an LV represented by heterogeneous parcels (LV_{Het}) will include true and specific variance (see Appendix A Figure A6). Consequently, a smaller proportion of the LV_{Het} will predict the dependent variable compared to the LV_{Hom} (assuming a non-standardized model), further decreasing γ estimates in the heterogeneous models. I hypothesize that *as the standardized higher-order loadings in the DG model increase, the model using the homogeneous parceling strategy will estimate* γ *with less bias.*

Model fit indices. Another important component of this study is how the two parceling strategies affect model fit indices. One of the biggest arguments raised by critics of parcels is that using parcels will always result in better model fit compared to models with items. Model fit improves with parcels, compared to items, because parceling models have fewer chances for cross-loadings and correlated disturbances. Even more alarming is the possibility that the choice of parceling strategy also affects the likelihood of properly accepting or rejecting the model. Arguably, the worst-case scenario would be if one parceling strategy had better model fit indices, despite having more biased estimates of γ than the other parceling strategy. Although that worst-case scenario occurred in Hall et al.'s study (1999), the current study will likely not have that problem because the structural components of the estimated models are properly specified.

Coffman and MacCallum (2005) found that the model fit for the heterogeneous parceling models were extremely close to the model fit of the homogeneous parceling models, despite having much more biased parameter estimates. In fact, the RMSEA for the homogeneous parceling model was 0.00, compared to 0.0032 for the heterogeneous parceling model. This is alarming considering the 25%- 67% parameter bias estimate for structural paths in the heterogeneous parceling model. Based on the previous findings as well as theoretical reasoning, the following is my hypothesis relating to goodness-of-fit indices:

[(mean estimated value – true parameter value) / true parameter value] x 100

The covariance matrix of the homogeneous parcels will have smaller covariances than the heterogeneous covariance matrix because homogeneous parcels do not share specific variance. Furthermore, when the observed covariances approach zero any model will fit well (Cole & Preacher, 2014). Thus, I hypothesize that *models using the homogeneous parceling strategy will have better model fit indices* than will the model using the heterogeneous parceling strategy.

CHAPTER II

Methods

I will use Mplus software to generate data from the model depicted in Figure 1 and estimate the parceling models depicted in Figure 2. Each DG model illustrates a simple multidimensional, hierarchical latent structure that predicts a single manifest variable. In total, there will be nine DG models that vary only in terms of their higher-order factor loadings (and consequently, the unique variance associated with the sub-factors). For each DG model, a sample with 500 cases will be drawn from a multivariate normal population (10 variables per case).

For each generated sample, I will use Mplus to create two 4x4 covariance matrices. The four variables in the covariance matrices are as follows: three parcels, each created by summing three items, and the outcome variable (y). The two covariance matrices will differ based on the parceling strategy utilized to create the three parcels. Specifically, one matrix will use the homogeneous parceling strategy while the other uses a heterogeneous strategy (see Figure 2 for specific item allocations). Each covariance matrix will be used to estimate the same structural model.

Study 1

The main inquiry for Study 1 is to compare homogeneous and heterogeneous parceling strategies by quantifying the likelihood of model convergence and model diagnostic problems under each simulated condition. For this section 1,500 samples were generated for each higher-order factor loading (i.e., 13,500 samples); and each sample was analyzed twice (using the homogeneous and heterogeneous parceling strategies). Results for the *Model Convergence* section report percentages based on the original 1,500 samples. The subsequent sections, *Diagnostic Information about the Model* and *Gamma Estimates*, report percentages based on only the models that converged for each condition.

Study 2

The main goal for Study 2 is to compare homogeneous and heterogeneous parceling strategies by statistically comparing the model fit indices and γ estimates of matched plausible solutions. For this section, the first 1,500 samples that converged and had plausible solutions (no Heywood cases) for both parceling strategies were used for analysis.

CHAPTER III

Results

Study 1 Results: 1,500 Repetitions / Converged Models

The results reported include the: (a) likelihood of models failing-to-converge, (b) likelihood of converged models having diagnostic problems (Heywood cases and/or poor model fit), and (c) γ estimates and SE of γ estimates for Plausible Solutions and Heywood cases. For this chapter Mplus generated 1,500 repetitions/samples per cell condition.

Model Convergence. This section contains information about model convergence in 1,500 samples drawn for each of the 18 cells (9 higher-order factor loadings x 2 parceling strategies). Table 1 reports the frequency and percentage of converged and failed-to-converge models. Failed-to-converge cases are models that exceeded the default number of iterations set by Mplus (10 random initial stage iterations and 1,000 iterations for the Quasi-Newton algorithm for continuous outcomes; Muthén & Muthén, 2010). Consequently, failed-to-converge cases have no parameter estimates to analyze. Converged cases do have solutions, although they may contain out-of-range parameter estimates (Heywood cases) and may not provide a good fit to the data. For easier identification of trends, information from Table 1 is illustrated in Figure C1 (see Appendix C).

Both parceling strategies had 100% convergence rates when higher-order factor loadings were 0.8 and 0.9. When the higher-order factor loadings in the DG model were smaller, the heterogeneous strategy maintained a higher likelihood of converged solutions compared to the homogeneous parceling strategy. When higher-order factor loadings were at their smallest, the difference between the strategies was especially large, 54.8% convergence for homogeneous versus 96.4% for heterogeneous parceling strategies. When factor loadings were 0.1, the homogeneous parceling strategy had nearly a 50% chance of converging.

Diagnostic Information about the Model. In this section, I focus on the subset of converged models from the original 1,500 cases. Consequently, the number of converged models varies across conditions; calculated as 1,500 minus the number of failed-to-converge models for that cell (see Table 1). I will focus on two specific diagnostic areas: Heywood cases and goodness-of-fit indices. Solutions with negative unique variance estimates, a certain type of Heywood case, were broken into two categories. The first category, called a Theta-Heywood case, occurred when a negative variance estimate was associated with the error term of any of the three parcels. The second category, called a Psi-Heywood case, occurred when a negative variance estimate was associated with the error term of the dependent variable (Y). Converged solutions that did not result in a negative variance estimates are called “Plausible Solutions.”

Table 1

Frequency and Percentage of Model Non-convergence for 1,500 Replications

Parceling Strategy	Factor Loading	Failed-to-Converge[1] N	Failed-to-Converge[1] %
Homogeneous	0.1	678	45.20
	0.2	633	42.20
	0.3	517	34.47
	0.4	343	22.87
	0.5	145	9.67
	0.6	31	2.07
	0.7	2	0.13
	0.8	0	0
	0.9	0	0
Heterogeneous	0.1	54	3.60
	0.2	32	2.13
	0.3	13	0.87
	0.4	5	0.33
	0.5	0	0
	0.6	0	0
	0.7	0	0
	0.8	0	0
	0.9	0	0

[1]Failed-to-Converge means non-convergence after 1,000 iterations. Converged solutions include Heywood cases.
Note: Numbers in the “Factor Loading” column represent the higher-order factor loading used in the data generative model.

Table 2 reports the frequency and percentage of Plausible Solutions versus the specific types of Heywood cases. For easier identification of trends, information from Table 2 is illustrated in Appendix C Figure C2. Results show that when a model converges, it is highly likely to generate a plausible solution (rather than a Heywood case), no matter the parceling strategy. A trend does seem to be apparent, however, in that Heywood cases are more likely to occur when the higher-order factor loadings are small and/or when homogeneous parceling is used. Furthermore, Psi-Heywood cases occurred only when models were estimated using the homogeneous parceling strategy. Theta-Heywood and Psi-Heywood cases occurred at approximately the same rate when the higher-order factor loadings were 0.3 or less; however, when the higher-order factor loadings were larger, the likelihood of a Psi-Heywood cases decreased compared to Theta-Heywood cases.

A second important set of diagnostics pertain to goodness-of-fit. Table 3 reports the means, standard deviations (SD), and ranges for five model fit indices: χ^2, RMSEA, TLI, SRMR, and CFI. All model fit indices, with the exception of the CFI, indicated that homogeneous parceling leads to equal or better model fit compared to heterogeneous parceling. Furthermore, a trend emerged: as the factor loadings increased, model fit tended to diminish on all indices except the CFI. This trend is consistent with my expectations; when the observed covariances approach zero, any model fit will improve (Cole & Preacher, 2014; Tomarken & Waller, 2003). The TLI had very large means when homogeneous parceling models were coupled with low higher-order factor loadings (i.e., 0.4 or less).

Table 2

Frequency and Percentage of Plausible Solutions and Heywood Cases for Converged Models

Parceling Strategy	Factor Loading	Plausible Solution[1] N	Plausible Solution[1] %	Negative Variance: Theta N	Theta (%)	Psi N	Psi (%)
Homogeneous	0.1	724	88.08	54	6.57	44	5.35
	0.2	750	86.51	69	7.96	48	5.54
	0.3	885	90.03	52	5.29	46	4.68
	0.4	1085	93.78	46	3.98	26	2.25
	0.5	1308	96.53	40	2.95	7	0.52
	0.6	1446	98.43	23	1.57	0	0
	0.7	1491	99.53	7	0.47	0	0
	0.8	1500	100	0	0	0	0
	0.9	1500	100	0	0	0	0
Heterogeneous	0.1	1411	97.58	35	2.42	0	0
	0.2	1439	98.02	29	1.98	0	0
	0.3	1465	98.52	22	1.48	0	0
	0.4	1486	99.40	9	0.60	0	0
	0.5	1495	99.67	5	0.33	0	0
	0.6	1499	99.93	1	0.07	0	0
	0.7	1500	100	0	0	0	0
	0.8	1500	100	0	0	0	0
	0.9	1500	100	0	0	0	0

[1]Plausible Solution refers to cases that converged and generated no Heywood cases.
Note: Numbers in the "Factor Loading" column represent the higher-order factor loading used in the data generative model.

Various cutoff values are often used to determine whether a model should be rejected. As each index has particular strengths and weaknesses, Hu and Bentler (1999) suggested a combination of index cutoffs for model rejection. This is called a "Two-Index Presentation Strategy" (TIPS). Hu and Bentler suggested that researchers reject models using any of the three index combinations: (a) TLI below 0.96 or an SRMR greater than 0.09; (b) RMSEA above 0.06 or an SRMR greater than 0.09; or (c) CFI below 0.96 or an SRMR greater than 0.09. Note that all of these combinations require an SRMR smaller than 0.09 to fail-to-reject a model. Figure 3 presents rejection rates percentages for of each of Hu and Bentler's TIPS as well as the χ^2 index (i.e., $p \leq .05$). Appendix C Table C1 reports the frequencies and percentages of model rejections.

Several key findings emerged. First, the SRMR was always 0.06 or smaller. Consequently, the SRMR never met the Hu and Bentler cutoff of 0.09 for any condition; thus the SRMR never contributed to any of Hu and Benter's TIPS criteria. Second,

Table 3

Model Fit Indices as a Function of Higher-Order Factor Loading and Parceling Strategy for Converged Models

Parceling Strategy	Factor Loading	χ^2 (df=2)			RMSEA			TLI			SRMR			CFI		
		M	*SD*	*Range*	*M*	*SD*	*Range*	*M*	*SD*	*Range*	*M*	*SD*	*Range*	*M*	*SD*	*Range*
Homogeneous	0.1	0.88	0.91	7.72	0.00	0.01	0.11	4.21	37.33	1035.20	0.01	0.01	0.04	0.96	0.16	1
	0.2	0.96	1.00	8.23	0.00	0.02	0.11	12.98	266.81	7833.32	0.01	0.01	0.04	0.96	0.15	1
	0.3	1.13	1.29	15.05	0.01	0.02	0.16	3.01	16.13	478.63	0.02	0.01	0.06	0.95	0.16	1
	0.4	1.32	1.38	10.90	0.01	0.02	0.13	2.62	11.19	232.21	0.02	0.01	0.05	0.95	0.14	1
	0.5	1.56	1.60	11.26	0.01	0.03	0.14	1.71	13.94	610.76	0.02	0.01	0.05	0.96	0.10	1
	0.6	1.74	1.77	12.19	0.02	0.03	0.14	1.29	2.93	99.05	0.02	0.01	0.05	0.97	0.08	1
	0.7	1.85	1.87	12.80	0.02	0.03	0.15	1.04	0.25	3.26	0.02	0.01	0.05	0.98	0.05	0.63
	0.8	1.91	1.92	13.20	0.02	0.03	0.15	1.02	0.14	1.90	0.02	0.01	0.05	0.98	0.03	0.23
	0.9	1.94	1.94	13.29	0.02	0.03	0.15	1.01	0.09	1.68	0.02	0.01	0.05	0.99	0.02	0.18
Heterogeneous	0.1	1.74	1.74	11.61	0.02	0.03	0.14	1.12	0.67	24.06	0.02	0.01	0.06	0.97	0.07	1
	0.2	1.79	1.79	13.59	0.02	0.03	0.15	1.25	5.39	280.24	0.02	0.01	0.05	0.97	0.06	0.58
	0.3	1.85	1.84	14.13	0.02	0.03	0.16	1.06	0.37	8.38	0.02	0.01	0.05	0.97	0.06	1
	0.4	1.89	1.89	14.38	0.02	0.03	0.16	1.03	0.23	3.26	0.02	0.01	0.05	0.98	0.05	0.60
	0.5	1.93	1.94	14.51	0.02	0.03	0.16	1.02	0.18	1.81	0.02	0.01	0.05	0.98	0.04	0.30
	0.6	1.95	1.96	14.57	0.02	0.03	0.16	1.01	0.14	1.69	0.02	0.01	0.05	0.98	0.03	0.24
	0.7	1.96	1.97	14.60	0.02	0.03	0.16	1.01	0.11	1.69	0.02	0.01	0.05	0.99	0.03	0.19
	0.8	1.97	1.98	14.60	0.02	0.03	0.16	1.00	0.09	1.73	0.02	0.01	0.05	0.99	0.02	0.15
	0.9	1.98	1.98	14.58	0.02	0.03	0.16	1.00	0.07	1.77	0.02	0.01	0.05	0.99	0.02	0.12

Note: Numbers in the “Factor Loading” column represent the higher-order factor loading used in the data generative model. df=degrees of freedom. M = mean. SD = standard deviation. RMSEA=root mean square error of approximation. TLI=Tucker Lewis index. SRMR=standardized root mean square residual. CFI=comparative fit index.

