FLORIDA STATE UNIVERSITY

COLLEGE OF EDUCATION

USE OF ITEM PARCELING IN STRUCTURAL EQUATION MODELING WITH MISSING DATA

By

FATIH ORCAN

A Dissertation submitted to the
Department of Educational Psychology and Learning Systems
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Fall Semester, 2013

Fatih Orcan defended this dissertation on November 1, 2013.

The members of the supervisory committee were:

Yanyun Yang
Professor Directing Dissertation

Adrian Barbu
University Representative

Betsy Becker
Committee Member

Russell Almond
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with university requirements.

ACKNOWLEDGMENT

I would like to express my great appreciation to Dr. Yanyun Yang for her guidance and encouragement during my dissertation. She was patient and willing to give her time to guide me though this dissertation. I also would like to thank my dissertation committee members, Dr. Adrian Barbu, Dr. Betsy Becker, and Dr. Russell Almond for their valuable feedbacks on my dissertation. Finally, I would like to thank my family and friends for their support and praying throughout my study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Parceling is referred to as a procedure for computing sums or average scores across multiple items. Parcels instead of individual items are then used as indicators of latent factors in the structural equation modeling analysis (Bandalos 2002, 2008; Little et al., 2002; Yang, Nay, & Hoyle, 2010). Item parceling may be applied to alleviate some problems in analysis with missing data (e.g., MCAR, MAR, and MNAR) and/or nonnormal data. No simulation study has been conducted to examine whether using parceling leads to better (at least not worse) results than individual items when there are missing values and nonnormality issues in the dataset. The purpose of this study is to investigate how item parceling behaves under various simulated conditions in structural equation modeling with missing and nonnormal distributed data. The design factors of the simulation study included sample size, missingness mechanism, percentage of missingness, degree of nonnormality for individual items, and magnitude of factor loadings. For each condition, 2000 datasets were generated. Each generated dataset was analyzed at both parcel and item levels using full information maximum likelihood estimation method. All analysis models were considered correctly specified. The results of the simulation showed that models based on parcels were less likely to be rejected than those based on individual items. Specifically, parcel analysis tended to result in smaller empirical alpha based on the chi-square test, greater CFI, and smaller RMSEA and SRMR. In addition, Parameter estimates from parcel level analysis performed equally well or slightly better than those from item level analysis in all conditions. In general, parcel level analysis yielded results as good as or better than those from item level analysis under any type of missing mechanisms, different degrees of nonnormality of data, and percentage of missingness.

CHAPTER 1

INTRODUCTION

Data missingness is an unavoidable issue in Structural Equation Modeling (SEM) for most social science studies (Allison, 2002; Davey, Savla, & Luo, 2005; Enders & Bandalos, 2001). In the SEM literature, many missing data techniques have been proposed. Some of the techniques involve removing cases (e.g., Listwise Deletion [LD], Pairwise Deletion [PD]), some involve adding data (e.g., Mean Imputation, Similar Response Pattern Imputation [SRPI]), while others use only available information in the dataset (e.g., Full Information Maximum Likelihood [FIML]).

All of the techniques listed above assume that data are missing randomly (i.e., missing at random or missing completely at random; more details later) (McKnight, McKnight, Sidani, & Figueredo, 2007). However, some techniques are better than the others in terms of yielding smaller parameter estimation bias or higher model convergence rates (Enders et al., 2001). Among these techniques, FIML could be considered as the best in general (Enders et al., 2001). However, FIML requires all assumptions underlying the maximum likelihood (ML) method. If any of the assumptions (e.g., multivariate normality) is not met, the results from ML may be biased, and the type I error rate for statistical testing may be inaccurate (e.g., chi-square test for evaluation of model-data fit may be inflated) (Enders, 2006). These conclusions are also applicable to FIML with missing data. For example, Enders (2001) found that FIML for data with missing values behaved similarly to ML (i.e., for a complete dataset) in terms of rejection rate and parameter estimation. Specifically, the rejection rate of the model-data fit test was inflated and standard errors of parameter estimates were negatively biased.

Parceling is one of the most commonly used modeling techniques in SEM. It uses the sum or average score of a subset of items instead of individual items as measured indicators for the latent factors in the analysis. Bandalos and Finney (2001) reviewed the literature on the use of item parceling in five journals published in 1989-1994. They found that about 20% of empirical studies (62 out of 317) used some kind of item parceling techniques. One of the most prominent advantages of using item parcels is that parcel scores tend to be more normally distributed compared to individual item scores (Bandalos, 2002; Little, Cunningham, Shahar, & Widaman, 2002; Matsunaga, 2008).

Both missing data and nonnormality affect estimation procedures in SEM (e.g., Andreassen, Lorentzen, & Olsson, 2006; Finney & DiStefano, 2006). The presence of nonnormality could lead to more serious issues when missing values are observed in the dataset (Enders, 2001). Use of item parceling may not only obviate some difficulties in meeting normality assumptions (e.g., Bandalos, 2002; Carter, 2006; Little et al., 2002; Matsunaga, 2008), but may also overcome some effects of missingness in the data on the estimation procedure. Typically, parcel scores are computed as the sum or average across a subset of item scores. When missing data are present on one or more individual items, parcel scores can be computed as the average of a subset of *available* item scores. Consequently, no missing data appear in the dataset containing the parcel scores. To illustrate the procedure, assume that we have a dataset containing a set of items with sample size of 20. For simplicity, values from only four items are presented in Table 1. Two of the items, X1 and X2, have some missing values. If a Confirmatory Factor Analysis (CFA) model is conducted on this set of items, only 10 cases are available for analysis if listwise deletion is used. The CFA results will be affected by the presence of missing data, especially if the missing pattern is not at random (i.e., MNAR). However, if parcel level

data are used, it is expected that the impact of the missingness on the results be smaller (e.g., smaller parameter estimation bias, better model-data fit indices, etc.). For example, a parcel can be created by computing the mean across items X1 to X4. The parcel score is then the mean of the items X2, X3, and X4 for the 1st case and the mean of the items X1, X3, and X4 for the 6th case, etc. In other words, the parcel scores are based on all items without missing values within the parcel. As a result, no values are missing in the dataset when analyzing the assumed model.

From one point of view, handling the missing data through parceling as demonstrated above is an imputation method (i.e., adding data for missing values). One of the well-known

Table 1: *An Example of Creating Parcel Scores for Data with Missing Values*

ID	X1	X2	X3	X4	Parcel	Non-missing Parcel
1	NA (3)	4	3	3	3.33	3.25
2	2	4	3	2	2.75	2.75
3	NA (1)	3	4	2	3.00	2.50
4	3	1	2	2	2.00	2.00
5	NA (1)	2	3	3	2.67	2.25
6	0	NA (3)	1	1	0.67	1.25
7	1	2	1	1	1.25	1.25
8	2	2	2	2	2.00	2.00
9	NA (2)	NA (1)	2	2	2.00	1.75
10	NA (1)	2	2	1	1.67	1.50
11	3	0	2	3	2.00	2.00
12	4	1	1	2	2.00	2.00
13	1	2	1	0	1.00	1.00
14	2	NA (3)	2	1	1.67	2.00
15	NA (0)	2	1	2	1.67	1.25
16	2	3	2	2	2.25	2.25
17	3	1	1	3	2.00	2.00
18	NA (4)	1	2	4	2.33	2.75
19	1	1	4	1	1.75	1.75
20	4	NA (4)	2	3	3.25	3.00

Note. NA indicates missing value. Values in parentheses indicate values of the missed data.

imputation methods is mean imputation. Mean imputation method takes the average of variable (e.g., X1 in Table 1), and then imputes this average value for all the missing cases on the variable. For creating parcel scores when missing data are present, the idea is to replace missing values with the average value within the case (Schafer & Graham, 2002). For example, consider the 3rd case in Table 1. The value for X1 is missing and the average of all other items in case 3 is [(3+4+2)/3] =3. With the missing value replaced by 3 for X1, the average score for this case retains the same [(3+3+4+2)/4] =3. In other words, the parcel score is based on all available data for a particular case.