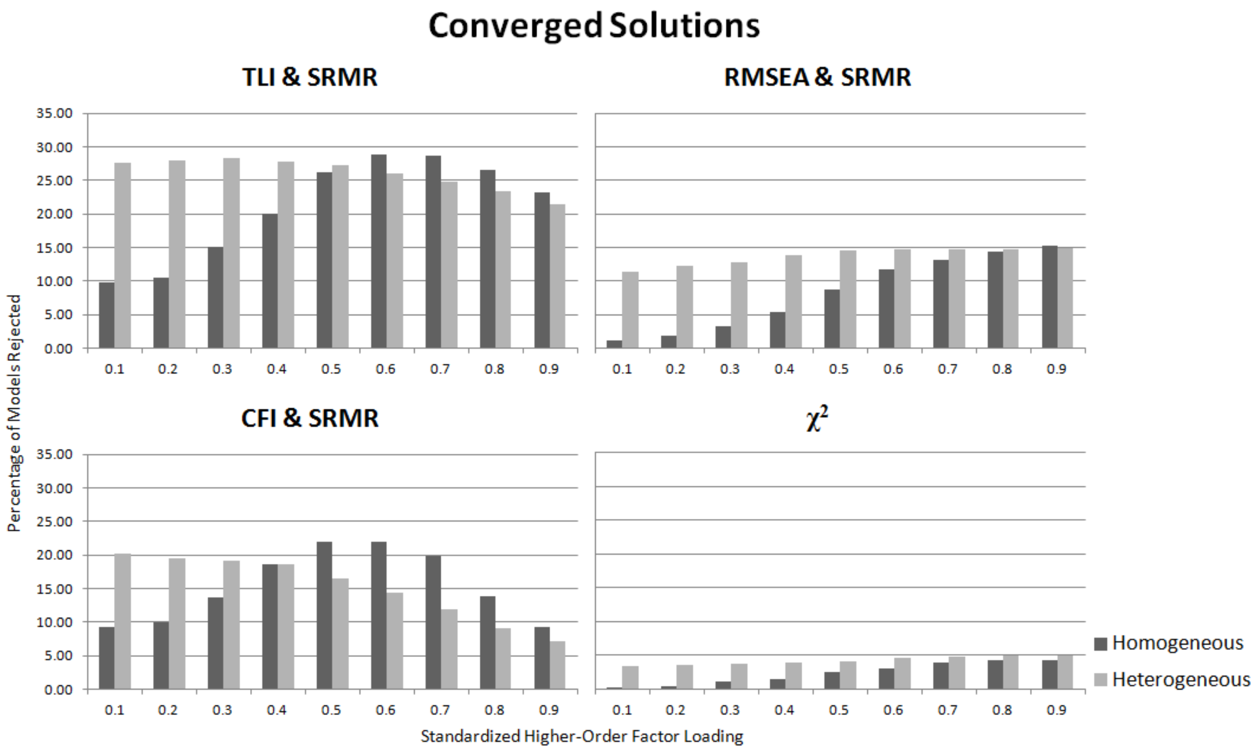


Figure 3. Percentage of converged models that were rejected by Hu and Bentler's (1999) Two-Index Presentation Strategies. TLI/SRMRS rejects models when TLI ≤ 0.96 or SRMR ≥ .09. RMSEA/SRMR rejects models when RMSEA ≥ 0.06 or SRMR ≥ .09. CFI/SRMR rejects models when CFI ≤ 0.96 or SRMR ≥ .09. χ^2 index rejects models when $p \leq .05$.

parceling strategy appears to have more influence on rejection rates when higher-order factor loadings are small. Specifically, when higher-order factor loadings were 0.4 or smaller, the homogeneous parceling strategy consistently led to lower rejection rates than did the heterogeneous parceling strategy. In contrast, when higher-order factor loadings were 0.5 or higher, the heterogeneous parceling strategy led to proportionally fewer rejected models, at least for TLI/SRMR and CFI/SRMR TIPS criteria. Even for the RMSEA/SRMR and χ^2 indices, the homogeneous parceling models showed an increase in the number of rejected models when the factor loadings were larger, whereas the heterogeneous parceling strategy had a model rejection rate that was relatively constant across the different levels of factor loadings. The only exception was the CFI/SRMR TIPS combination, where the rejection rate of heterogeneous models decreased as loadings increased.

The next point of interest is how Heywood cases and goodness-of-fit indices co-occurred as diagnostic tools. Figure 4 shows stacked bar graphs of the four possible outcomes for each model. The bottom portion of each bar (in purple) represents solutions with Heywood cases but good model fit. The second part of each bar (in light blue) represents solutions with a poor model fit by one or more TIPS criteria, yet no Heywood cases. The third part of each bar (in red) represents solutions that had a poor model fit and a Heywood case. The top part of each bar (in green) reflects solutions in which model fit was good and there

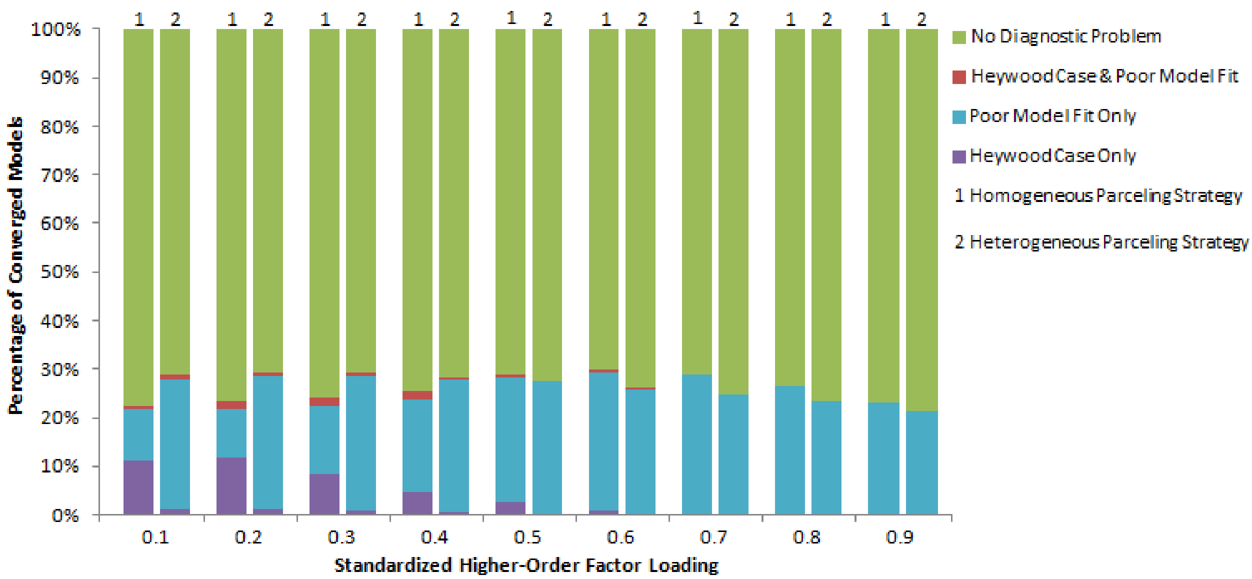


Figure 4. Percentage of converged models by diagnostic situation.

were no Heywood cases. These four outcomes are mutually exclusive and exhaustive; therefore, each of the bars total to 100%. For each higher-order factor loading the homogeneous parceling model is always on the left and the heterogeneous parceling model is on the right. Appendix C Table C2 reports the frequency and percentage of these diagnostic situations.

Several important results emerged. First, when higher-order factor loadings are 0.4 or greater, poor model fit alone appears to be the most common diagnostic problem for heterogeneous and homogeneous parceling models. In contrast, when higher-order factor loadings are 0.3 or smaller and a homogeneous parceling strategy is utilized, a Heywood case or poor model fit seems equally likely. Although poor model fit remains the most likely diagnostic problem for heterogeneous parceling strategies, no matter the higher-order factor loadings. Despite the increase of Heywood cases at smaller values of the higher-order factor loadings, the percentages of converged models that had no diagnostic problems were relatively constant across all conditions, ranging from 70.18% to 78.6%. Furthermore, very few of the models had both diagnostic problems. In fact, at most 1.73% of models accounted for poor fitting models with a Heywood case within any condition (this was the 0.3 homogeneous parceling condition).

Gamma Estimates. Analyses in this section again used all converged models. Figure 5 graphs the mean standardized estimates of γ and their mean standard errors as a function of parceling strategy, higher-order factor loading, and whether the

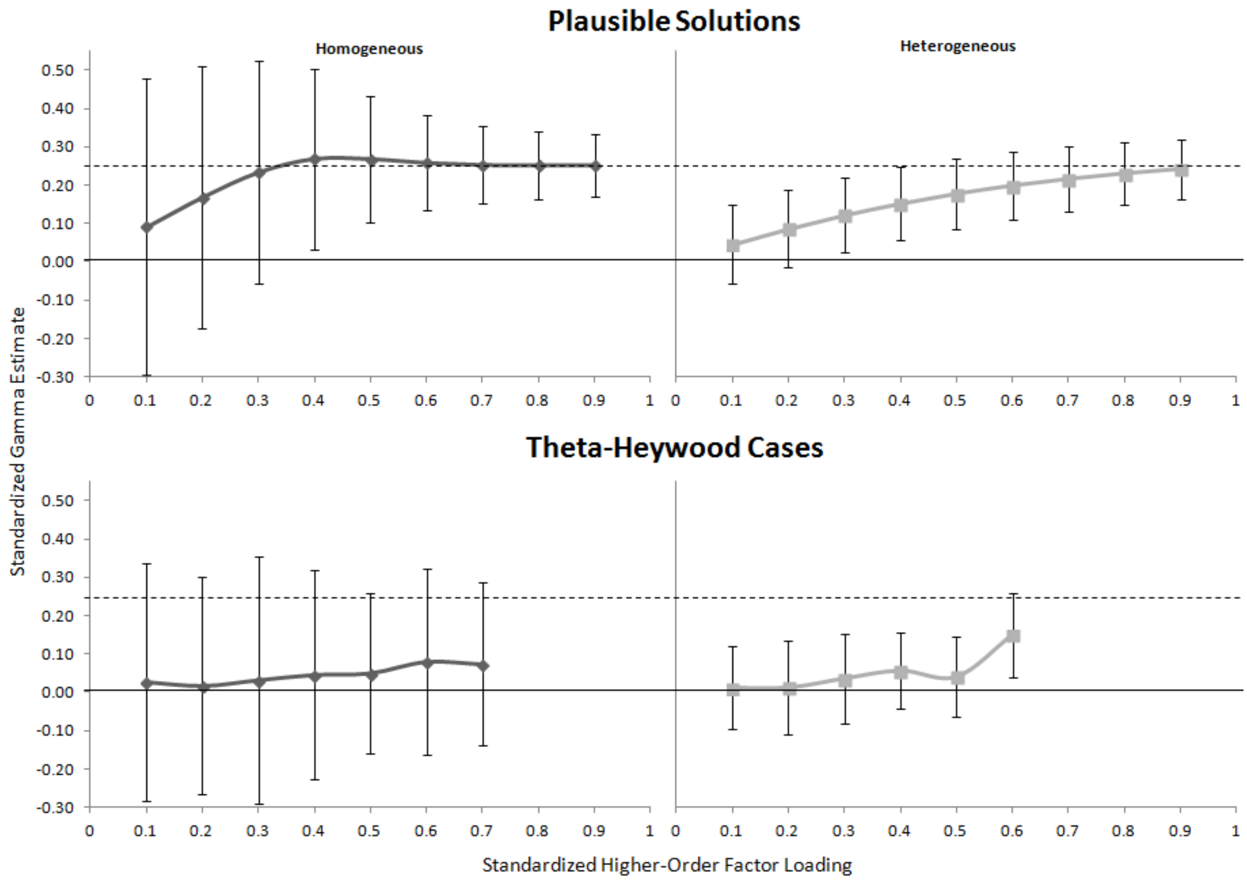


Figure 5. Mean standardized gamma estimates for converged models that resulted in Plausible Solutions or Theta-Heywood cases as function of higher-order factor loadings and parceling strategy. Error bars depict the mean standard error associated with that particular condition. The solid black line indicates a standardized gamma estimate of 0. The dotted line indicates 0.25, the population value of gamma.

converged model resulted in Plausible Solutions or Theta-Heywood cases[2]. Two horizontal lines have been added to the graphs, the solid black line marks a γ value of 0; the dotted line at 0.25 represents the population value of γ. Error bars have been drawn around the mean γ estimates to represent the average SE of γ estimates for each condition. Appendix C Tables C3 and C4 reports the mean, SD, and range of γ estimates and SE of γ estimates respectively.

When comparing the γ estimates from the Plausible Solutions to the Theta-Heywood cases, we see that for both parceling strategies, the Plausible Solutions tend to become less biased with respect to γ as the higher-order factor loadings increase. The estimates from the Theta-Heywood cases tended to remain close to 0 even at higher-order factor loadings of 0.5 and 0.6.

[2] Note that the results for solutions with Psi-Heywood cases were not reported in this section due to their extremely inflated standardized γ estimates (e.g. 8.03). The excessive inflation of γ estimates in Psi-Heywood cases is expected given that the variance of the Y is constrained: $var(Y) = var(D_{Hom}) * \gamma^2 + var(error_Y) * 1^2$. Because the results are standardized, $var(Y)$ and $var(D_{Hom}) = 1$; so the equation reduces to $1 = \gamma^2 + var(error_Y)$. If the error variance of Y is negative then the γ estimate must be inflated to bring their sum to +1.

Additionally, for the homogeneous parceling strategy the SE of γ estimates for the Plausible Solutions tends to decrease as the higher-order factor loadings increase; whereas the SE of γ estimates tends to remain constant for Theta-Heywood cases.

Comparing the parceling strategies for the Plausible Solutions only, we see that the homogeneous parceling strategy covers the population value for gamma (i.e., 0.25) more consistently than did the heterogeneous parceling strategy across all factor loading levels. Specifically, the homogeneous strategy estimates of γ were less biased compared to the heterogeneous factor loadings, and the SE bars contained 0.25 for all factor loadings. Furthermore, when higher-order factor loadings were 0.7 or greater, the homogeneous parceling strategy recovered an unbiased γ estimate. In contrast, the error bars for the heterogeneous parceling strategy only included 0.25 when factor loadings were 0.5 or greater, and revealed negatively biased γ estimates even when higher-order factor loadings were quite large (i.e., 0.9). Interestingly, when factor loadings were small, the heterogeneous parceling strategy yielded small, relatively consistent SEs for γ estimates; whereas the homogeneous parceling strategy had very large SEs This means that although the homogeneous parceling strategy showed better coverage compared to the heterogeneous parceling strategy, it also had less power to detect a significant effect. Coverage, power, and Type I error are examined more completely in Study 2.

Study 2 Results: 1,500 Matched Plausible Solutions

For the results section, Mplus generated samples and estimated models using the same procedure and syntax as described in Study 1; however, I discarded samples that resulted in non-convergence or a Heywood case for either parceling strategy. For example, when higher-order factor loadings were 0.1, Mplus generated 3,303 replications to obtain 1,500 samples for which both parceling strategies converged on Plausible Solutions; however, for conditions in which the higher-order factor loadings were large (0.8 and 0.9), 100% of the solutions were plausible for both parceling strategies, meaning that only 1,500 repetitions were needed. Samples in this section are referred to as "Matched Plausible Solutions."

Results contain (a) descriptive statistics about the goodness-of-fit indices and statistical comparisons across parceling strategies, (b) the likelihood of model rejection, (c) descriptive statistics about γ estimates and their SEs and statistical comparisons across parceling strategies (d) coverage, power, and Type I error rates and (e) the difference between the SD of γ estimates and the SE of γ estimates of as a function of parceling strategy and higher-order factor loading.

Model Fit Indices. Table 4 reports the descriptive statistics for model fit indices (similar to Table 3) for the 1,500 Matched Plausible Solutions of each condition. Unlike Table 3, Table 4 does not include data from Heywood cases. Despite these differences, the results reported in Table 3 and Table 4 are very similar. For the latter, however, I was able to conduct statistical tests to enhance our understanding of the differences between parceling strategies.

A mixed-ANOVA (parceling strategy was a within-group factor, whereas higher-order factor loading was a between-group factor) was conducted for the five goodness-of-fit indices. Table 5 reports the results of significance tests as well as effect sizes (ω^2). *P-values* were significant for all main and interaction effect tests, with the exception of the TLI main effect for higher-order factor loadings and the interaction effect (both of which had large SDs). Cohen (1988) suggested that ω^2 be interpreted as small, medium, and large when values are 0.01, 0.06, and 0.14, respectively. Effect sizes for the goodness-of-fit indices ranged from small to negligible (a.k.a., “too small to even acknowledge”). Overall, results indicate that parceling strategy, higher-order factor loading, and their interaction had small effects on χ^2 , RMSEA, SRMR, and CFI goodness-of-fit indices. Parceling strategy and higher-order factor loadings had little to no impact on the TLI.