Schafer and Graham (2002) raised an idea similar to item parceling when missing data are present but labeled it in an ambiguous manner. First, they suggested averaging scores across a subset of items when multiple items are available in measuring the same/similar construct (i.e., the same latent variable). "If a participant has missing values for one or more items, it seems more reasonable to average the items that remain rather than report a missing value for the entire scale" (p. 158). They suggested labeling this method as "*case-by-case item deletion*" and "*ipsative mean imputation*". This method was "widespread, but its properties remained largely unstudied; it does not even have a well-recognized name" although "the method can be reasonably well behaved" (Schafer et al., 2002, p. 158). Some authors have applied this method in their empirical studies (e.g., Achenbach, Bernstein, & Dumenci, 2005; Signorella, & Cooper, 2011; Yoder, Snell, & Tobias, 2012) without discussing its pros and cons (Schafer et al., 2002). To the best of my knowledge, no previous study has examined the effects of the use of item parceling on SEM analysis (e.g., parameter estimate, model-data fit) with nonnormally distributed incomplete data. The purpose of this study is to investigate how item parceling behaves in SEM analysis with missing data under a variety of simulation conditions. Because

parceling not only results in complete data for analysis but also tends to yield more normally distributed data, I expect that item parcels will yield results at least as good as those from item level data under all three missing data mechanisms (i.e., MCAR, MAR, and MNAR) with smaller parameter estimation bias or more accurate model rejection rates.

CHAPTER 2

LITERATURE REVIEW

In this chapter, principals and fundamentals of SEM are first briefly introduced. The use of item parceling techniques in SEM are then reviewed, followed by review of missing data mechanisms and data handling techniques. Last, purposes of the study and research questions are proposed.

Structural Equation Modeling

SEM techniques are frequently used when the focus of the study is to understand the inter-relationships among variables. SEM techniques are also called latent variable modeling techniques, which tend to involve multiple observed variables and latent variables. SEM techniques include a family of models such as Path Analysis, Confirmatory Factor Analysis (CFA), Full Structural Regression Model (SR model), Multiple-sample SEM, and Latent Growth Modeling. Structural equation models can be represented using diagrams to convey the hypothesized inter-relationships among variables. Below I briefly introduce the fundamental principles of SEM using CFA and SR models as examples because these two types of models are relevant to my study.

Confirmatory Factor Analysis

A CFA model is also called a measurement model. It is used to examine *a priori* relationships between latent factors and a set of measured indicators (or items). Researchers typically have hypotheses about the factor structure before conducting the analysis. Specifically, they posit a model with a number of latent factors and know which measured variable is an

indicator of which factor (Kline, 2010). A CFA model assesses whether the proposed relationships among the measured indicators and latent factors are supported by the data.

To illustrate, a CFA model with two latent factors each measured by three items is shown in Figure 1. Latent variables in a CFA model are represented by circles or ovals, whereas the measured variables are represented by squares or rectangles. The directional line from a latent factor to an observed item indicates a directional relationship between them. The weight applied to the latent factor to obtain the observed item, called factor loading, indicates the strength of the relationship. Two-headed lines (or curves) indicate covariances between two variables or variances of variables. In this CFA model, some of the factor loadings are fixed at a constant (i.e., 1) for identification purposes to assign a scale to each of the latent variables. This method is called Unit Loading Identification (ULI). In Figure 1, the loadings for Item 1 and Item 4 are fixed to 1. Consequently, the first factor and the second factor are assigned to a scale that is related to the scale of Item 1 and Item 4, respectively. For the same reason, all the directional effects from residuals to items are fixed to 1. Consequently, the observed scores for the three items (X1 to X3) measuring the first factor can be expressed as:

$$X1 = 1 * F1 + e_1$$

$$X2 = Loading_2 * F1 + e_2$$

$$X3 = Loading_3 * F1 + e_3$$

where *F* denotes factor score and *e* denotes residual. Although not shown, the three items measuring the second factor can be expressed in a similar format.

A CFA model should be identified before conducting the analysis. "A model is identified if it is theoretically possible for the computer to derive a unique estimate of every model parameter" (Kline, 2010, p. 93).

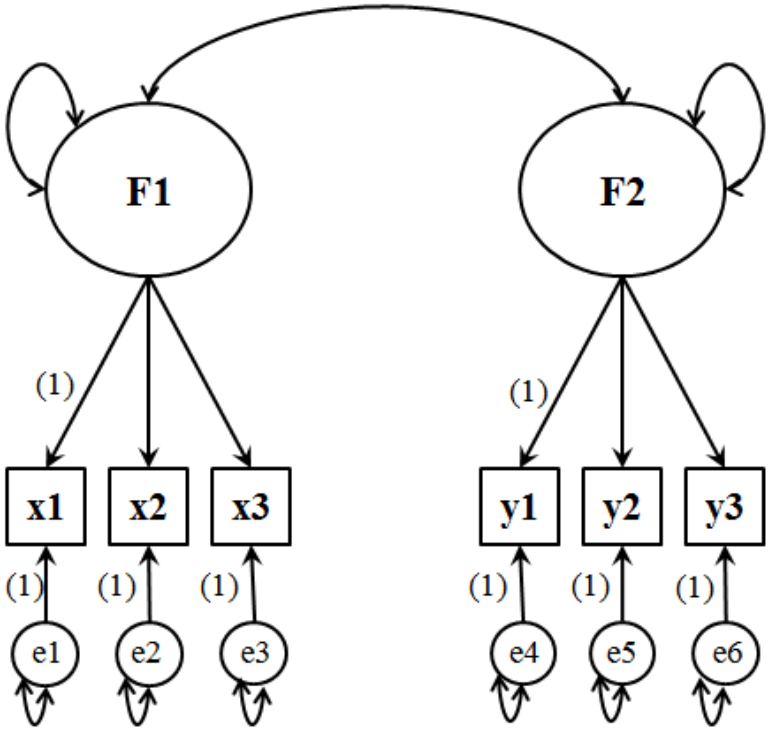

Figure 1: CFA with Two Factors

To meet this requirement, in addition to assigning a scale to each latent variable, a CFA model should have degrees of freedom (*df*) greater than or equal to zero. Model *df* is determined by the number of pieces of information minus the number of free parameters. The amount of information in a model is the number of unique elements in the variance-covariance matrix among the observed variables when mean structure is not considered. For example, the two-factor CFA model given in Figure 1 has six observed variables. The number of unique elements in the covariance matrix among these six variables is 21 as computed by 6*(6+1)/2.

The number of freely estimated parameters in the model is 13 consisting of:

- Factor variances for factors: 2
- Covariance between the two factors: 1
- Variances of disturbances for each item: 6

- Factor loadings: 4

The degrees of freedom for this model is thus 8 (= 21-13) and the model is theoretically identified.

Full Structural Equation Models

Full structural equation models are also known as structural regression (SR) models. A SR model is a model describing the relationships among several latent variables each with a measurement component. The model used for the simulation study, which is shown in Figure 2, is an example of SR model. In this model, the hypothesized relationships between the latent factors F1 and F2, and F1 and F3 are directional (i.e., F1→ F2; F1→ F3). The disturbances associated with F2 and F3 are correlated. In addition, each of the three latent factors is measured by a set of measured variables. That is, items X1 to X15 are measured indicators of F1, items X16-X18 are measured indicators of F2, and items X19-X21 are measured indicators of F3. The number of freely estimated parameters in this SR model is 45 consisting of:

- Variance of F1: 1
- Variance of disturbances of F2 and F3: 2
- Covariance of disturbances: 1
- Direct effect among the factors: 2
- Factor loadings (one loading for each factor is fixed at 1): 18
- Variance of measurement error for each item: 21

The amount of information for the model is 231 (= 21*22/2). Therefore, degrees of freedom *(df)* for the model are 186 and the model is identified.