The prime focus of this paper was comparing the parceling strategies; therefore, post hoc dependent t-tests were conducted comparing the parceling strategies at each factor loading level. Table 6 reports the results of the nine dependent t-tests as well as the associated effect sizes (Cohen’s *d*). Note that *p-values* in the table reflect a Bonferroni correction. A table reporting the mean difference and SD of the differences for each of the goodness of fit indices (Table C3) and a table reporting the correlation between the homogeneous and heterogeneous parceling fit indices (Table C4) can be found in Appendix C. Again, many of the significance tests resulted in *p-values* $\leq$ .05; however, results tended towards non-significance as the higher-order factor loadings increased. Specifically, the χ^2 , RMSEA, and SRMR had non-significant *p-values* when the factor loadings were 0.7 or larger but were significant when factor loadings were 0.4 or less. Cohen (1988) suggested that effect sizes be interpreted as small, medium, and large when values of Cohen’s *d* are 0.2, 0.5, and 0.8 respectively. Effect sizes for comparing the parceling strategies’ goodness-of-fit indices again ranged from small to negligible. Small effect sizes occurred for the χ^2 , RMSEA, and SRMR when higher-order factor loadings were 0.3 or less and for the CFI when factor loadings were between 0.4 and 0.6.

In summary, regarding the χ^2 , RMSEA, TLI, and SRMR, results show that homogeneous parceling led to statistically better model fit compared to heterogeneous parceling when higher-order factor loadings were small to medium. Although statistically significant, effect sizes were small for χ^2 , RMSEA, and SRMR when factor loadings were small and negligible under the other conditions. The CFI behaved differently than the other indices. The heterogeneous parceling strategy resulted in statistically better model fit for most of the higher-order factor loadings, although this effect was small when factor loadings were medium strength and negligible at all other strengths.

Table 4

Model-Fit Indices Descriptive Statistics for Matched Plausible Solutions as a Function of Higher-Order Factor Loading and Parceling Strategy

Parceling	Factor	χ^2 (*df*=2)			RMSEA			TLI			SRMR			CFI		
Strategy	Loading	*M*	*SD*	*Range*	*M*	*SD*	*Range*	*M*	*SD*	*Range*	*M*	*SD*	*Range*	*M*	*SD*	*Range*
Homogeneous	0.1	0.90	0.94	8.23	0.00	0.01	0.11	3.27	28.76	1035.30	0.01	0.01	0.05	0.96	0.15	1.00
	0.2	0.98	1.02	8.22	0.01	0.02	0.11	8.41	202.92	7820.78	0.01	0.01	0.04	0.96	0.16	1.00
	0.3	1.13	1.27	15.05	0.01	0.02	0.16	4.55	51.74	1909.64	0.02	0.01	0.06	0.95	0.16	1.00
	0.4	1.35	1.46	10.90	0.01	0.02	0.13	2.36	9.71	242.74	0.02	0.01	0.05	0.95	0.15	1.00
	0.5	1.57	1.62	11.26	0.01	0.03	0.14	1.79	13.78	610.75	0.02	0.01	0.05	0.96	0.10	1.00
	0.6	1.73	1.75	12.19	0.02	0.03	0.14	1.29	2.90	99.05	0.02	0.01	0.05	0.97	0.07	1.00
	0.7	1.84	1.86	12.80	0.02	0.03	0.15	1.05	0.25	3.26	0.02	0.01	0.05	0.98	0.05	0.63
	0.8	1.91	1.92	13.19	0.02	0.03	0.15	1.02	0.14	1.28	0.02	0.01	0.05	0.98	0.03	0.23
	0.9	1.94	1.94	13.29	0.02	0.03	0.15	1.01	0.09	0.74	0.02	0.01	0.05	0.99	0.02	0.18
Heterogeneous	0.1	1.78	1.84	15.94	0.02	0.03	0.17	1.13	0.57	11.36	0.02	0.01	0.06	0.97	0.06	0.46
	0.2	1.83	1.90	16.12	0.02	0.03	0.17	1.22	5.31	206.77	0.02	0.01	0.06	0.97	0.06	0.51
	0.3	1.84	1.87	15.04	0.02	0.03	0.16	1.06	0.39	8.07	0.02	0.01	0.06	0.98	0.05	0.43
	0.4	1.86	1.87	14.38	0.02	0.03	0.16	1.03	0.20	2.19	0.02	0.01	0.05	0.98	0.04	0.34
	0.5	1.92	1.92	14.50	0.02	0.03	0.16	1.02	0.17	1.52	0.02	0.01	0.05	0.98	0.04	0.27
	0.6	1.92	1.94	14.57	0.02	0.03	0.16	1.01	0.13	1.07	0.02	0.01	0.05	0.99	0.03	0.22
	0.7	1.96	1.97	14.60	0.02	0.03	0.16	1.01	0.11	0.85	0.02	0.01	0.05	0.99	0.03	0.19
	0.8	1.97	1.98	14.60	0.02	0.03	0.16	1.00	0.09	0.65	0.02	0.01	0.05	0.99	0.02	0.15
	0.9	1.98	1.98	14.58	0.02	0.03	0.16	1.00	0.07	0.51	0.02	0.01	0.05	0.99	0.02	0.12

Note: Numbers in the "Factor Loading" column represent the higher-order factor loading used in the data generative model. df=degrees of freedom. M = mean. SD = standard deviation. RMSEA=root mean square error of approximation. TLI=Tucker Lewis index. SRMR=standardized root mean square residual. CFI=comparative fit index.

Table 5

Mixed ANOVA Results: Significance Testing and Effect Sizes of Goodness-of-Fit Indices

Fit Index	Source	Design	*F*	*p-value*	ω^2
χ^2	Parcel	Within	380.71	0.000	0.03
	FactorLoad	Between	54.39	0.000	0.03
	Parcel * FactorLoad	Within	28.15	0.000	0.02
RMSEA	Parcel	Within	294.00	0.000	0.02
	FactorLoad	Between	50.57	0.000	0.03
	Parcel * FactorLoad	Within	23.32	0.000	0.01
TLI	Parcel	Within	7.75	0.005	0.00
	FactorLoad	Between	1.89	0.05	0.00
	Parcel * FactorLoad	Within	1.69	0.09	0.00
SRMR	Parcel	Within	159.06	0.000	0.01
	FactorLoad	Between	12.43	0.000	0.01
	Parcel * FactorLoad	Within	36.80	0.000	0.02
CFI	Parcel	Within	209.44	0.000	0.02
	FactorLoad	Between	41.90	0.000	0.02
	Parcel * FactorLoad	Within	9.79	0.000	0.01

Note. RMSEA=root mean square error of approximation. TLI=Tucker Lewis index. SRMR=standardized root mean square residual. CFI=comparative fit index.

Using the same Hu and Bentler Two-Index Presentation cutoffs outlined in Study 1, Figure 6 graphs the percentage of Matched Plausible Solution models that meet criteria for rejection (actual frequencies and percentages are reported in Appendix C Table C5). Note that Figure 6 is nearly identical to Figure 3. Thus, the same conclusions maybe drawn for rejection rates for the converged models and the Matched Plausible Solutions: (a) the SRMR never met the Hu and Bentler cutoff of .09 for any condition, (b) parceling strategy appears to have had the most influence on rejection rates when higher-order factor loadings were small, (c) across all higher-order factor loadings the homogeneous parceling strategy had fewer rejections for the RMSEA and χ^2 compared to the heterogeneous parceling strategy, and (d) homogeneous parceling has lower model rejection rates for the TLI and CFI when the factor loadings are small, but heterogeneous parceling had fewer model rejection rates when factor loadings were large.

Gamma Estimates. Figure 7 presents the mean γ estimates and average SE of γ estimates for the Matched Plausible Solutions (Appendix C Tables C5 and C6 report the mean, SD, and range of γ estimates and SE of γ estimates, respectively[3]). The

[3] A table reporting the mean and SDs of parcel factor loadings is in Appendix C, Table C9. A table reporting the mean difference and SD of the difference of the combined parcel factor loadings is in Appendix C, Table C10.

Table 6

Dependent t-test Results Comparing Parceling Strategies: Significance Testing and Effect Sizes for Goodness-of-Fit Indices

Factor	χ^2			RMSEA			TLI			SRMR			CFI		
Loading	*t*	*p-value*	Cohen's *d*	*t*	*p-value*	Cohen's *d*	*t*	*p-value*	Cohen's *d*	*t*	*p-value*	Cohen's *d*	*t*	*p-value*	Cohen's *d*
0.1	16.53	0.000	0.43	14.96	0.000	0.39	2.89	0.004	0.07	14.44	0.000	0.37	1.77	0.077	0.05
0.2	15.60	0.000	0.40	14.01	0.000	0.36	1.37	0.171	0.04	12.72	0.000	0.33	3.64	0.000	0.09
0.3	12.22	0.000	0.32	11.38	0.000	0.29	2.61	0.009	0.07	8.94	0.000	0.23	5.63	0.000	0.15
0.4	8.45	0.000	0.22	7.46	0.000	0.19	5.30	0.000	0.14	4.85	0.000	0.13	7.89	0.000	0.20
0.5	5.39	0.000	0.14	4.70	0.000	0.12	2.17	0.030	0.06	1.76	0.079	0.05	8.27	0.000	0.21
0.6	2.89	0.004	0.07	2.32	0.021	0.06	3.71	0.000	0.10	0.55	0.584	0.01	8.35	0.000	0.22
0.7	1.75	0.081	0.05	1.36	0.173	0.04	5.75	0.000	0.15	1.23	0.217	0.03	6.85	0.000	0.18
0.8	0.90	0.368	0.02	0.63	0.526	0.02	2.72	0.007	0.07	1.37	0.172	0.04	4.89	0.000	0.13
0.9	0.57	0.570	0.01	0.34	0.732	0.01	1.19	0.233	0.03	0.74	0.461	0.02	2.33	0.020	0.06

Note. Numbers in the "Factor Loading" column represent the higher-order factor loading used in the data generative model. *p*-values reflect Bonferroni correction.

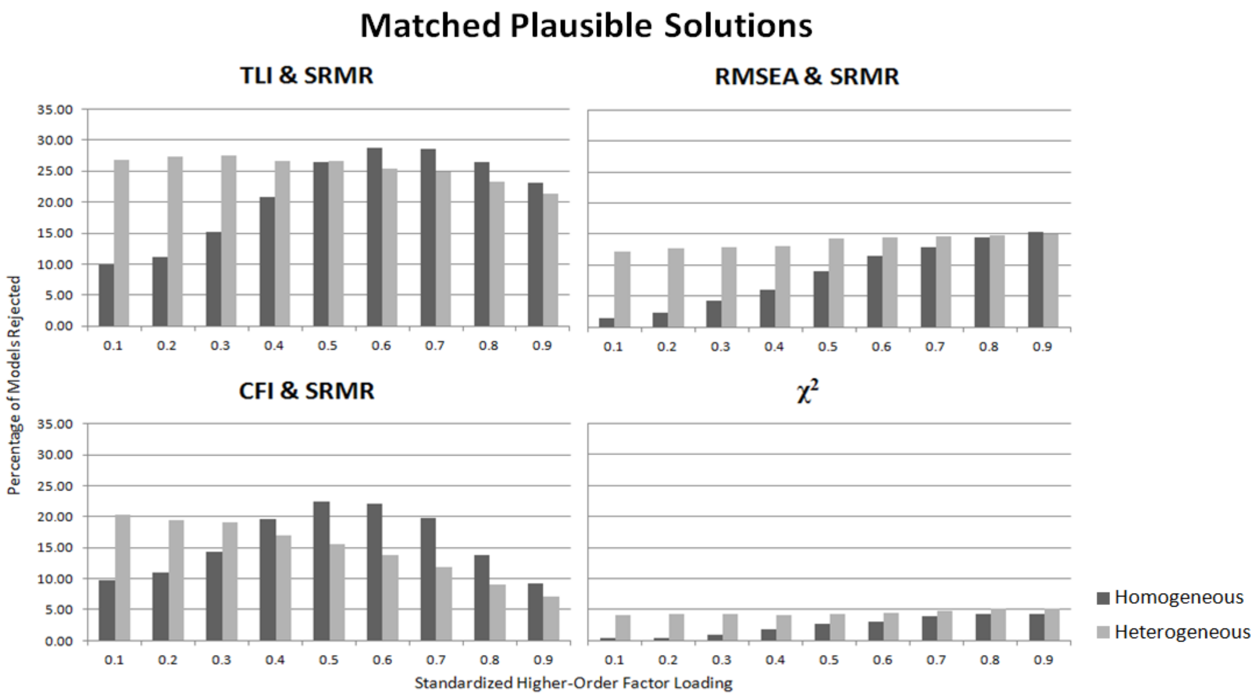


Figure 6. Percentage of Matched Plausible Solutions that were rejected by Hu and Bentler's (1999) Two-Index Presentation Strategies. TLI/SRMRS rejects models when TLI ≤ 0.96 or SRMR ≥ .09. RMSEA/SRMR rejects models when RMSEA ≥ 0.06 or SRMR ≥ .09. CFI/SRMR rejects models when CFI ≤ 0.96 or SRMR ≥ .09. χ^2 index rejects models when $p \leq .05$.

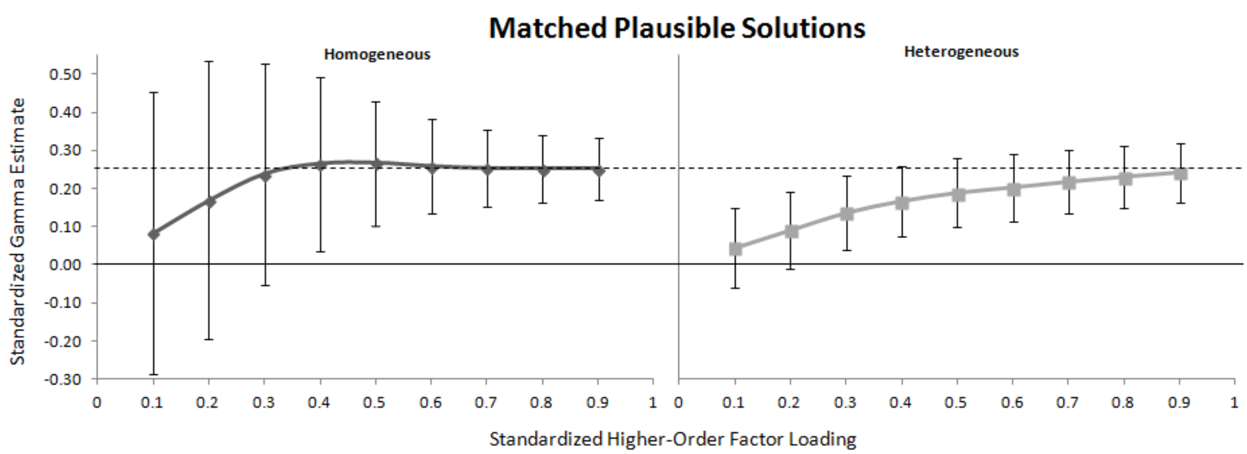


Figure 7. Mean standardized gamma estimates for Matched Plausible Solutions as function of higher-order factor loadings and parceling strategy. Error bars depict the mean standard error associated with that particular condition. The solid black line indicates a standardized gamma estimate of 0. The dotted line indicates 0.25, the population value of gamma.

Table 7

Mixed ANOVA Results: Significance Testing and Effect Sizes of Gamma Estimates and Standard Error of Gamma Estimates

Variable	Source	Design	F	*p-value*	ω^2
Gamma	Parcel	Within	2556.02	0.000	0.15
	FactorLoad	Between	340.16	0.000	0.17
	Parcel * FactorLoad	Within	94.33	0.000	0.04
SE of Gamma	Parcel	Within	1335.62	0.000	0.08
	FactorLoad	Between	188.89	0.000	0.10
	Parcel * FactorLoad	Within	140.91	0.000	0.07

Note. SE = standard error.

mean γ estimates and average SE of γ estimates for the Converged Plausible Solutions are very similar to the Matched Plausible Solutions. Table 7 summarizes the results of mixed-ANOVAs and associated effect sizes. All main effects and interactions were significant and had effect sizes that ranged from small to large. Gamma estimates were largely affected by parceling strategy and higher-order factor loading (the interaction effect was small). Higher-order factor loadings had a large effect on the SE of γ estimates, but parceling strategy and the interaction term only had medium effects.