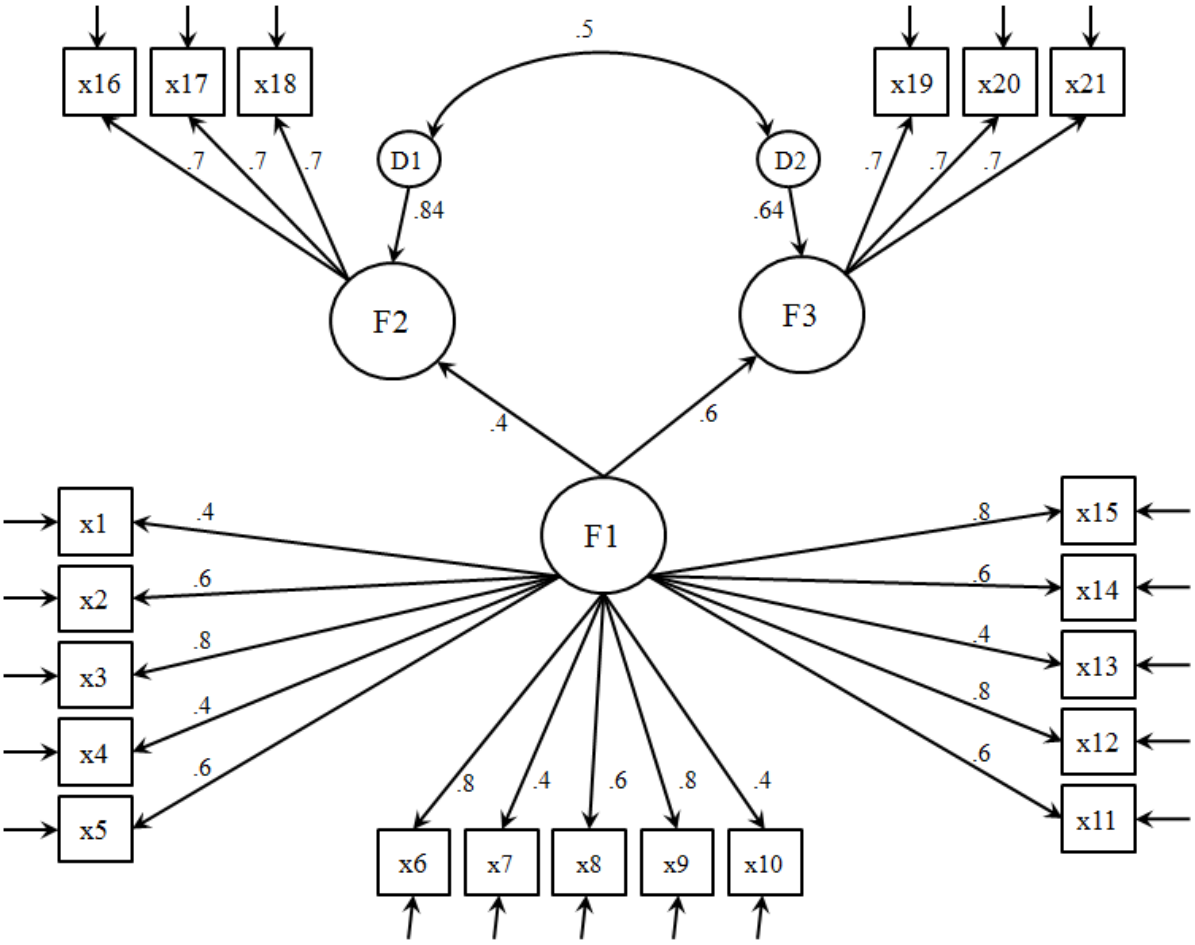

Figure 2: The SR Model Used for the Data Generation

Null Hypothesis and Fit Indices

The null hypothesis for statistical testing for the overall fit of a SEM is that the population covariance matrix among observed variables is identical to the reproduced covariance matrix derived from the model. The mathematical representation of the null is:

$$\Sigma = \Sigma(\theta)$$

where Σ represents the observed covariance matrix, $\Sigma(\theta)$ represents the reproduced covariance matrix, and θ represents the model parameters. The chi-square (χ^2) test is used to test this null

hypothesis. The smaller the chi-square is, the better the model-data fit is. However, it is well known that the χ^2 test is a function of sample size. Therefore, it is necessary to consider supplemental fit indices to assess the model-data fit. Numerous supplemental fit indices are available in the SEM literature but no single fit index works well for all type of models (Davey et al., 2005). Therefore, using multiple fit indices has been recommended to assess the model-data fit (Davey et al., 2005).

The most commonly used fit indices can be categorized into two groups: incremental fit indices (a.k.a. comparative fit indices) and absolute fit indices. However, some of the indices can be categorized into multiple categories (Kline, 2010). Incremental fit indices indicate the relative fit of the researcher's model to a baseline model (Davey et al., 2005; Kline, 2010), including Comparative Fit Index (CFI; Bentler, 1990) and Tucker and Lewis Index (TLI; Tucker & Lewis, 1973). The larger the values for both CFI and TLI are, the better the fit is. The range of CFI is 0 to 1. However, values of TLI may go outside this range. Values of CFI and TLI greater than .95 indicate good fit (Hu & Bentler, 1999). Root Mean Square Error of Approximation (RMSEA; Steiger & Lind, 1980) and Standardized Root Mean Square Residual (SRMR; Bentler, 1995) are two absolute fit indices. RMSEA smaller than .06 and SRMR smaller than .08 indicate good model-data fit (Hu & Bentler, 1999).

In addition to the typical assumptions (e.g., multivariate normality assumption), complexity of models has a major effect on the estimation procedure in SEM. Model complexity increases as the number of parameters estimated in the model increases. A more complex model generally requires a larger sample size in order to reach stable parameter estimation. Some conventional rules about the ratio of sample size to the number of free parameters are available in the literature. For example, Kline (2010) indicated that ratios less than 10:1 need attention

especially for sample size smaller than 100. In practice, when a model involves a large number of measured indicators, applied researchers may use some item parceling techniques to reduce model complexity when the purpose of modeling is to investigate the relationships among latent factors, particularly when the sample size is relatively small. Next, I summarize item parceling techniques as well as their pros and cons.

Item Parceling

Item parceling has been often used in empirical SEM analyses since it was introduced by Cattell in 1956 (Bandalos, 2002; Kishton & Widaman, 1994; Little et al., 2002). Parceling is referred to as a procedure for computing sums or average scores across multiple items. The sum or average scores (called parcel scores) instead of the individual item scores are then used as indicators of latent factors in the SEM analysis (Bandalos 2002, 2008; Little et al., 2002; Yang, Nay, & Hoyle, 2010). In addition to Bandalos et al. (2001) as I summarized in the previous chapter, Plummer (2000) reviewed three journals published from 1996 to 1999 and found that 102 articles used confirmatory factor analysis (CFA) and 50 of them used some kind of parceling techniques in the analysis.

Use of parcels is appealing because it reduces model complexity; that is, the number of indicators of a latent factor is reduced to a smaller number (Nasser & Takahashi, 2003). Researchers have argued that using parcels also help reach optimal reliability, avoid violation of normality assumptions (particularly when the individual items are measured with a limited number of response categories), reduce the requirements on sample sizes, reduce influences of individual items' systematic errors on the model estimation, and help obtain better model-data fit (e.g., Little et al., 2002; Yang et al., 2010). However, researchers have also pointed out

disadvantages of using parcels. For example, it may blur the dimensionality of original measures and produce biased estimates of model parameters (Bandalos, 2002; Matsunaga, 2008).

There have been discussions about whether parceling techniques should be used. Methodological and simulation studies have shown that parceling might not always lead to the correct model. "Parceling improves model fit regardless of whether the fitted model is correctly specified or not" (Matsunaga, 2008, p. 289). Since the late 1990s, a number of authors have investigated optimal techniques for creating parcels while minimizing the bias of parameter estimates and the distortion of dimensionality of measures. Little and his colleagues (Little et al., 2002) stressed that researchers should understand the structure and dimensionality of the original measures before creating parcels. They also summarized various parceling techniques for both unidimensional and multidimensional measures. In this study, I focused only on unidimensional measures because parceling procedures for unidimensional structure are less complicated than those for multidimensional structures. In future studies, based on the results from the current work, multidimensional structures for parcel creation with missing data will be considered. For parceling techniques for multidimensional measures, see Matsunaga (2008) and Little et al. (2002).

Parceling Techniques for Unidimensional Measures

While it is controversial whether item parceling should or should not be used, there are some other discussions on use of parceling such as the types of the parceling techniques and number of items needed for a parcel. In this section several item parceling techniques for unidimensional measures are introduced including random assignment, factorial algorithm, and correlational algorithm. It is important for researchers to understand the structure of the items (Little et al., 2002). "If the dimensionality of a set of items is not known, the items could be

prescreened using an exploratory factor analysis algorithm that uses an iterative estimator with an oblique rotation [e.g., the analogous algorithm of the SEM measurement model]" (Little et al., 2002, p. 165).

Random Assignment

This method is easy to understand and to apply in the real world by randomly assigning items to parcels. However, Matsunaga (2008) recommended that it would be better to have parcels with approximately equal communality and error variances. "If the items evince unequal variances because the scales, or metrics, differ across items, the resulting parcel would be biased in favor of the items with the larger variances" (Little et al., 2002, p. 165). Little and his colleagues have also recommended standardizing the item scores as a solution for this problem.

Factorial Algorithm (Item-to-Construct Balance)

The factorial algorithm is another parceling method for unidimensional measures. This technique is also known as "single-factor" method (Landis, Beal, & Tesluk, 2000; Matsunaga, 2008). Different from the random assignment technique that assigns "item specific components" to parcels randomly (Matsunaga, 2008), the factorial algorithm technique decomposes "item specific components" and combines them within different parcels purposefully. For this technique, parcels are created based on the magnitude of the factor loadings. First, a factor analysis is executed. Parcels are then created with respect to the magnitude of the loadings. Figure 3 illustrates an example of the factorial algorithm adopted from Matsunaga (2008). Let's assume that three parcels are created from twelve items. First, a factor analysis is conducted on these 12 items to obtain their loadings ordered from the largest to smallest as shown in Figure 3. The first parcel would

then consist of items 1, 6, 7 and 12. The second parcel has items 2, 5, 8 and 11. The rest of the items are assigned to the third parcel.

In the above example, factor loadings are utilized to determine which items to be included in which parcel because only the covariance structure is of interest. When the SEM analysis involves a mean structure, balance on the means or intercepts should also be considered when creating parcels (Little et al., 2002):

> Under conditions in which the intercept information is also important, this procedure can be extended to include the intercepts by specifying a single construct model as mentioned previously, but request that the means, or intercepts, be estimated. In this case, one has to consider the relative balance between the discrimination parameter of the item (e.g., its loading) and its difficulty parameter (e.g., its intercept) in constructing balanced parcels (p. 166).

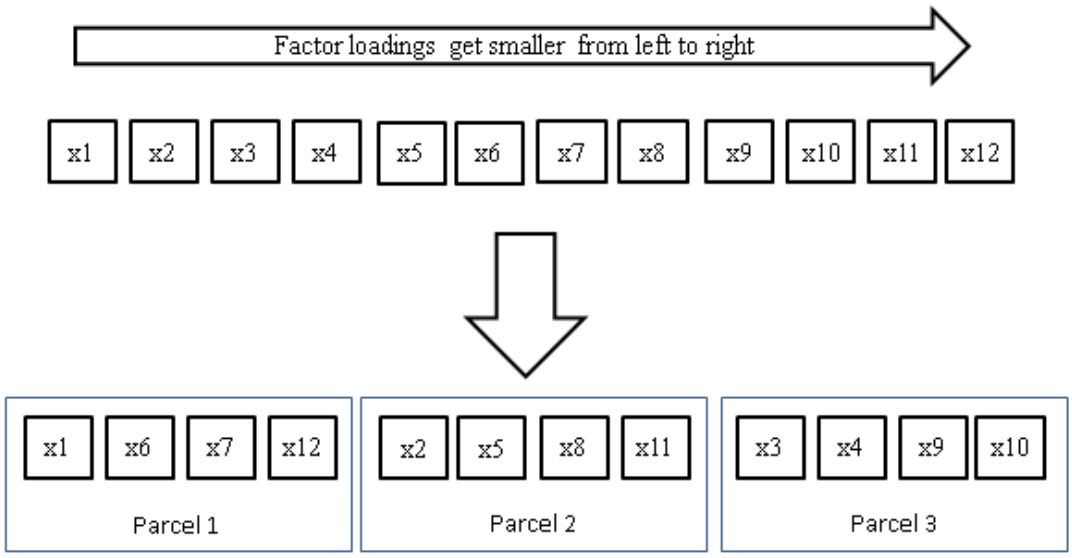

Figure 3: Factorial Algorithm

Correlational Algorithm

Another parceling method is correlational algorithm. In this method parcels are created based on the magnitude of correlations among the items. Starting from the highest correlation, items are assigned to the parcels, until all the parcels have been assigned once. Next, correlations for the remaining items with the parcels are calculated. Then, starting from the highest correlation, items are assigned to parcels based on the correlation between the items and the parcel. The same procedure is applied again until all the items are assigned into one parcel. Figures 4-6 illustrate this method.

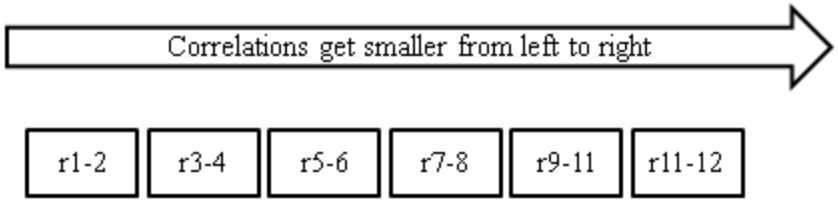

Figure 4: Correlational Algorithm (Step 1)

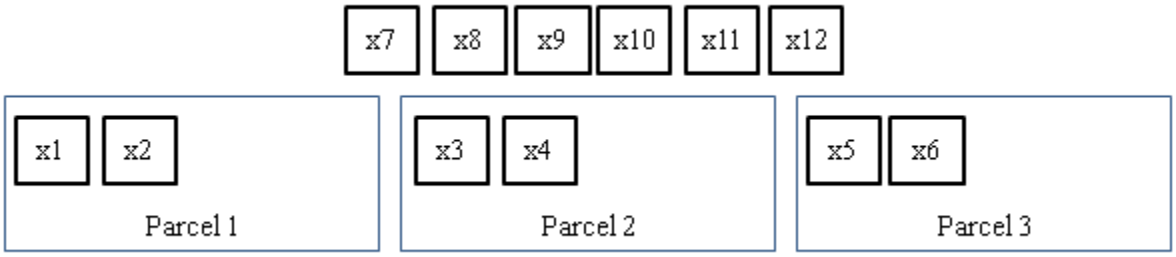

Figure 5: Correlational Algorithm (Step 2)

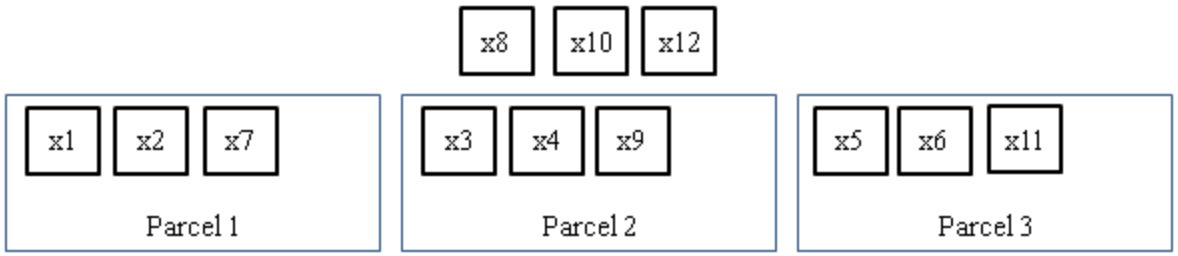

Figure 6: Correlational Algorithm (Step 3)

Pros and Cons of Item Parceling

To parcel or not to parcel is still a controversial topic. There are some good reasons for using parcels, such as its helpfulness to reduce the number of indicators for latent variables and increase the likelihood to get closer to normally distributed indicators. For example, Sass and Smith (2006) clearly indicated that item parcels are usually preferred over individual items because the smaller the number of indicators is, the smaller the chance of encountering estimation error is, and the higher the chance of meeting the normality assumption is.

Little et al. (2002) summarized two different philosophical perspectives on item parceling. The first was described as "*conservative view*". For this philosophy, "parceling is akin to 'cheating' because modeled data should be as close to the response of the individual as possible in order to avoid the potential imposition, or arbitrary manufacturing, of a false structure" (p. 152). Therefore, using parcels alters the data structures. The second perspective of parceling was labeled as "*liberal perspective*." Based on this view, measurement process is defined by the researcher; therefore, choosing to use parcels is a "matter of choice" and it is a part of the investigation. Next I summarize pros and cons of parceling in the literature.

Pros of Parceling

Pros of parceling could be classified into two main classes: the first one is based on "psychometric characteristics of parceling" and the second one is related to "factor-solution and model-fit" issues (Little et al., 2002; Matsunaga, 2008).