Table 8 summarizes the results of dependent t-tests for γ estimates and their SEs between the parceling strategies for each of the higher-order factor loadings. All t-tests were statistically significant (using Bonferroni corrections). Effect sizes for γ estimates were medium to large when factor loadings were 0.3 or larger. The use of homogeneous parceling led to a large reduction in the bias of γ estimates when higher-order factor loadings were between 0.5 and 0.8. Effect sizes for the SE of γ estimates ranged from small to large. The use of heterogeneous parceling led to large effect sizes when higher-order factor loadings were 0.7 and 0.8, and medium effects when factor loadings were 0.3, 0.5, and 0.9.

Table 9 reports the Type I error, coverage (the probability of the parameter value being in the CI), and power for the Matched Plausible Solutions, as well as significance tests comparing parceling strategies at each factor loading level. When higher-order factor loadings were 0.5 or smaller, the homogeneous parceling strategy was more likely than the heterogeneous parceling strategy to result in Type I errors; furthermore, the rate of Type I errors was closer to 5%. The homogeneous parceling strategy also had better coverage when higher-order factor loadings were 0.6 or smaller; however, the large SEs that created the good coverage also led to poor power. When higher-order factor loadings were 0.3 and 0.4, the power difference between homogeneous and heterogeneous parceling was about 0.18. In other words, heterogeneous parceling was 1.65 to 2.34 times more likely to reject the null hypothesis than was homogeneous parceling. That said, the power was still less than 0.5; in fact, power

Table 8

Dependent t-test Results: Significance Testing and Effect Sizes of Gamma Estimates and Standard Error of Gamma Estimates

Factor	Gamma			SE of Gamma		
Loading	*t*	*p-value*[a]	*Cohen's d*	*t*	*p-value*[a]	*Cohen's d*
0.1	6.68	0.000	0.17	15.42	0.000	0.40
0.2	14.45	0.000	0.37	15.42	0.000	0.40
0.3	22.96	0.000	0.59	22.65	0.000	0.58
0.4	28.28	0.000	0.73	17.67	0.000	0.46
0.5	31.74	0.000	0.82	19.53	0.000	0.50
0.6	33.57	0.000	0.87	15.46	0.000	0.40
0.7	36.69	0.000	0.95	35.88	0.000	0.93
0.8	31.71	0.000	0.82	45.83	0.000	1.18
0.9	19.83	0.000	0.51	29.90	0.000	0.77

[a]*p*-values reflect Bonferroni correction.
Note. Numbers in the "Factor Loading" column represent the higher-order factor loading used in the data generative model.

Table 9

Type I Error, Coverage, and Power Based on 95% Confidence Intervals for Matched Plausible Solutions

Factor	Type 1 Error		Coverage		Power		Coverage & Power[a]	
Loading	Hom	Het	Hom	Het	Hom	Het	Hom	Het
0.1	0.02	0.06 ***	0.80	0.51 *	0.03	0.08 ***	0.02	0.06 ***
0.2	0.03	0.08 ***	0.88	0.66 ***	0.05	0.18 ***	0.05	0.18 ***
0.3	0.03	0.10 ***	0.93	0.79 ***	0.13	0.30 ***	0.12	0.30 ***
0.4	0.05	0.10 ***	0.95	0.86 ***	0.27	0.44 ***	0.26	0.44 ***
0.5	0.06	0.09 ***	0.96	0.90 ***	0.45	0.55 ***	0.44	0.54 ***
0.6	0.07	0.08	0.94	0.92 **	0.60	0.65 ***	0.58	0.64 ***
0.7	0.07	0.07	0.94	0.93	0.71	0.72	0.68	0.71 ***
0.8	0.07	0.07	0.94	0.93	0.80	0.80	0.77	0.79 **
0.9	0.06	0.07	0.94	0.94	0.86	0.85	0.83	0.83

[a]The probability that a 95% confidence interval around the gamma estimates would include the parameter value of 0.25 but not zero.
Note. *p-value ≤ .05. **p-value ≤ .01. ***p-value ≤ .001. p-values reflect Bonferroni correction.

reached 0.8 for either parceling strategy only when higher-order factor loadings were 0.8 or 0.9. Again, not surprisingly, the coverage, power, and simultaneous occurrence of coverage and power increased as higher-order factor loadings became larger.

The "Coverage & Power" column in Table 9 identifies the probability that the 95% CI around the γ estimate included the parameter value of 0.25 while simultaneously excluding zero. The computed difference between the "Coverage & Power" and "Coverage" column can be interpreted as the percentage of models that would properly reject the hypothesis that $\gamma = 0$ and yet yield CIs that do not include the true value of γ. More than 3% of these cases occurred for homogeneous parceling models when higher-order factors were 0.9, compared with only 2.5% of heterogeneous parceling models. Thus, homogeneous models are slightly less likely to reject properly the null of $\gamma = 0$, and even when they do reject the null, they are slightly less likely to have a CI that includes the true value of γ.

Finally, it is important to compare the average SE of γ estimates to the SD of γ estimates. Specifically, the mean column of C6 should approximate the corresponding value in the SD column of Table C5 (both tables are in Appendix C). The homogeneous parceling strategy overestimated the SD of γ estimates, meaning the SEs were inflated when higher-order factor loadings were 0.5 or smaller. In contrast, the SD of γ estimates were slightly underestimated by the SE of γ estimates for the heterogeneous parceling strategy at all higher-order factor loadings and for the homogeneous parceling strategy at factor loadings of 0.6 or greater. If we considered higher-order loadings of 0.1 - 0.3 to be small, 0.4 - 0.6 to be medium, and 0.7 - 0.9 to be large, then the heterogeneous parceling strategy had less biased estimates of the SD of γ estimates when factor loadings were small and medium. When higher-order factor loadings were small, homogeneous parceling models overestimated the SD of γ estimates by an average of 0.083; whereas heterogeneous parceling models underestimated the SD of γ estimates by an average of 0.004. When higher-order factor loadings were medium, the bias in SD of γ estimates diminished for both parceling strategies. Specifically, the homogeneous parceling strategy overestimated the SD by an average of 0.15 and the heterogeneous parceling model underestimated the SD by an average of .001. When higher-order factor loadings were large, both parceling strategies underestimated the SD of γ estimates by an average of .002.

CHAPTER IV

Discussion

Four main topics are discussed below. The initial three topics reiterate the primary results of this study: model convergence, diagnostic information about converged models, and gamma estimates (and associated gamma statistics). These topics are explored further by discussing their practical implication for researchers, how they fit with the existing body of literature, and possible explanations for their existence. The last topic deals with the broader concept of parceling as a statistical tool and offers advice to researchers interested in utilizing homogeneous or heterogeneous parceling when their construct has a hierarchical structure.

Model Convergence

Results about model convergence supported my hypotheses. When higher-order factor loadings were small or medium, the homogeneous parceling models were less likely to converge than were heterogeneous parceling models. When higher-order factor loadings increased to 0.8 or 0.9, however, the additional specific variance in the latent variable extracted from the heterogeneous parcels were no longer helpful (hence equivalent convergence rates between parceling strategies). Researchers who decide to use homogeneous parceling should be aware that model with small factor loadings (i.e., 0.4 or smaller) have a 25% or greater chance of non-convergence compared to heterogeneous parceling models. Furthermore, homogeneous parceling models with very small higher-order factor loadings have virtually a coin-toss chance of model convergence. When higher-order factor loadings are 0.6 or larger, however, homogeneous parceling appears to result in reasonable convergence rates (i.e., greater than 97.9%). On the other hand, researchers who decide to use heterogeneous parceling need not be very concerned about model convergence. Even when higher-order factor loadings were 0.1, heterogeneous models converged 96.4% of the time.

Diagnostic Information about Converged Models

Four main findings emerged about model diagnostic problems and their resultant γ estimates. First, converged models for either parceling strategy had a 70% or greater likelihood of having a well-fitting model without negative variance estimates, even when higher-order factor loadings were 0.1. Second, Heywood cases were more likely to occur in homogeneous parceling models rather than heterogeneous parceling models, particularly when higher-order factor loadings were small. Third, γ estimates for models with Heywood cases, for both parceling strategies, were more biased than were γ estimates from plausible solutions (models without Heywood cases). Fourth, as higher-order factor loadings increased, the likelihood of Heywood cases decreased while the likelihood of poor model fit increased; consequently, the overall percentage of problematic models remained somewhat stable (approximately 22%-30% of the converged models). These findings are further explored below.

The first finding was that converged models have a 70% or greater likelihood of having plausible solutions and acceptable model fit indices than have a diagnostic problem (i.e., a Heywood case and/or having poor model fit). Although it is comforting to know that once a model converges, it has a 70% or greater chance of having an acceptable fit and in-range parameter estimates, one must consider convergence and diagnostic problems simultaneously. For example, when factor loadings were 0.1, converged homogeneous parceling models were 77.62% likely to have an acceptable fit and in-range parameter estimates; however, only 54.6% of the original 1,500 models converged in the first place. In the end, when higher-order factor loadings were 0.1, only 42.53% of homogeneous parceling models converged on a well-fitting model with plausible solutions. When higher-order factor loadings were small (0.1-0.3) homogeneous parceling models had an average of 45.63% diagnostic-free models, compared to an average of 69.31% for heterogeneous parceling models. The percentage of well-fitting models with plausible solutions increased as higher-order factor loadings increased for both parceling strategies. More heterogeneous parceling models than homogeneous parceling models had acceptable model fit with plausible solutions at every level of higher-order factor loadings. Nevertheless, researchers should still be aware that even when higher-order factor loadings are extremely large, both parceling strategies have more than a 20% chance of a properly specified structural model being rejected due to poor model fit or convergence on a Heywood case.

A second finding was that homogeneous parceling models were more likely to result in Heywood cases than were heterogeneous parceling models when higher-order factor loadings were less than 0.8 (Heywood cases did not occur when factor loadings were 0.8 or larger). Furthermore, Psi-Heywood cases, when the negative variance was associated with the error term of the dependent variable, only occurred in homogeneous parceling models. Over the years, researchers have cited various causes for Heywood cases: (a) outliers (Bollen, 1987), (b) empirical under-identification due to small factor loadings (Dillon, Kumar, & Mulani, 1987; Rindskopf, 1984), (c) structurally misspecified models (Dillon et al., 1987; Kolenikov & Bollen, 2012; van Driel, 1978), and (d) sampling fluctuations coupled with a near zero population error variance (Anderson & Gerbing, 1984; Chen, Bollen, Paxton, Curran, & Kirby, 2001; Dillon et al., 1987; van Driel, 1978). For the current study, we can eliminate structural misspecification and near-zero error variances in the population as causes of these Heywood cases. Although the error variance associated with the sub-factors in our data-generative model became smaller when higher-order factor loadings were 0.9, the error variances associated with the parcels in the estimated models were not unreasonably close to zero.

For two reasons, outliers also seem like an unlikely cause of the Heywood cases. First, Heywood cases for homogeneous parceling models were more likely than for heterogeneous parceling models applied to the same samples. If outliers were the major cause of Heywood cases, parceling strategies would have resulted in Heywood cases at approximately the same rate. Second, further inspection of a handful of samples that resulted in Heywood cases did not appear to have z-scores that were more extreme than samples that did not result in a Heywood case[4].

In the end, empirical under-identification due to very small factor loadings seems like the most likely cause of Heywood cases in this study (and likely the cause of model convergence problems as well). Specifically, the combination of the sampling variance and only three parcels loading into the LV likely problems for the model. As the higher-order factor loadings approach zero, sampling variance likely caused some sample factor loadings to become zero. Because only three MVs loaded onto the LV, if any one of the MVs had a factor loading of zero the LV would be empirically under-identified. In the future, it would interesting to have a DG model that allows for more the creation of more parcels in the estimation models. It is conceivable that had more parcels been created, homogeneous parceling models may have had better convergence rates and fewer Heywood cases. Empirical under-identification explains the increase in Heywood cases when homogeneous parcels were used and when higher-order factor loadings decrease. As noted above, homogeneous parcels do not share specific variance leading to proportionally less common variance in the extracted latent variable, contributing to smaller factor loadings compared to heterogeneous parceling models. As the higher-order factor loadings became smaller, the homogeneous parceling models were more likely than the heterogeneous parceling models to become empirically under-identified and consequently result in Heywood cases.

Note that Heywood cases are not intrinsically bad. In situations where estimates of the structural coefficients from models containing Heywood cases are not statistically different from plausible solution estimates, the existence of a Heywood case may not be a considered a diagnostic problem and parameter estimates may be as interpretable as results from a plausible solution (Nasser & Wisenbaker, 2003). In short, the decision to interpret results with a Heywood case depends largely on whether key parameter estimates are biased or unreliable. This leads to the third finding: models with Heywood cases appeared to yield more biased γ estimates compared to models without Heywood cases. This was true for both parceling strategies and across all higher-order factor loadings. In other words, this study suggests that researchers who parcel items from a hierarchical factor model should

[4] Bollen (1987) suggested a variety of procedures to identify outliers, some included multivariate methods. This paper used only bivariate methods.

refrain from interpreting structural path coefficient estimates from models with Heywood cases, no matter which parceling strategy is chosen.

The fourth finding was that although percentages of converged models increase as higher-order factor loadings increase, the percentage of converged models that had acceptable fit and plausible solutions remained somewhat stable across higher-order factor loadings. In other words, the likelihood of converging on a Heywood case decreased as the higher-order factor loadings increased; however, the likelihood of poor model fit increased – essentially balancing each other out. This combination of trends was particularly apparent for the homogeneous parceling models, and was less evident for heterogeneous parceling models with large higher-order factor loadings. The percentage of heterogeneous parceling models that converged and had good model-fit and plausible solutions seemed to increase when higher-order factor loadings were very large. Although, it is important to remember that the total percentage of diagnostic-free models (out of the 1,500 replications) increased for both parceling strategies as the higher-order factor loadings increased. Furthermore, the heterogeneous parceling method consistently led to more well-fitting models with plausible solutions than did the homogeneous parceling method, at least for conditions where the higher-order factor loadings were not large.

Model-Fit Indices

There were three main findings about model fit indices. First, when higher-order factor loadings were small, homogeneous parceling models had a slightly but significantly better model fit than did heterogeneous parceling models for χ^2, RMSEA, and SRMR indices. Second, when higher-order factor loadings were medium in strength, heterogeneous parceling models led to a small but statistically significant improvement in model fit (over homogeneous parceling models) for the CFI. Third, the TLI often had a significantly better model fit for homogeneous parceling models than heterogeneous parceling models, although effect sizes fell short of the criteria for being considered "small." These results are discussed below.

The first finding supported my hypothesis that homogeneous parceling models would have better model fit than heterogeneous parceling models, especially when higher-order factor loadings were smaller. Specifically, parceling strategy choice had a small impact on the χ^2, RMSEA, and SRMR indices when higher-order factor loadings were 0.3 or smaller. Cole and Preacher (2014) emphasized that model fit will improve as the observed covariances approach zero. In the current data generative model, having smaller factor loadings generally results in smaller covariances, making it easier for any model (including the current one) to fit the data. As my original hypothesis stated that all model fit indices would favor homogeneous parceling models, the next two findings were unexpected. The second finding was that the CFI favored heterogeneous over homogeneous parceling models, especially when higher-order factor loadings were medium in strength. Also surprising was the (third) finding that the TLI was only negligibly impacted by parceling strategy. Perhaps the CFI and TLI had unexpected results because they are both incremental fit indices, whereas the other model fit indices in this paper are not. Marsh, Hau, & Grayson (2005) noted that when the null model is able to fit the data or the estimated model is saturated the TLI can be become unstable. These two circumstances were more likely to occur in the current study when homogeneous parcels were used and when higher-order factor loadings were small. Previous simulation studies have also reported TLIs with unexpectedly large values and large SDs (Anderson & Gerbing, 1984; Marsh et al., 2005; e.g., Sterba & MacCallum, 2010), although none had values as extreme as those in the current study. Investigating the sometimes anomalous behavior of the CFI and TLI would be an interesting avenue for future research.

Gamma Estimates, Standard Error of Gamma Estimates, Coverage, Power, and Type I Error Rates

Three major findings summarize the contribution of this paper to our knowledge about the impact of parceling strategy on γ estimates and associated statistics. First, when higher-order factor loadings were 0.3 or larger, homogeneous parceling models generated estimates of γ that were less biased but had larger SEs compared to heterogeneous parceling models. Second, 95% CIs around γ estimates were generally more likely to include the parameter value of 0.25 without including zero for heterogeneous parceling models than for homogeneous parceling models. Third, Type I errors regarding the significance of γ were less likely for homogeneous parceling models than for heterogeneous parceling models (at least when higher-order factor loadings were small or medium).

The first finding pertains to γ estimates and their SEs. Compared to heterogeneous parceling models, homogeneous parceling models were significantly less biased across all higher-order factor loadings. This difference was especially large when higher-order factor loadings were 0.3 or larger. This advantage, however, was offset by the fact that homogeneous parceling models had significantly larger SEs around gamma, compared to heterogeneous parceling models.

The second finding pertains to power and coverage. In general, heterogeneous parceling models were more likely to generate confidence intervals that simultaneously included the population value of .25 for gamma and excluded zero, compared to homogeneous parceling models. In other words, heterogeneous parceling models simultaneously had the better combination of coverage and power. These two findings taken together indicate that although the homogeneous parceling strategy will give unbiased estimates of γ in the long run, any homogeneous parceling has a poorer combination of power and coverage than heterogeneous parceling models. When factor loadings are small to medium in size, the use of heterogeneous parceling could increase power and coverage by 18.7% over homogeneous parceling. Interestingly, neither parceling strategy reached 80% power and coverage until higher-order factor loadings were 0.9. Even at moderate higher-order factor loadings of 0.5, both parceling strategies were essentially a coin-flip in terms of power. As γ plays an essential part in the power of this model, future research into the impact of γ on power in hierarchical factor models is encouraged.

The third finding was that homogeneous parceling models tended to be more conservative in terms of Type I error rates. Homogeneous parceling models were overly conservative (failing to reject more than 5% of models) when higher-order factor loadings were 0.3 or smaller. In contrast, heterogeneous parceling models generally had larger Type I error rates, rejecting the null hypothesis that $\gamma = 0$ at a rate greater than that implied by a nominal alpha level of .05. Overall, homogeneous parceling models had Type I error rates closer to the expected 5% value.

Homogeneous versus Heterogeneous Parceling Strategies: The Researcher's Decision

The goal of this paper was to explore how a researcher's decision to choose one parceling strategy over another, when the true model was hierarchical in structure, could affect model convergence rates, diagnostic information gathered from converged models, model fit indices, and γ estimates. One of the main points made in the parceling literature is that the decision to parcel, and ultimately which parceling strategy to use, should be decided based on the nature of the researcher's question and the true structure of the underlying LV. Decisions about whether to use homogeneous or heterogeneous parceling should always be made a priori based on the research question and the researcher's conceptualization of the LV.

The parceling literature has often denigrated heterogeneous parceling as generating biased estimates of structural parameters and yielding better-than-deserved model fit (Bandalos, 2002; Hall et al., 1999; Marsh et al., 2013). The conclusions of these articles emphasize that the decision to parcel ought not to be taken likely and that homogeneous parcels should be chosen over heterogeneous parcels. The main argument is that the LV extracted from homogeneous parcels contains only the true variance; whereas the LV extracted from heterogeneous parcels contains true and specific variance (Gibson, 2012; Little et al., 2002, 2013; Marsh et al., 2013). This argument, however, ignores the real-world applications of SEM.

The following is an example of how heterogeneous parcels can be more representative of the intended construct than many quantitative methodologists have suggested. Consider a researcher interested in capitalizing on the advantages of SEM, who is studying the construct of depression in relation to other constructs. Let us narrow our focus to only the measure of depression for the moment. Imagine true depression has a hierarchical structure (like the independent variable in Figure 1) where the sub-factors A, B, and C are *Affect, Behavior,* and *Cognition.* The researcher obtains multiple measures of depression from which to extract a latent depression factor. Most opponents of parceling would approve this scenario despite the fact that each of these measures is essentially a heterogeneous parcel! For example, the Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) contains items about affect, behavior, and cognition. Assuming that the four measures are congeneric measures of the intended construct, the extracted factor will represent the construct of interest. So, why is it that when a researcher can only afford the use of only one measure, parceling critics argue against heterogeneous parceling? Why is it acceptable to use a heterogeneously created total score in one scenario but not the other? The opinion of this author is that the use of heterogeneous parcels can be appropriate, depending upon the nature of the underlying construct and research question.

I believe it is also important to draw attention to the word “bias” used in this paper and other papers comparing homogeneous and heterogeneous parceling strategies. I chose to use the word “bias” with respect to γ estimates for ease of writing and comprehension; however, the word may be subtly inappropriate for use in relation to heterogeneous parceling models. Relative bias refers to the systematic over- or under-estimation of a population parameter, compared to another method of estimating *the same parameter*. If homogeneous and heterogeneous parceling strategies result in somewhat different latent variables, then the paths that lead to or away from them in SEM are also different. In this study, the LV extracted from homogeneous parcels contains only true variance, just like the LV represented in the DG model; thus, the population value of γ is 0.25 in homogeneous parceling models. In contrast, the LV extracted from heterogeneous parcels contains true and specific variance, which means the population value of γ is likely something different from 0.25 (probably something smaller than 0.25). Consequently, bias for heterogeneous parceling models should be regarded as the systematic mis-estimation of this unknown value, not 0.25.

In the end, the decision about which parceling strategy to use should be based on the LV that the researcher wishes to extract. If the goal is to extract only the true variance of the construct and to partition out the specific variance of the sub-factors, then homogeneous parceling is the correct strategy. If the goal is to extract a latent variable that contains both true variance and the specific variance of the sub-factors, then heterogeneous parcels are more appropriate. After making this decision, the results of the current study can alert researchers to specific kinds of problems they may expect (see Table 10 for study summary).

Limitations and Future Directions

Limitations of this study offer opportunities for future research. The first limitation echoes a limitation inherent to all simulation studies: generalizability. In this case, it is unclear how generalizable the results are to other hierarchical models. Future directions could be to explore how homogeneous and heterogeneous parceling models differ under other hierarchical structures (e.g., altering lower-order factor loadings, using a more complex structural model when generating data). Second, over half of the originally generated samples failed to converge on plausible solutions for both parceling strategies when higher-order factor loadings were small. Samples that converge for only one parceling strategy may quantitatively and qualitatively different than samples that converged for both; thus causing systematic unmatching. This study chose to continue generating samples until each condition had the same number of cases; however, other methodological strategies or sophisticated statistical analyses maybe imposed in the future.

Table 10

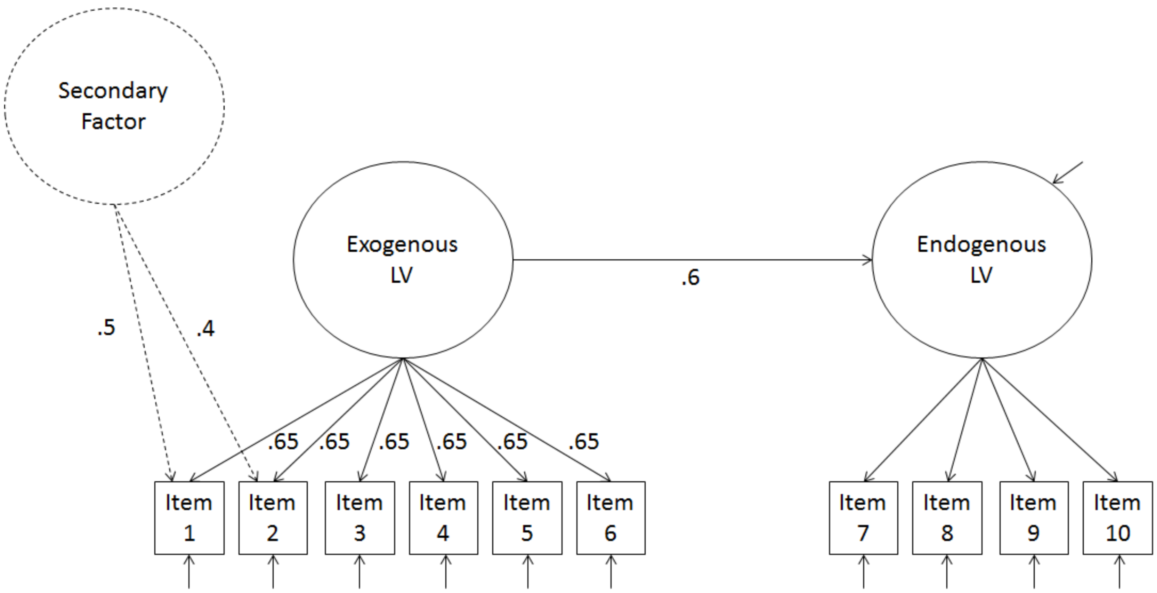
Once Parceling Strategy is Chosen A Priori What Are the Consequences?

	Homogeneous Parceling Strategy	Heterogeneous Parceling Strategy
Model Convergence	Poor model convergence when higher-order factor loadings are small and medium.	Poor model convergence when higher-order factor loadings are small, but convergence is more likely than homogeneous parceling models.
Diagnostic Problems	Theta- and Psi-Heywood cases are possible when higher-order factor loadings are small to medium. Models tended to have slightly better χ^2, RMSEA, and SRMR compared to heterogeneous parceling models, when higher-order factor loadings were small. Models tended to have marginally better TLI across all higher-order factor loadings.	Theta-Heywood cases are possible when higher-order factor loadings are small to medium, but are nearly half as likely to occur compared to homogeneous parceling models. Psi-Heywood case did not occur. Models tended to have slightly better CFI, compared to homogeneous parceling models, when higher-order factor loadings were medium.
Gamma Estimates	Gamma estimates are less biased, but have large SEs when higher-order factor loadings are small. Simultaneous power and coverage is smaller than heterogeneous parceling models, but increase as higher-order factor loadings increase. Type I Error is smaller than 5% when higher-order factor loadings are small, larger than 5% when factor loadings are medium to large, and are always smaller than heterogeneous parceling models.	Gamma estimates are more biased across all higher-order factor loadings; however, gamma estimates have smaller SEs, especially when higher-order factor loadings are large. Simultaneous power and coverage is larger than homogeneous parceling models. Type I Error is consistently larger than 5%.

APPENDIX A

Models from Previous Research

Model 1: Secondary Factor Unrelated to Endogenous LV



Model 2: Secondary Factor Predicts to Endogenous LV

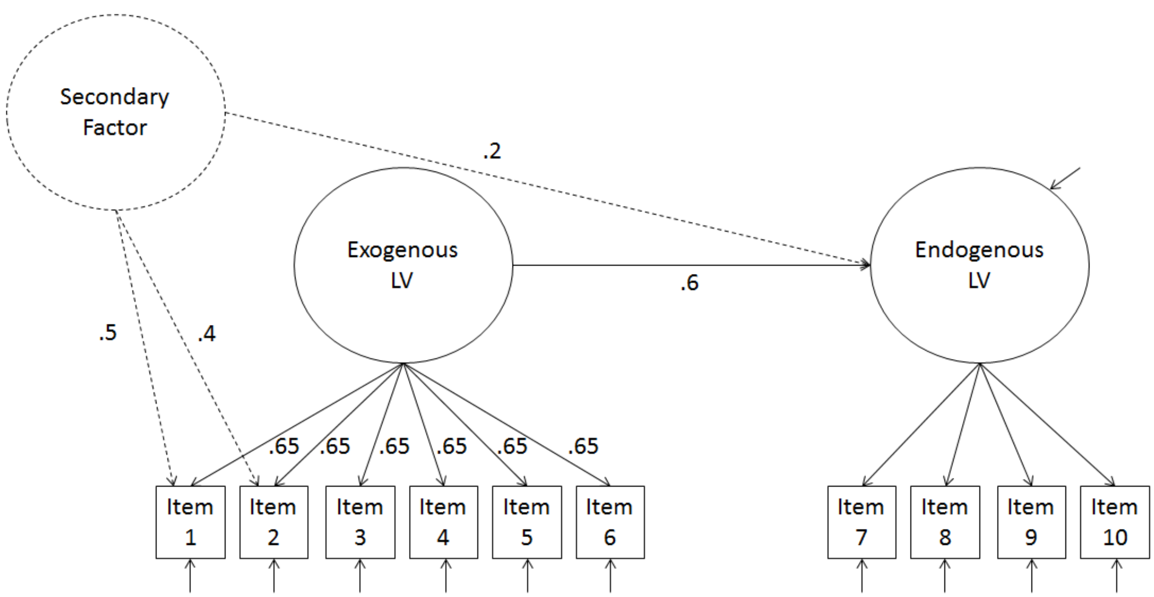


Figure A1. Hall et al. (1999) data generative model 1. Dashed secondary factor and associated paths were not included in the estimated models. Homogeneous parceling strategy combined Items 1 and 2 into a single parcel. Heterogeneous parceling strategy split Items 1 and 2 into separate parcels. LV= latent variable.

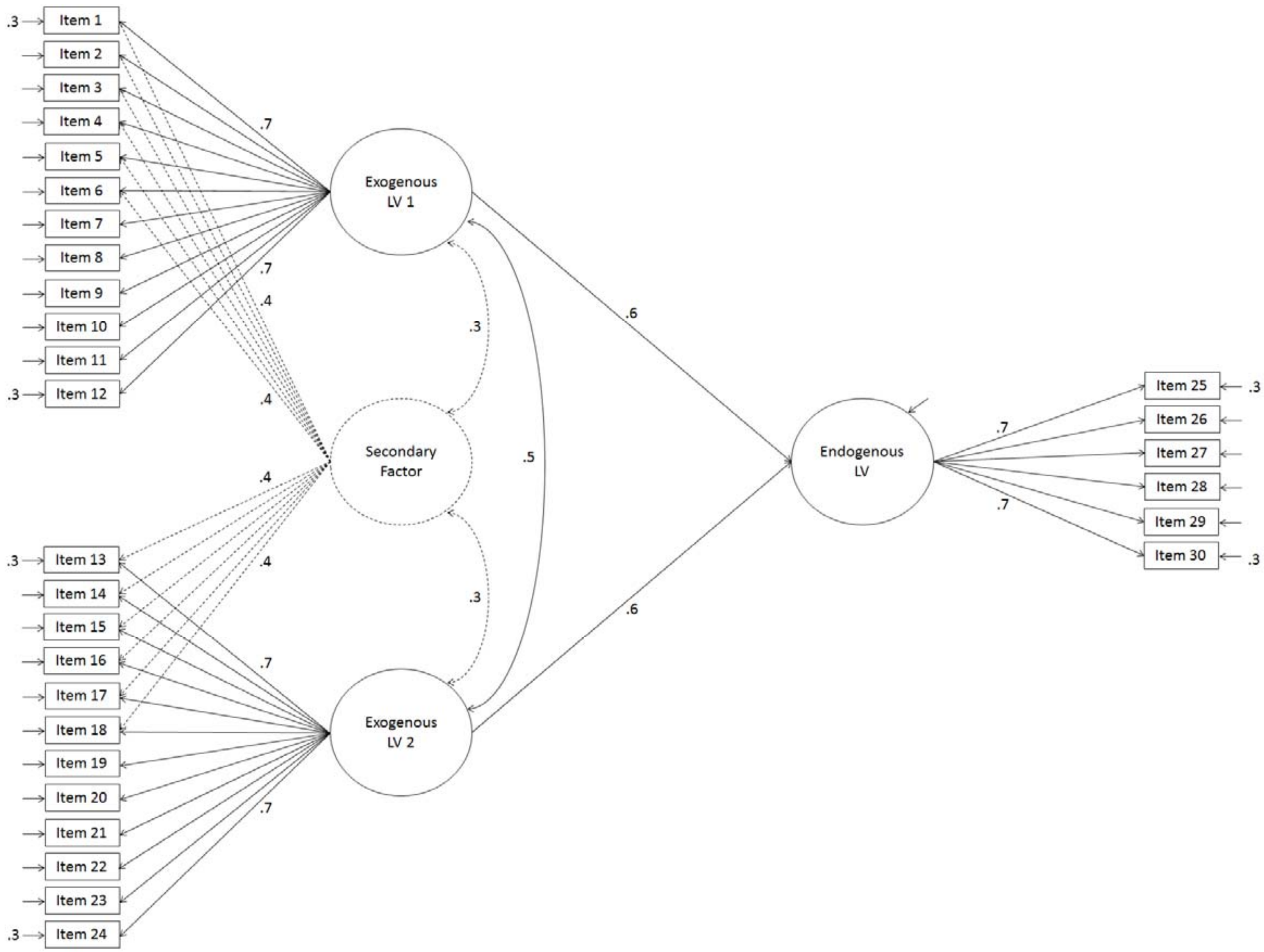


Figure A2. Bandalos (2002) data generative model for Study 2. Dashed secondary factor and associated factor loadings were not included in the estimated models. LV= latent variable.