Psychometric characteristics of parceling. Matsunaga (2008) listed two psychometric advantages of parceling. The first one is that it increases the communality, increases the commonality to uniqueness ratio for each parcel, and reduces the random error. Because individual items usually do not cover all part of the construct but capture only restricted feature of a construct, by parceling more aspects of the construct would be covered (Matsunaga, 2008). From classical test theory perspective, the observed scores consist of true score, systematic error, and the random error (e.g., McDonald, 1999). The mathematical representation is:

$$X_i = T + S_i + e_i$$

where X is a score from an individual item i, T is the true score, S is the systematic error and e is the random error. S represents idiosyncratic component that is specific, but unrelated to the construct (Matsunaga, 2008). If the individual items are used to measure a construct, both specific and random errors will cause a reduction on the communality. However, if aggregated items are used, "the random errors, e_i, and the specific components, S_i, would be drastically reduced because e_i and S_i are defined as uncorrelated sources of variance within an item and across all items in a domain; e_i and S_i components are also assumed to be uncorrelated" (Little et al., 2002, p. 156). In other words, when a large number of items are used, the estimation of a person's true score would be more accurate (Little et al., 2002).

The second psychometric advantage of parceling is that parceling increases the chance of approaching normality of the distribution of scores (Bandalos, 2002; Matsunaga, 2008; Meade &

Kroustalis, 2005; Little et al, 2002). “Parceling can normalize the distribution because, as far as the assembled items tap various aspects of the construct of interest, aggregating them into parcels may well distinguish individuals at subtly varying levels in regard to that construct” (Matsunaga, 2008, p. 265).

Factor-solution and model-fit. Using parcels increases the model fit (Nasser & Wisenbaker, 2003; Plummer, 2000). Also, it may decrease the item-specific biases and random error (Matsunaga, 2008). On the contrary, using a larger number of individual items tends to result in increase in the measurement error (Matsunaga, 2008). In addition, as number of indicator per construct is increased, the stability of the results of the model decreases (Little et al., 2002). “Thus, there is a dilemma: from a psychometric standpoint, the more items the better (Marsh, Hau, Balla, & Grayson, 1998); from a modeling perspective, however, more items means increasingly bad model fit, and hence, undesirable” (Matsunaga, 2008, p. 268).

Little et al. (2002) perceived model-level issue from type I error point of view. When individual items are used, the number of “spurious correlation” is higher than when the parcels are used. They also pointed out that, using more items in item-based analysis may increase the possibility of sharing specific source of variance. In addition, item-based models are more unstable, causing larger standard errors of measurement. Moreover, as the number of indicators per factor increases, the model tends to be more complex (i.e., with more degrees of freedom). Little et al. (2002) argued that a just-identified model (with zero degrees of freedom) is better than an over identified model (with more degrees of freedom). “Depending on the number of items that have been measured to represent a construct, parcels can be used to effectively reduce the number of indicators to an optimal, just-identified level” (Little et al, 2002, p. 162). On the

other hand, an over-identified model is more complex, resulting in unstable solution, unless required sample size is increased (Matsunaga, 2008).

Cons of Parceling

Matsunaga (2008) listed two main critiques for use of parceling. One is related to dimensionality and the second is parameter estimation bias.

Dimensionality. When the dimensionality of the items is not clear, parceling should not be used (Little et al., 2002). Dimensionality of a set of items should be known so that a proper parceling technique can be decided. It was even suggested that parceling could be used only for unidimensional item structure (Bagozzi, & Edwards, 1994; Bandalos et al., 2001). When items are multidimensional, parceling may blur the factor structure, rather than clarifying it (Bandalos, 2002; Matsunaga, 2008).

Parameter estimation bias. Research on item parceling yields conflict results in parameter estimation bias (Matsunaga, 2008). Some researchers demonstrated that using parceling increases the estimation bias (Hall, Snell, & Foust, 1999; Stephenson, & Holbert, 2003). However, other researchers disagreed with this conclusion (Bandalos, 2002) and found totally opposite findings. For example, Matsunaga (2008) found that parceled data resulted in less bias path coefficients than item level data. He argued that "previous findings indicate that parceling is particularly useful when the given data are not optimal in light of the assumptions invoked for SEM such as multivariate normality and uncorrelated errors" (p. 276) and the reason the use of parceling resulted in more bias in some studies is that the researchers used extremely well condition which was unrealistic in most application studies.

So far, there is no clear conclusion about the usefulness of item parceling. It is still a controversial topic. However, the controversy also indicates that the potential of item parceling is still to be discovered. More research is needed to examine the issues regarding the use of item parceling. For example, item parceling technique may be used with missing data. "If a participant has missing values for one or more items, it seems more reasonable to average the items that remain rather than report a missing value for the entire scale" (Schafer et al., 2002, p. 157). For the three parceling techniques used for unidimensional measurement as I summarized above, Rogers and Schmitt (2004) suggested that none of them is better than the other in obtaining accurate parameter estimates. However, factorial algorithm technique has shown to be superior to others in terms of model-data fit (Rogers et al., 2004). Therefore, factorial algorithm was used in this study. In the next section, I summarized missing data mechanisms and techniques.

Missing Data Mechanisms

Missing data mechanisms were first proposed by Rubin (1976). Rubin explicitly defined missing mechanisms in a probabilistic manner (Enders 2010). "Missing data mechanisms describe relationships between measured variables and probability of missing data" (Enders, 2010, p. 2). These probability-based mechanisms can also be considered as assumptions for missingness (Allison, 2002; Peugh & Enders, 2004). Broadly, the assumptions could be divided into two groups: ignorable and non-ignorable missing data (Chen & Astebro, 2003). Although some other types of mechanisms have been discussed in the literature (e.g., "missing by definition of the subpopulation" [Acock, 2005]), most widely discussed missing mechanisms are missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR), with the later mechanism indicating more relaxed or weaker assumption. In addition,

the first two mechanism (MCAR and MAR) are considered as ignorable whereas the last one (MNAR) as non-ignorable.

Missing Completely at Random (MCAR)

MCAR could be described as the mechanism that has the strongest assumption compared to the other two mechanisms. Assume that *X* (focal variable) is the variable that has some missing values. If the missingness is not related to the variable itself (*X*) or any other variables (*Y*s) in the dataset, missingness on the variable *X* is called MCAR. By definition, missingness on variable *X* is not related to any other variable in the dataset. However, this does not guarantee that *X* is not related to any other variables which are not included in the dataset.

A dummy variable *R* can be created to indicate missingness for a particular variable. Whenever there is a missing value on the focal variable, *R* is coded as 0. Therefore, the variable *R* can be called as "missingness" variable (Schafer et al., 2002). Assume that we have the dataset shown in Table 2. *X* is the observed variable that does not have any missing values on it. X_{mis} is the variable with missing values. Even though it is not specifically indicated in Table 2 (for simplicity), let's assume that there are some other measured variables (*Y*s) in the dataset. Also, let's assume that some other possible variables are not included in the dataset (*Z*s – un-measured variables). Therefore, the complete dataset, *D*, that consists of observed and missing data can be expressed as:

$$D = (X_{mis}, Y_s).$$

If a variable is missing completely at random (MCAR), the probability of the missingness is the same for the data with or without missing. That is, the probability of missingness does not depend on any variables in the dataset.

$$P(R|D) = P(R|X_{mis}, Y) = P(R)$$

Table 2: *An Example of R-variable*

ID	X	X_{mis}	R	Y_1-Y_n
1	3	NA	0	…
2	2	2	1	…
3	1	NA	0	…
4	3	3	1	…
5	1	NA	0	…
6	0	0	1	…
7	1	1	1	…
8	2	2	1	…
9	2	NA	0	…
10	1	NA	0	…
11	3	3	1	…
12	4	4	1	…
13	1	1	1	…
14	2	2	1	…
15	0	NA	0	…
16	2	2	1	…
17	3	3	1	…
18	4	NA	0	…
19	1	1	1	…
20	4	4	1	…

Schafer et al. (2002) and Enders (2010) depicted the idea of MCAR as shown in Figure 7, where *X* is the variable that has missing values, *Y* is other measured variables in the dataset (*Y* – auxiliary), *R* is a dummy variable for missingness on *X*, and *Z* is any possible un-measured auxiliary variables. The two headed arrows indicate existence of correlation. For example, the two-headed arrow between *X* and *Y* indicates possible correlation between the focal variable and any auxiliary variables, whereas, the two-headed arrow between *R* and *Z* indicates possible correlation between missingness and any other un-measured variables. For MCAR, there exists no direct or indirect effect between *X* and *R*. In other words, MCAR mechanism is not affected by the correlation between missingness and un-measured variables.