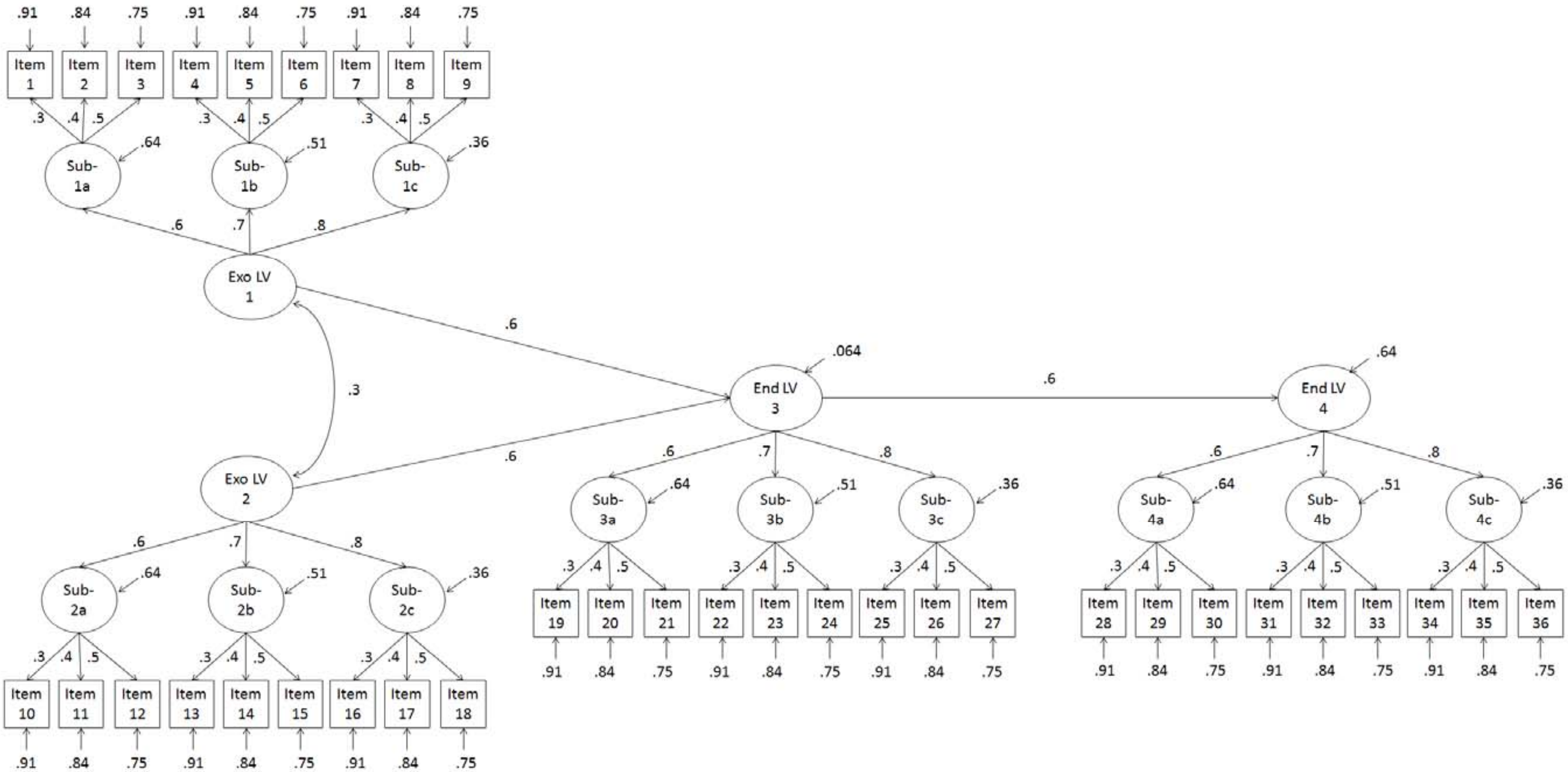
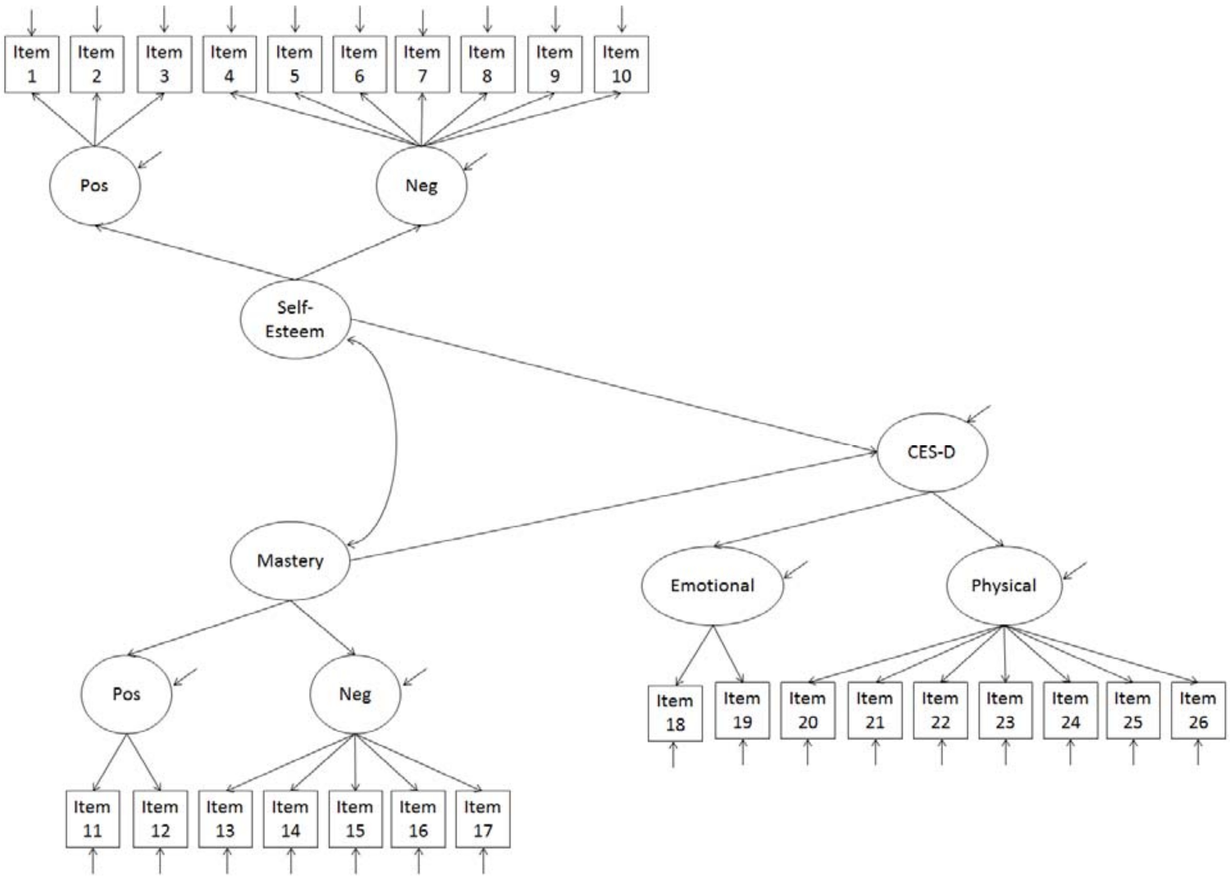


Figure A3. Coffman & MacCallum (2005) data generative model. Exo = exogenous. End=endogenous. LV= latent variable.

Model 1



Model 2

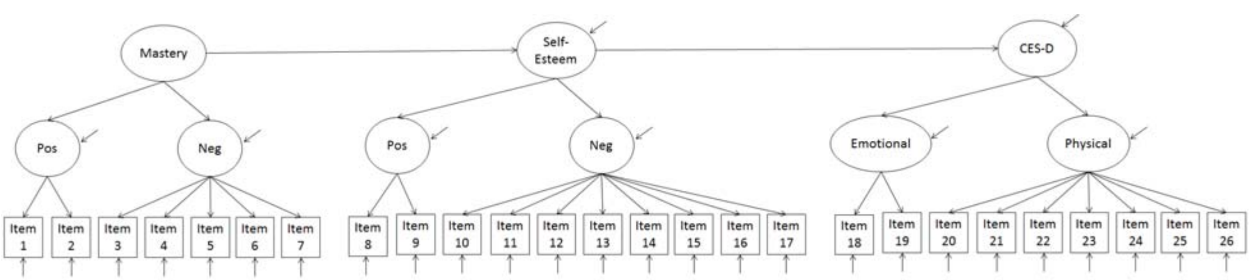


Figure A4. Coffman and MacCallum (2005) models for empirical data. Pos=positively worded items. Neg=negatively worded items. CES-D=Center for Epidemiological Studies-Depression.

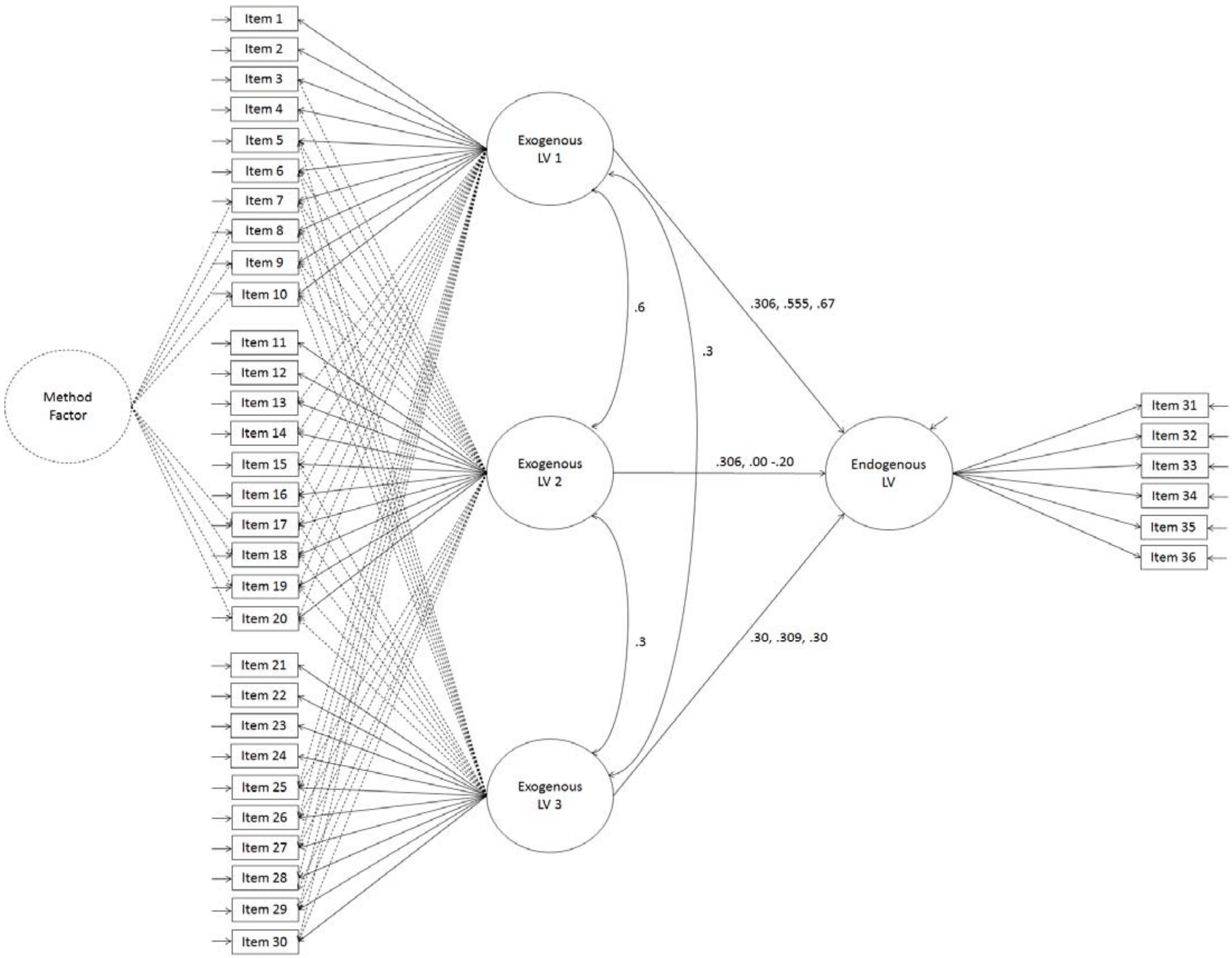


Figure A5. Marsh et al. (2013) data generative model for Study 4. Dashed secondary factor and associated factor loadings were not included in the estimated models. Structural paths from exogenous latent variables (LVs) to the endogenous LV indicate path coefficients for three simulated situations respectively.