MCAR is the only mechanism that may be tested (Enders, 2010). Allison (2002) explained the testing procedure for MCAR. The first step is to create a dummy variable for the focal variable that shows '1' for non-missing and '0' for missing values. Then, test if the average values of the focal variable for 0-coded and 1-coded cases are significantly different from each other. Non-significant mean difference may indicate MCAR mechanism. However, even though the test indicates a non-significant mean difference, it does not necessarily mean that data are MCAR, deeper investigation is necessary to confirm data are MCAR (Allison, 2002) because subgroups with equal means can also be obtained under MNAR and MAR (Enders, 2011).

Missing at Random (MAR)

MAR requires a weaker assumption than MCAR (Allison, 2002). In MAR, missingness on the focal variable (*X*) is related to the other variables in the dataset but not to *X*. If any variable (*Y*) in the dataset is related to the missingness on *X*, the missingness on *X* is called MAR. The name of MAR is somewhat confusing. Although it is called missing "at random," the probability of the missingness on *X* is related to other variable/s in the dataset; that is, there exists a "systematic relationship" between focal variable and other measured variables (Enders, 2010).

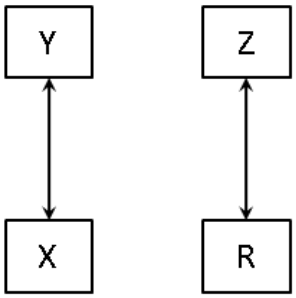

Figure 7: Missing Completely at Random

Schafer at al. (2002) and Enders (2010) pictured MAR cases as shown in Figure 8. MAR can be formulated as follow:

$$P(R|D) = P(R|X_{mis}, Y) = P(R|Y).$$

All the terms are the same as defined earlier. In words, if a variable is missing at random (MAR), the probability of the missingness is the same for the data with or without missing observation, controlling for some variables (*Y*) in dataset. As also shown in Figure 8, different from the MCAR case, MAR has a correlation between the missingness and a measured auxiliary variable as indicated by the two-headed arrow between *Y* and *R*. In addition, there may be a correlation between *Y* and *Z*, which is indicated by a dashed line in Figure 8. Enders (2010) and Schafer et al. (2002) did not place this dashed line in their works; however, having this dashed line is necessary since it is possible that there is a correlation between *Y* and *R* through *Z*.

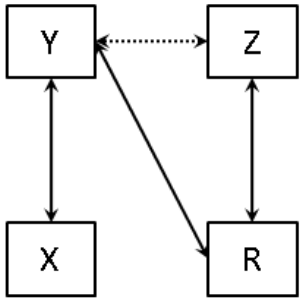

Figure 8: Missing at Random

The problem with the MAR mechanism is that it is not possible to test it (Allison, 2002; Enders, 2010). We are never sure if the missingness only depends on the measured variables (e.g., *Y*) (Enders, 2010). "Because we do not know the values of the missing data, we cannot compare the values of those with and without missing data to see if they differ systematically on that variable" (Allison, 2002, p. 4). One way to satisfy the MAR assumption is to include auxiliary variables in the dataset that may potentially be related to the missingness on the focal

variable (Enders, 2010). If the cause of missingness is a measured variable, the MAR assumption is fulfilled (Enders, 2010). The strategy is called "inclusive analysis". Inclusive analysis strategy suggests having as many auxiliary variables as possible. By including these auxiliary variables in the analyses, statistical power may be increased and non-response bias may be reduced. In addition, including auxiliary variables may convert possible MNAR cases to be MAR (Enders, 2010). Consequently, having auxiliary variables may possibly relieve the assumption of most of the missing data handling methods (i.e., maximum likelihood and multiple imputation) since missingness would be considered MAR with certain auxiliary variables. For example, assume that we have a dataset which contains scores from a test measuring individuals' math and science abilities. A dummy variable is used to indicate if a student is attending after-school activities or not. The after-school program is assumingly specific for lower-level achieving students. Also, let's assume that the math scores for some of the students are missing. The missingness on the math scores might be considered as MAR if we think students' math scores are missing for those who do not attend after-school activities. However, if we did not have the dummy variable to indicate if a student attended after-school activities in the dataset, the missingness on the math scores might be thought of as MNAR since the after-school program is only for lower-level students. Therefore, including the auxiliary variable (e.g., after-school activities) may turn MNAR into MAR.

Missing Not at Random (MNAR)

The missing mechanism is called MNAR if the missingness on a variable is related to the variable itself. Consider that *X* is the focal variable with missing values on it. When the data present MNAR, the probability of missingness on *X* depends on its value. For example, if all the values of X1 on Table 1 are missing for X1= 4, it can be concluded that X1 is MNAR. If values

on a variable are missing not at random (MNAR), the probability of the missingness is different for the cases with or without missing elements. The mathematical representation for MNAR is:

$$P(R|D) = P(R|X_{mis}, Y) = P(R|X_{mis})$$

The graphical representation of the missing mechanism for MNAR is demonstrated in Figure 9. In contrast to Figure 8, Figure 9 has one additional two-headed arrow between focal variable *X* and missingness *R*, indicating that the missingness is related to itself.

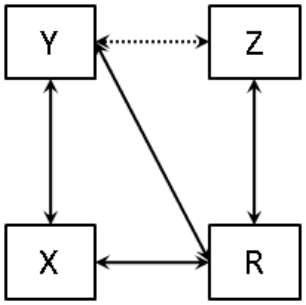

Figure 9: Missing Not at Random

Missing Data Techniques

Data missingness is an issue frequently encountered in the experimental and observational studies in education and psychology. The reasons behind the missingness vary. Researchers might be aware of the reasons for missingness in some cases, but not in most cases. McKnight et al. (2007) classified the reasons for missingness into three categories: missingness caused by the participant, caused by the study design, and caused by both the participant and the study design. If certain questions are offensive for some of the examinees, the examinees may decide to not answer the questions (McKnight et al., 2007). For example, a student who is participating in a survey may not want to answer a question related to his or her GPA score from previous year just because his or her GPA is low. If a question is too long to read, examinees may not read all questions thus leaving some questions un-answered.

In literature, many methods have been proposed to handle missing data (Peugh & Enders, 2004). These methods can be categorized into two broader groups: traditional (ad hoc) methods and modern methods. Traditional methods include listwise deletion, pairwise deletion, and mean imputation, whereas full information maximum likelihood, expectation maximization, and multiple imputation are considered as modern methods. Below I provided a brief review of each of these methods.

Listwise (Case) Deletion

Listwise deletion (LD) is the most commonly used method in the literature (Acock, 2005; Schafer et al., 2002). For example, it is the default method in SPSS 20 (1993 – 2012) to handle missing data. LD removes all the cases as long as there is a missing value on any of the variables on these cases. That is, the dataset, after LD, will contain only cases with complete data on all variables. Consider that we have the dataset presented in Table 1. After LD is applied the dataset will be the one presented in Table 3, in which the same size decreased by ten.

LD is a commonly used ad hoc method probably because it is easy to apply. However, it has disadvantages. The most obvious disadvantage is it reduces the sample size in the analysis, and so the power (Acock, 2005; Enders, 2006; McKnight, 2007; Peugh et al. 2004). The LD method assumes MCAR, which is the strictest assumption for the missing data mechanism (Carter, 2006; Enders, 2006). If the assumption of MCAR is not met, parameter estimates are usually biased (Enders, 2006, 2010; Peugh et al. 2004), inefficient (Chen et al., 2003), and Type II error may be increased (Acock, 2005). However, the power is not a concern for the LD method when the sample size is big enough and the data are missing completely at random (Acock, 2005).

Table 3: *Listwise Deletion*

ID	X1	X2	X3	X4
2	2	4	3	2
4	3	1	2	2
7	1	2	1	1
8	2	2	2	2
11	3	0	2	3
12	1	1	1	2
13	1	2	1	0
16	2	3	2	2
17	3	1	1	3
19	1	1	4	1

Pairwise Deletion

Pairwise deletion (PD) method is also known as available case analysis (Allison, 2002). It is also one of the methods available in SPSS 20. Instead of removing all the cases with missing values at once as in LD, PD removes the cases only for variables that are involved in the analysis. Take the dataset from the Table 1 as an example. If we are interested in computing the correlation between variables X1 and X3, the data used in the computation would look like those presented in Table 4. Although there are some missing values in X2, the analysis does not involve X2, thus the cases with missing values on X2 but not on X1 and X3 (i.e., cases # 6, 14, and 20) are not excluded.