Table A1

Factor Loadings for Marsh et al. (2013) Data Generative Model in Study 4

Item	Exogenous LV 1	Exogenous LV 2	Exogenous LV 3	Endogenous LV	Method Factor	Residual
1	.70	.00	.00			.508
2	.70	.00	.00			.509
3	.70	.20	.00			.401
4	.70	.20	.00			.357
5	.70	.20	.10			.375
6	.70	.20	.10			.378
7	.60	.20	.10		.20	.128
8	.60	.30	.10		.20	.220
9	.50	.30	.10		.20	.496
10	.50	.30	.10		.20	.370
11	.00	.70	.00			.509
12	.00	.70	.00			.511
13	.20	.70	.00			.302
14	.20	.70	.00			.356
15	.20	.70	.10			.377
16	.19	.70	.10			.298
17	.20	.60	.10		.20	.460
18	.30	.60	.10		.20	.393
19	.30	.50	.10		.20	.160
20	.30	.50	.10		.20	.372
21	.00	.00	.70			.510
22	.00	.00	.70			.513
23	.00	.00	.70			.506
24	.00	.00	.70			.509
25	.10	.10	.70			.337
26	.10	.10	.70			.397
27	.10	.10	.60			.569
28	.10	.10	.60			.538
29	.10	.10	.50			.681
30	.10	.10	.50			.682
31				.80		.357
32				.80		.358
33				.80		.358
34				.80		.357
35				.80		.360
36				.80		.360

Note. LV=latent variable

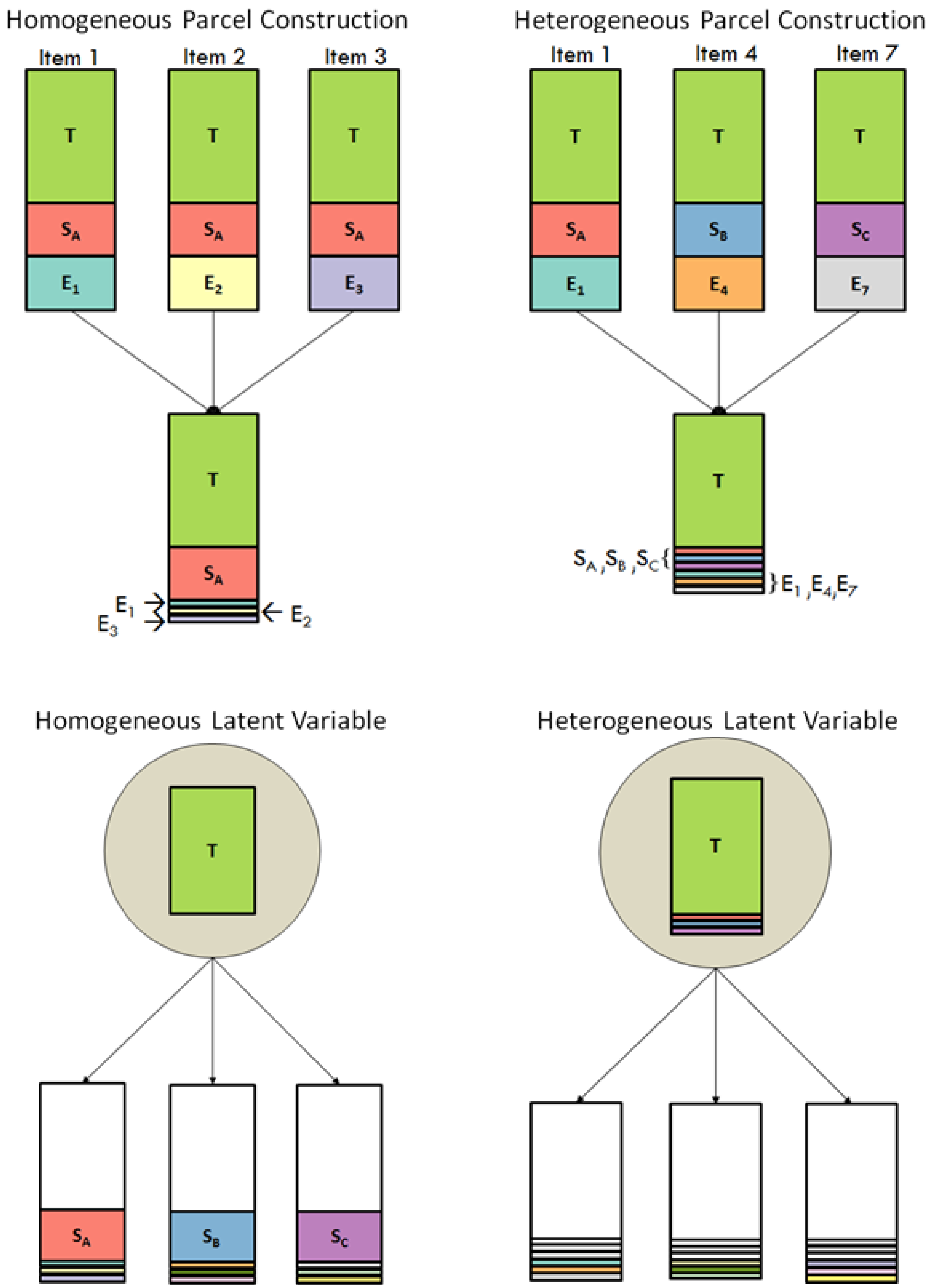


Figure A6. Illustration of variance of homogeneous and heterogeneous parcels as well as the latent variable represented by these same parceling strategies. Adapted from Gibson (2012).

Appendix B

Mplus Syntax

The Mplus code for data generation and model estimation based on parceling strategy is below. The syntax for data generation seen below is appropriate when higher-order factor loadings are 0.1. Comments in the syntax, preceded by an explanation mark, identify lines that need to be changed as higher-order factor loadings change. Table B1 reports the unique variance for sub-factors A, B, and C (seen in Figure 1) as the higher-order factor loadings change.

Example of Mplus Syntax for Data Generative Model

```
TITLE: Monte Carlo Data Generative Model Higher-Order Factor Loadings 0.1

MONTECARLO:
      NAMES = y1-y10;
      NOBS = 500;
      NREPS = 1500;
      SEED = 5256856;
      REPSAVE = all;
      SAVE = zDG.1.*.dat;

MODEL MONTECARLO:

      f1 BY y1-y3*.4;
      f2 BY y4-y6*.4;
      f3 BY y7-y9*.4;

      y1-y9*.84;

      f4 BY f1-f3*.1;  !HIGER-ORDER FACTOR LOADINGS

      f1-f3*.99;       !UNIQUE VARIANCE OF SUB-FACTORS A, B, AND C IN FIGURE 1

      f4@1;

      y10 ON f4*.25;

      y10*.9375;

ANALYSIS: TYPE = BASIC;

OUTPUT: TECH9;
```

Example of Mplus Syntax for Homogeneous Parceling Model Estimation

```
TITLE: Estimating the Model Using HOMOGENEOUS Parcels when Factor Loadings were 0.1

DATA: FILE IS zDG.1.list.dat;
TYPE IS MONTECARLO;
DEFINE: par1=(y1+y2+y3)/3;
        par2=(y4+y5+y6)/3;
        par3=(y7+y8+y9)/3;

VARIABLE:  NAMES ARE y1-y10 par1-par3;
           USEVARIABLES ARE y10 par1-par3;

MODEL:

     f4 BY  par1*(lpar1)
            par2*(lpar2)
            par3*(lpar3);

     par1*(epar1);
     par2*(epar2);
     par3*(epar3);

     f4@1;

     f5 BY y10*; y10@0;

     f5 ON f4*(g);

     f5*(ey10);


MODEL CONSTRAINT:
0 = 1 - g^2 - ey10;


OUTPUT: TECH9;
SAVEDATA:   RESULTS ARE Results.HOM.1.txt;
```

Example of Mplus Syntax for Heterogeneous Parceling Model Estimation

```
TITLE: Estimating the Model Using HETEROGENEOUS Parcels when Factor Loadings were 0.1

DATA: FILE IS zDG.1.list.dat;
TYPE IS MONTECARLO;
DEFINE: par1=(y1+y4+y7)/3;
        par2=(y2+y5+y8)/3;
        par3=(y3+y6+y9)/3;

VARIABLE:  NAMES ARE y1-y10 par1-par3;
           USEVARIABLES ARE y10 par1-par3;

MODEL:

     f4 BY  par1*(lpar1)
            par2*(lpar2)
            par3*(lpar3);

     par1*(epar1);
     par2*(epar2);
     par3*(epar3);

     f4@1;

     f5 BY y10*; y10@0;

     f5 ON f4*(g);

     f5*(ey10);


MODEL CONSTRAINT:
0 = 1 - g^2 - ey10;


OUTPUT: TECH9;
SAVEDATA:   RESULTS ARE Results.HET.1.txt;
```

Table B1

Unique Variance of Sun-Factors A, B, and C as Higher-Order Factor Loadings Change

Factor Loadings	Unique Variance
0.1	0.99
0.2	0.96
0.3	0.91
0.4	0.84
0.5	0.75
0.6	0.64
0.7	0.51
0.8	0.36
0.9	0.19

Note: Numbers in the “Factor Loading” column represent the higher-order factor loading used in the data generative model.

Appendix C

Further Graphics

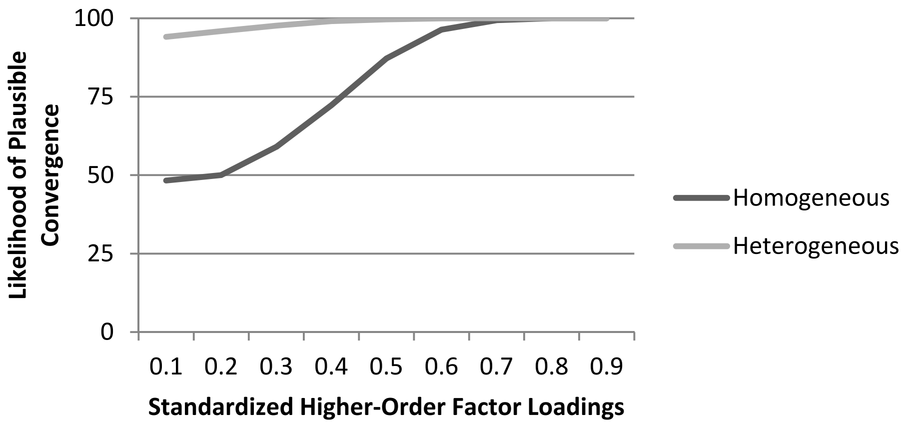


Figure C1. Percentage of plausible convergence across parceling strategies and changes to the higher-order factor loadings in the data generative model.

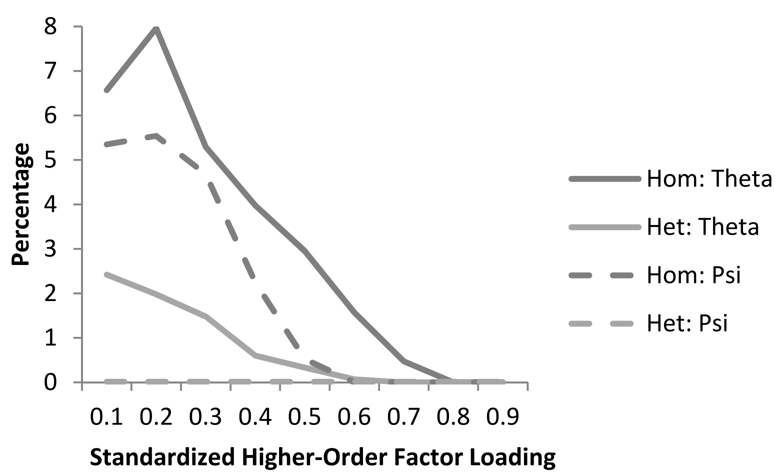


Figure C2. Percentage of Heywood cases across parceling strategies and changes to the higher-order factor loadings in the data generative model. Hom=homogeneous parceling strategy. Het=heterogeneous parceling strategy. Theta=negative unique variance associated with one of the parcels. Psi=negative unique variance associated with the dependent variable (Y).

Table C1

Frequency and Percentage of Model Rejection for Converged Models

Parceling Strategy	Factor Loading	Two-Index Presentation from Hu & Bentler: 1 N	1 %	2 N	2 %	3 N	3 %	Any N	Any %	χ^2 (*df*=2) N	χ^2 (*df*=2) %
Homogeneous	0.1	80	9.73	10	1.22	76	9.25	92	11.19	3	0.36
	0.2	91	10.50	16	1.85	86	9.92	99	11.42	4	0.46
	0.3	148	15.06	33	3.36	135	13.73	155	15.77	12	1.22
	0.4	232	20.05	62	5.36	215	18.58	240	20.74	18	1.56
	0.5	356	26.27	119	8.78	298	21.99	357	26.35	34	2.51
	0.6	425	28.93	172	11.71	324	22.06	426	29.00	45	3.06
	0.7	430	28.70	196	13.08	297	19.83	430	28.70	59	3.94
	0.8	398	26.53	216	14.40	207	13.80	398	26.53	65	4.33
	0.9	348	23.20	229	15.27	138	9.20	348	23.20	64	4.27
Heterogeneous	0.1	399	27.59	164	11.34	291	20.12	399	27.59	49	3.39
	0.2	410	27.93	179	12.19	287	19.55	410	27.93	52	3.54
	0.3	422	28.38	190	12.78	286	19.23	423	28.45	55	3.70
	0.4	416	27.83	206	13.78	279	18.66	416	27.83	60	4.01
	0.5	408	27.20	218	14.53	248	16.53	408	27.20	63	4.20
	0.6	390	26.00	221	14.73	215	14.33	390	26.00	69	4.60
	0.7	373	24.87	220	14.67	179	11.93	373	24.87	73	4.87
	0.8	350	23.33	221	14.73	136	9.07	350	23.33	76	5.07
	0.9	321	21.40	225	15.00	107	7.13	321	21.40	76	5.07

Note: Numbers in the "Factor Loading" column represent the higher-order factor loading used in the data generative model. Hu and Bentler's (1999) Two-Index Presentation Strategies rejects models when: (1) TLI ≤ 0.96 or SRMR ≥ .09; (2) RMSEA ≥ 0.06 or SRMR ≥ .09; (3) CFI ≤ 0.96 or SRMR ≥ .09. χ^2 index led to model rejection when $p \leq .05$.

Table C2

Frequency and Percentage of Converged Models Separated into Diagnostic Situations

Parceling Strategy	Factor Loading	No Diagnostic Problem N	No Diagnostic Problem %	Heywood Only N	Heywood Only %	Poor Model Fit Only N	Poor Model Fit Only %	Heywood & Poor Model Fit N	Heywood & Poor Model Fit %
Homogeneous	0.1	638	77.62	92	11.19	86	10.46	6	0.73
	0.2	665	76.70	103	11.88	85	9.80	14	1.61
	0.3	747	75.99	81	8.24	138	14.04	17	1.73
	0.4	863	74.59	54	4.67	222	19.19	18	1.56
	0.5	963	71.07	35	2.58	345	25.46	12	0.89
	0.6	1031	70.18	12	0.82	415	28.25	11	0.75
	0.7	1064	71.03	4	0.27	427	28.50	3	0.20
	0.8	1102	73.47	0	0	398	26.53	0	0
	0.9	1152	76.80	0	0	348	23.20	0	0
Heterogeneous	0.1	1027	71.02	20	1.38	384	26.56	15	1.04
	0.2	1040	70.84	18	1.23	399	27.18	11	0.75
	0.3	1052	70.75	12	0.81	413	27.77	10	0.67
	0.4	1073	71.77	6	0.40	413	27.63	3	0.20
	0.5	1088	72.53	4	0.27	407	27.13	1	0.07
	0.6	1110	74.00	0	0	389	25.93	1	0.07
	0.7	1127	75.13	0	0	373	24.87	0	0
	0.8	1150	76.67	0	0	350	23.33	0	0
	0.9	1179	78.60	0	0	321	21.40	0	0

Note. Numbers in the “Factor Loading” column represent the higher-order factor loading used in the data generative model.

Table C3

Difference Between Parceling Strategies: Model Fit Indices Descriptive Statistics for Matched Plausible Solutions

Factor	χ^2 *(df =2)*		RMSEA		TLI		SRMR		CFI	
Loading	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
0.1	-0.88	2.07	-0.01	0.03	2.14	28.76	0.00	0.01	-0.01	0.16
0.2	-0.85	2.12	-0.01	0.03	7.18	202.99	0.00	0.01	-0.02	0.17
0.3	-0.71	2.25	-0.01	0.04	3.49	51.75	0.00	0.01	-0.02	0.16
0.4	-0.51	2.33	-0.01	0.04	1.33	9.72	0.00	0.01	-0.03	0.15
0.5	-0.35	2.50	0.00	0.04	0.77	13.78	0.00	0.01	-0.02	0.10
0.6	-0.19	2.60	0.00	0.04	0.28	2.89	0.00	0.01	-0.02	0.08
0.7	-0.12	2.69	0.00	0.04	0.04	0.27	0.00	0.01	-0.01	0.06
0.8	-0.06	2.74	0.00	0.04	0.01	0.16	0.00	0.01	0.00	0.04
0.9	-0.04	2.76	0.00	0.04	0.00	0.11	0.00	0.01	0.00	0.03

Note. Difference is homogeneous parceling value minus heterogeneous parceling value. *M*= mean of differences. *SD* = standard deviation of differences.