PD is more powerful and more efficient (e.g., has smaller sampling variability) than LD (Allison, 2002; Enders, 2010), however, parameter estimates may still be biased when missingness is not MCAR (Allison, 2002; Enders, 2010). Another important issue with PD is that it uses different sub-sets of the data for different analyses (Enders, 2010; Peugh et al., 2004). For example, consider the dataset in Table 1 again. The sample size for computing the correlation between X1 and X3 is 13, whereas the sample size for computing the correlation

Table 4: *Pairwise Deletion*

ID	X1	X2	X3	X4
2	2	4	3	2
4	3	1	2	2
6	0	NA (3)	1	1
7	1	2	1	1
8	2	2	2	2
11	3	0	2	3
12	1	1	1	2
13	1	2	1	0
14	2	NA (3)	2	1
16	2	3	2	2
17	3	1	1	3
19	1	1	4	1
20	0	NA (4)	2	3

Note. NA indicates missing value. Values enclosed in parentheses indicate values of the missing data.

between X2 and X3 is 16. Related to this, it is difficult to compute the model degrees of freedom[1] with PD (Acock, 2002; Schafer et al., 2002). In addition, it is likely to have a singular (non-invertible) variance-covariance matrix (Enders, 2006; McKnight et al., 2007) as well as correlation out of the boundary of -1 to 1 (Schafer et al., 2002) with the PD method. Some approaches have been suggested to handle the sample size issue. One is to use the smallest sample size as if it is the actual sample size; however, that reduces the power (Acock, 2002). Another is to use the average of the sample sizes, "but this approach is likely to underestimate the standard error of some variables and overestimate the standard error for others" (Enders, 2002, p. 41).

Mean Imputation

Mean imputation method is easy to implement. Any missing values are replaced by the average score based on all the available cases on the variable. The mean value of the variable is

[1] In this case, the term "degrees of freedom" is not the same as the one defined in SEM. It is related to the sample size, for example, the df for correlation is n-1 where n is sample size.

accurate under the MCAR data (Schafer & Graham, 2002). However, mean imputation not only attenuates variances and covariance of the variables (Acock, 2002; Allison, 2002; Enders, 2006; McKnight et al. 2007), but also may change the underling distribution of the variable (Kline, 2010). Using mean imputation, the parameter estimates will be biased "regardless of whether data are MCAR, MAR or MNAR" (Peugh et al., 2004, p. 529). Allison (2002) recommended avoiding this method since it produces biased estimates in general.

Regression Imputation

Regression imputation (RI, a.k.a. conditional mean imputation) is another method that replaces each of the missing values with values based on regression analysis. Specifically, missing value on a variable is replaced by a predicted value that comes from a regression equation (Enders, 2010; Peugh et al., 2004). Assume that we have missing values only on X1 in Table 1. Since there are no missing values on X2, X3 and X4, they can be used as predictors to predict X1. Then, the equation for the regression would be

$$\widehat{X1} = B_0 + B_1 * X2 + B_2 * X3 + B_3 * X4.$$

Based on all non-missing values of X2, X3 and X4, a predicted X1 value will be produced for each case and then replaced the missing values on X1.

Even though RI uses more information than mean imputation (Kline, 2010), it underestimates the variance (Enders, 2006; Enders, 2010). Since the missing values are replaced by the values coming from regression equation, RI tends to overestimate correlations (between the dependent and independent variables) and R^2 value; therefore, it is recommended not to use this method for covariance and correlation analysis (Enders, 2010; Schafer & Graham, 2002). To solve the problems, researchers recommended adding a random error component to each predicted value. This new method is called stochastic regression imputation (SRI) (Enders, 2010).

By adding the random error, the problem of having too high correlation and R^2 is softened, and the attenuation for variance effect might be removed (Enders, 2010). Finally, even under MCAR, RI overestimates the correlation (Allison, 2002; McKnight et al., 2007), whereas SRI gives un-biased parameter estimates under MAR.

Figures 10-14 give an example for each of the methods discussed above under MNAR mechanism. Figure 10 represents the original dataset (i.e., no missing values) on two variables of X and Y. The sample size was 50. Twenty of cases at the lower values of Y were deleted to simulate the missing values under MNAR. Figure 11 shows the scatterplot for the data with missing values. It represents listwise deletion method (also pairwise deletion method for this instance because there are only two variables). Figure 12 shows the scatterplot for mean imputed dataset. In Figure 13, regression-imputed dataset example was depicted. Finally, an example of stochastic regression imputation was shown in Figure 14.