Table C4

Correlation Between Parceling Strategies: Model Fit Indices Descriptive Statistics for Matched Plausible Solutions

Factor Loading	χ^2 (*df* =2)	RMSEA	TLI	SRMR	CFI
0.1	0.00	0.01	0.01	-0.02	-0.02
0.2	0.04	0.06	0.00	0.00	0.04
0.3	0.01	0.03	-0.02	-0.01	0.00
0.4	0.03	0.03	-0.01	0.02	0.02
0.5	0.01	0.02	0.04	0.01	0.05
0.6	0.01	0.02	0.04	0.01	0.03
0.7	0.01	0.02	0.03	0.02	0.03
0.8	0.01	0.02	0.03	0.03	0.03
0.9	0.01	0.02	0.03	0.03	0.03

Note: Numbers in the “Factor Loading” column represent the higher-order factor loading used in the data generative model.

Table C3

Descriptive Statistics for Gamma Estimates for Converged Models

	Plausible Solutions						Theta-Heywood					
Factor	Homogeneous			Heterogeneous			Homogeneous			Heterogeneous		
Loading	M	SD	Range	M	SD	Range	M	SD	Range	M	SD	Range
0.1	0.09	0.29	1.90	0.05	0.11	0.75	0.03	0.05	0.29	0.01	0.07	0.30
0.2	0.17	0.26	1.73	0.09	0.11	0.71	0.02	0.05	0.25	0.01	0.07	0.30
0.3	0.23	0.24	1.93	0.12	0.10	0.66	0.03	0.05	0.29	0.04	0.08	0.29
0.4	0.27	0.19	1.91	0.15	0.10	0.61	0.05	0.06	0.29	0.06	0.10	0.28
0.5	0.27	0.16	1.67	0.18	0.09	0.59	0.05	0.06	0.27	0.04	0.10	0.28
0.6	0.26	0.13	1.09	0.20	0.09	0.57	0.08	0.06	0.23	0.15[a]		
0.7	0.25	0.10	0.68	0.22	0.09	0.54	0.07	0.10	0.26			
0.8	0.25	0.09	0.59	0.23	0.08	0.52						
0.9	0.25	0.08	0.52	0.24	0.08	0.51						

[a]This value represents a single case.
Note. Numbers in the "Factor Loading" column represent the higher-order factor loading used in the data generative model.

Table C4

Descriptive Statistics for the Standard Error of Gamma Estimates for Converged Models

	Plausible Solutions						Theta-Heywood					
Factor	Homogeneous			Heterogeneous			Homogeneous			Heterogeneous		
Loading	M	SD	Range	M	SD	Range	M	SD	Range	M	SD	Range
0.1	0.39	0.70	7.80	0.10	0.02	0.41	0.31	0.26	1.32	0.11	0.07	0.29
0.2	0.34	0.44	4.17	0.10	0.03	0.91	0.28	0.21	1.12	0.12	0.08	0.28
0.3	0.29	0.33	4.13	0.10	0.01	0.19	0.32	0.25	1.32	0.12	0.07	0.24
0.4	0.24	0.33	5.97	0.10	0.01	0.13	0.27	0.25	1.36	0.10	0.04	0.12
0.5	0.17	0.18	4.35	0.09	0.01	0.09	0.21	0.15	0.62	0.10	0.07	0.16
0.6	0.12	0.10	2.58	0.09	0.01	0.06	0.24	0.15	0.49	0.11[a]		
0.7	0.10	0.02	0.44	0.08	0.01	0.05	0.21	0.27	0.75			
0.8	0.09	0.01	0.07	0.08	0.00	0.04						
0.9	0.08	0.01	0.04	0.08	0.00	0.03						

[a]This value represents a single case.
Note. Numbers in the "Factor Loading" column represent the higher-order factor loading used in the data generative model.

Table C5

Descriptive Statistics for Gamma Estimates for Matched Plausible Solutions

Factor Loading	Homogeneous *M*	Homogeneous *SD*	Homogeneous *Range*	Heterogeneous *M*	Heterogeneous *SD*	Heterogeneous *Range*	Difference[1] *M*	Difference[1] *SD*
0.1	0.08	0.28	1.96	0.04	0.11	0.69	0.04	0.23
0.2	0.17	0.27	1.89	0.09	0.11	0.67	0.08	0.21
0.3	0.24	0.23	1.93	0.14	0.10	0.62	0.10	0.17
0.4	0.26	0.19	1.91	0.17	0.09	0.61	0.10	0.13
0.5	0.27	0.15	1.67	0.19	0.09	0.58	0.08	0.10
0.6	0.26	0.13	1.09	0.20	0.09	0.57	0.06	0.06
0.7	0.25	0.10	0.68	0.22	0.09	0.54	0.04	0.04
0.8	0.25	0.09	0.59	0.23	0.08	0.52	0.02	0.03
0.9	0.25	0.08	0.52	0.24	0.08	0.51	0.01	0.02

[1]Difference is homogeneous parceling value minus heterogeneous parceling value.
Note. *M* = mean. *SD* = standard deviation. Numbers in the "Factor Loading" column represent the higher-order factor loading used in the data generative model.

Table C6

Descriptive Statistics for Standard Error of Gamma Estimates for Matched Plausible Solutions

Factor Loading	Homogeneous *M*	Homogeneous *SD*	Homogeneous *Range*	Heterogeneous *M*	Heterogeneous *SD*	Heterogeneous *Range*	Difference[1] *M*	Difference[1] *SD*
0.1	0.37	0.67	9.76	0.10	0.02	0.31	0.27	0.67
0.2	0.37	0.66	12.00	0.10	0.03	0.91	0.26	0.66
0.3	0.29	0.33	4.13	0.10	0.01	0.19	0.19	0.33
0.4	0.23	0.30	6.01	0.09	0.01	0.13	0.14	0.30
0.5	0.16	0.15	2.10	0.09	0.01	0.08	0.07	0.14
0.6	0.12	0.09	2.58	0.09	0.01	0.06	0.04	0.09
0.7	0.10	0.02	0.44	0.08	0.01	0.05	0.02	0.02
0.8	0.09	0.01	0.07	0.08	0.00	0.04	0.01	0.01
0.9	0.08	0.01	0.04	0.08	0.00	0.03	0.00	0.00

[1]Difference is homogeneous parceling value minus heterogeneous parceling value.
Note. Numbers in the "Factor Loading" column represent the higher-order factor loading used in the data generative model.

Table C9

Mean and Standard Deviations of Parcel Factor Loadings

		Parcel Factor Loading							
		1		2		3		Combined[1]	
Parceling Strategy	Factor Loading	M	SD	M	SD	M	SD	M	SD
Homogeneous	0.1	0.09	0.16	0.09	0.17	0.14	0.36	0.11	0.25
	0.2	0.11	0.15	0.12	0.15	0.17	0.36	0.13	0.24
	0.3	0.14	0.14	0.15	0.13	0.18	0.26	0.16	0.19
	0.4	0.17	0.12	0.18	0.12	0.20	0.19	0.18	0.15
	0.5	0.21	0.11	0.21	0.11	0.22	0.17	0.21	0.13
	0.6	0.24	0.09	0.24	0.10	0.25	0.13	0.25	0.11
	0.7	0.28	0.08	0.28	0.08	0.28	0.10	0.28	0.09
	0.8	0.32	0.07	0.32	0.07	0.32	0.07	0.32	0.07
	0.9	0.36	0.06	0.36	0.06	0.36	0.06	0.36	0.06
Heterogeneous	0.1	0.24	0.08	0.24	0.09	0.26	0.12	0.24	0.10
	0.2	0.25	0.08	0.25	0.08	0.26	0.11	0.25	0.09
	0.3	0.26	0.08	0.26	0.07	0.27	0.10	0.26	0.09
	0.4	0.27	0.07	0.27	0.07	0.27	0.07	0.27	0.07
	0.5	0.29	0.07	0.29	0.07	0.29	0.07	0.29	0.07
	0.6	0.31	0.06	0.30	0.06	0.30	0.06	0.30	0.06
	0.7	0.33	0.06	0.33	0.06	0.32	0.06	0.33	0.06
	0.8	0.35	0.06	0.35	0.06	0.35	0.06	0.35	0.06
	0.9	0.38	0.05	0.37	0.06	0.37	0.06	0.37	0.06

[1]Combined means that the three parcel factor loadings for each estimated model were averaged. The mean and SD of this new variable is reported.

Note. Numbers in the "Factor Loading" column represent the higher-order factor loading used in the data generative model.

Table C10

Difference Between Parceling Strategies: Combined Parcel Factor Loadings for Matched Plausible Solutions

Factor Loading	M	SD
0.1	-0.14	0.12
0.2	-0.12	0.11
0.3	-0.10	0.08
0.4	-0.09	0.05
0.5	-0.07	0.05
0.6	-0.06	0.03
0.7	-0.04	0.02
0.8	-0.03	0.01
0.9	-0.01	0.01

Note. Numbers in the "Factor Loading" column represent the higher-order factor loading used in the data generative model.

REFERENCES

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*(2), 155–173. doi:10.1007/BF02294170

Bagozzi, R. P., & Edwards, J. R. (1998). A General Approach for Representing Constructs in Organizational Research. *Organizational Research Methods*, *1*(1), 45–87. doi:10.1177/109442819800100104

Bagozzi, R. P., & Heatherton, T. F. (1994). A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modeling: A Multidisciplinary Journal*, *1*(1), 35–67. doi:10.1080/10705519409539961

Bandalos, D. L. (2002). The Effects of Item Parceling on Goodness-of-Fit and Parameter Estimate Bias in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(1), 78–102. doi:10.1207/S15328007SEM0901_5

Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. Marcoulidesd & R. Schumacker, *New developments and techniques in structural equation modeling* (pp. 269–296). Mahwah, N.J.: Lawrence Erlbaum Associates. Retrieved from http://search.proquest.com/psycinfo/docview/619580051/13E94D8B0FE1C7BF326/3?accountid=14816

Beck, A. T., Ward, C. T., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571. doi:10.1001/archpsyc.1961.01710120031004

Bollen, K. A. (1987). Outliers and Improper Solutions A Confirmatory Factor Analysis Example. *Sociological Methods & Research*, *15*(4), 375–384. doi:10.1177/0049124187015004002

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

Cattell, R. B. (1956). Validation and intensification of the sixteen personality factor questionnaire. *Journal of Clinical Psychology*, *12*(3), 205–214. doi:10.1002/1097-4679(195607)12:3<205::AID-JCLP2270120302>3.0.CO;2-0

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper Solutions in Structural Equation Models Causes, Consequences, and Strategies. *Sociological Methods & Research*, *29*(4), 468–508. doi:10.1177/0049124101029004003

Coffman, D. L., & MacCallum, R. C. (2005). Using Parcels to Convert Path Analysis Models Into Latent Variable Models. *Multivariate Behavioral Research*, *40*(2), 235–259. doi:10.1207/s15327906mbr4002_4

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, *19*(2), 300–315. doi:http://dx.doi.org/10.1037/a0033805

Dillon, W. R., Kumar, A., & Mulani, N. (1987). Offending estimates in covariance structure analysis. Comments on the causes of and solutions to Heywood cases. *Psychological Bulletin*, *101*(1), 126–135.

Gagne, P., & Hancock, G. R. (2006). Measurement Model Quality, Sample Size, and Solution Propriety in Confirmatory Factor Models. *Multivariate Behavioral Research*, *41*(1), 65–83. doi:10.1207/s15327906mbr4101_5

Gibson, K. A. (2012a). *Covariance algebra of parcels* (M.A.). University of Kansas, United States -- Kansas. Retrieved from http://search.proquest.com/docview/1033566790/abstract/1404976F94FD0AEC35/1?accountid=14816

Gibson, K. A. (2012b). *Covariance algebra of parcels* (M.A.). University of Kansas, United States -- Kansas. Retrieved from http://search.proquest.com/docview/1033566790/abstract/1404976F94FD0AEC35/1?accountid=14816

Gribbons, B. C., & Hocevar, D. (1998). Levels of aggregation in higher level confirmatory factor analysis: Application for academic self-concept. *Structural Equation Modeling: A Multidisciplinary Journal*, *5*(4), 377–390. doi:10.1080/10705519809540113

Hall, R. J., Snell, A. F., & Foust, M. S. (1999). Item Parceling Strategies in SEM: Investigating the Subtle Effects of Unmodeled Secondary Constructs. *Organizational Research Methods*, *2*(3), 233–256. doi:10.1177/109442819923002

Hau, K.-T., & Marsh, H. W. (2004). The use of item parcels in structural equation modelling: Non-normal data and small sample sizes. *British Journal of Mathematical and Statistical Psychology*, *57*(2), 327–351. doi:10.1111/j.2044-8317.2004.tb00142.x

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. doi:10.1080/10705519909540118

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

Kishton, J. M., & Widaman, K. F. (1994). Unidimensional Versus Domain Representative Parceling of Questionnaire Items: An Empirical Example. *Educational and Psychological Measurement*, *54*(3), 757–765. doi:10.1177/0013164494054003022

Kolenikov, S., & Bollen, K. A. (2012). Testing Negative Error Variances Is a Heywood Case a Symptom of Misspecification? *Sociological Methods & Research*, *41*(1), 124–167. doi:10.1177/0049124112442138

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: A Multidisciplinary Journal*, *9*(2), 151–173. doi:10.1207/S15328007SEM0902_1

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the Items Versus Parcels Controversy Needn't Be One. *Psychological Methods*. doi:http://dx.doi.org/10.1037/a0033266

MacCallum, R. C., & Austin, J. T. (2000). Applications of Structural Equation Modeling in Psychological Research. *Annual Review of Psychology*, *51*(1), 201–226. doi:10.1146/annurev.psych.51.1.201

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*(2), 181–220. doi:10.1207/s15327906mbr3302_1

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of Fit in Structural Equation Models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary Psychometrics*. Florence, KY: Psychology Press.

Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J. S., & Von Davier, M. (2013). Why Item Parcels Are (Almost) Never Appropriate: Two Wrongs Do Not Make a Right—Camouflaging Misspecification With Item Parcels in CFA Models. *Psychological Methods*. doi:http://dx.doi.org/10.1037/a0032773

Muthén, L. K., & Muthén, B. O. (2010). *Mplus Statistical Analysis with Latent Variables: User's Guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Rindskopf, D. (1984). Structural Equation Models Empirical Identification, Heywood Cases, and Related Problems. *Sociological Methods & Research*, *13*(1), 109–119. doi:10.1177/0049124184013001004

Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin*, *94*(1), 18–38. doi:10.1037/0033-2909.94.1.18

Sterba, S. K., & MacCallum, R. C. (2010). Variability in Parameter Estimates and Model Fit Across Repeated Allocations of Items to Parcels. *Multivariate Behavioral Research*, *45*(2), 322–358. doi:10.1080/00273171003680302

Tomarken, A. J., & Waller, N. G. (2003). Potential Problems With "Well Fitting" Models. *Journal of Abnormal Psychology*, *112*(4), 578–598. doi:http://dx.doi.org/10.1037/0021-843X.112.4.578

Van Driel, O. P. (1978). On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika*, *43*(2), 225–243. doi:10.1007/BF02293865