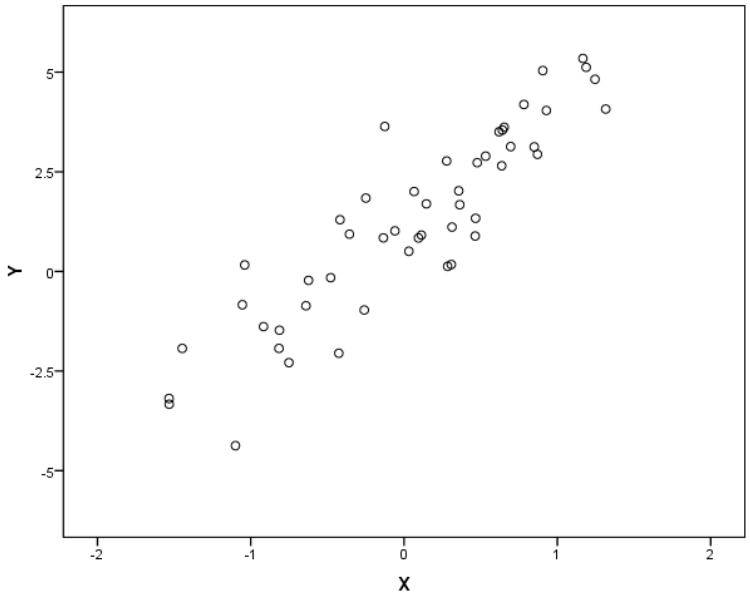

Figure 10: No Missing Values

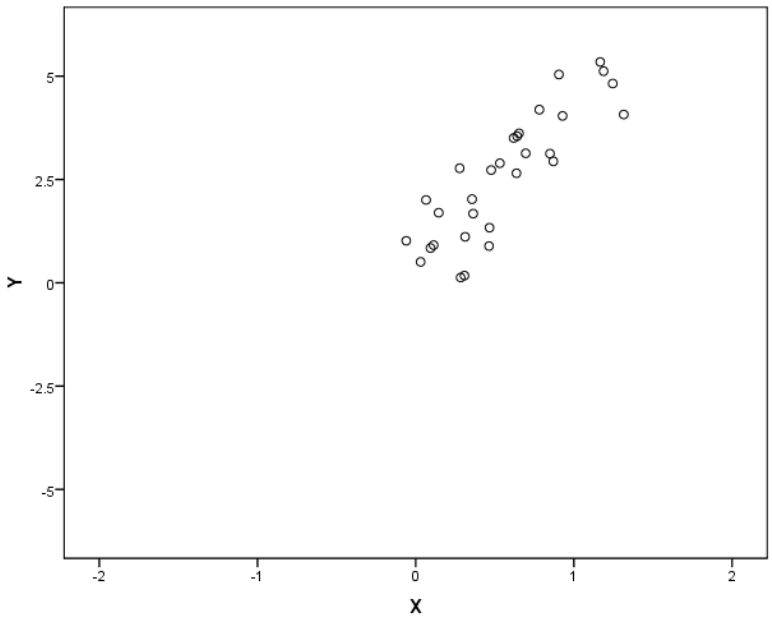

Figure 11: Listwise and Pairwise Deletion

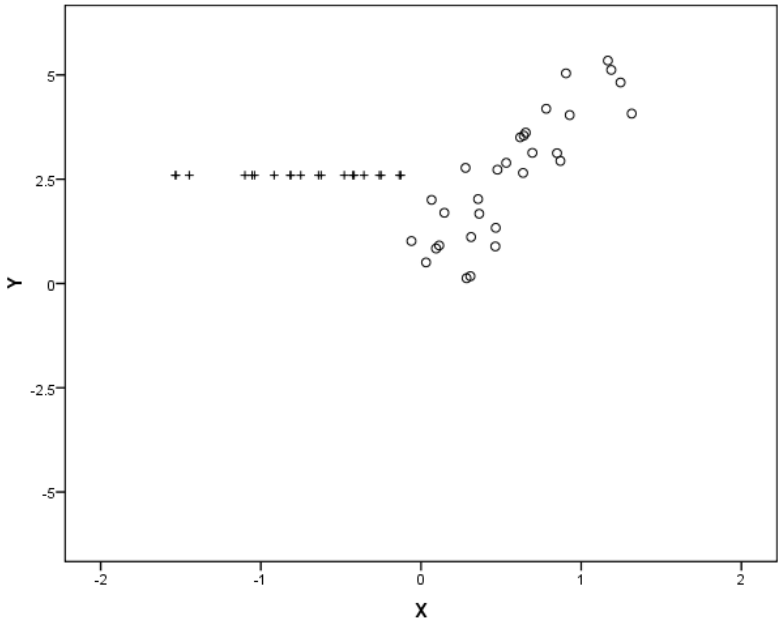

Figure 12: Mean Imputation

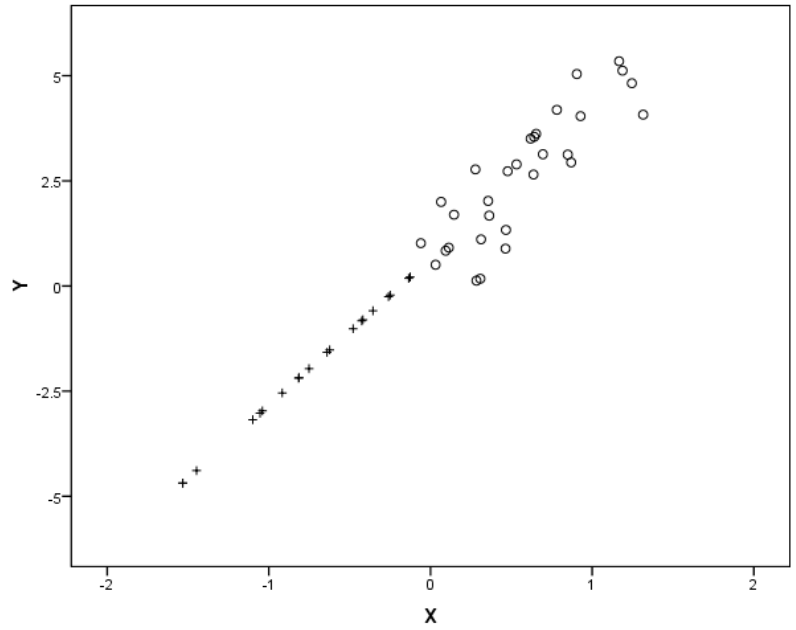

Figure 13: Regression Imputation

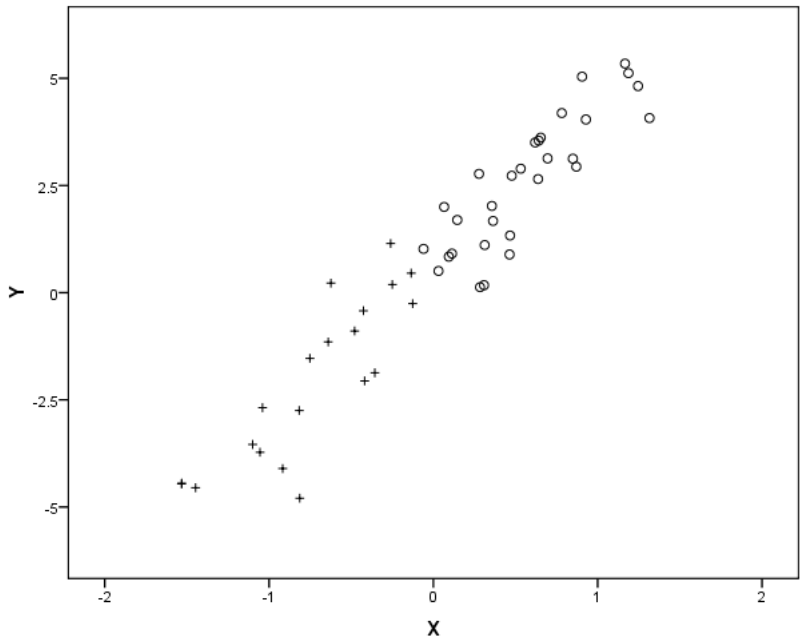

Figure 14: Stochastic Regression Imputation

Full Information Maximum Likelihood

Different from all the methods summarized above, full information maximum likelihood (FIML) neither imputes nor deletes any values from the dataset. Instead, FIML uses all the available information to estimate the parameters (Acock, 2002; Chen & Astebro; 2003). "Briefly, the FIML approach computes a casewise likelihood function using only those variables that are observed" for each case (Enders & Bandalos, 2001, p. 434). The likelihood for each of the cases is calculated and then the sum of them gives the likelihood of all available cases. The log-likelihood function for each case is

$$\text{Log}(L_i) = K_i - \frac{1}{2}\log|\sum_i| - \frac{1}{2}(x_i - \mu_i)'\sum_i^{-1}(x_i - \mu_i).$$

Log-likelihood for all available cases is

$$Log(L) = \sum_{i=1}^{N} \log(L_i).$$

The constant K_i depends on the number of non-missing data points, x_i is the observed value for case i, whereas μ_i is the mean vector and $\sum_i$ is the covariance matrix for the parameters (Enders et al., 2001).

The multivariate normality assumption plays an essential role in FIML estimation under MAR (Allison, 2002; Chen et al., 2003; Enders, 2004). Researchers found that results of FIML showed unbiased parameter estimates under MCAR and MAR and were less biased and more efficient than both PD and LD even under MNAR (Enders & Bandalos, 2001). Moreover, FIML produces accurate model-data fits under normality assumptions (Enders, 2001).

Expectation Maximization

The expectation maximization (EM) method can also be considered an iterative ML method. That is because the "EM algorithm produces a maximum likelihood estimate of the covariance matrix" (Enders, 2006, p. 319). This method consists of two steps: expectation (E)

and maximization (M). Allison (2003) explained how the EM algorithm works for missing data. The first step is the expectation (E) step. In this step, a starting value for the covariance matrix is chosen. The expected values for the covariance matrix are then determined given the initial value. "The purpose of the E step is to estimate the missing values, and this is accomplished with regression imputation" (Peugh & Enders, 2004, p. 533). In the second step, maximization (M), the covariance matrix is updated with the values from the previous E-step, by using the current values of the parameters calculated by linear regression of focal variable on the other variables. "This is done separately for each pattern of missing data (e.g., each set of variables present and variables missing)" (Allison, 2003, p. 549). Then the estimated regression equation values of X for the missing values are imputed. Finally, we recalculate the values of the covariance matrix based on the full dataset. If the convergence criterion does not meet after M-step, repeat the E and M steps until the convergence criteria meets.

The EM algorithm assumes that the data are at least MAR (Allison, 2003; Enders, 2004). The EM algorithm produces similar parameter estimates to FIML, but the standard errors and model-data fits may not be accurate (Enders, 2004). It is because no single sample size gives correct (consistent) standard error for all the parameters (Allison, 2003).

Multiple Imputation

Multiple Imputation (MI) could be considered as equitant to doing single imputation multiple times. In MI, each of the missing values is replaced *k* times. McKnight et al. (2007) suggested replacing missing values between 3 or 10 times. Therefore, after MI is applied there would be 3 to 10 replicated full (with no-missing values) datasets. Then, results of the analysis based on these *k* datasets are averaged for parameter estimates, and standard error (SE) is

calculated using an acceptable procedure. Allison (2002) provided this formula for SE calculation:

$$S.E(\bar{\theta}) = \sqrt{\frac{1}{k}\sum_{m=1}^{k} s_m^2 + (1+\frac{1}{k})(\frac{1}{k-1})\sum_{m=1}^{k}(\theta_m - \bar{\theta})^2}$$

where k is number of replication, $\bar{\theta}$ symbolizes the mean of the parameter estimates, θ_m is parameter estimates from the m^{th} replication, and S_m is estimated standard error from the m^{th} replication.

With MI, the problem related to the uncertainty ("because conventional uncertainty measures ignore the fact the imputed values are only guesses" (Schafer & Graham, 2002, p. 161)) in single imputation is solved. Moreover, MI shares all properties of ML, but removes some of the disadvantages of it (Allison, 2002). The "multiple imputation, when used correctly, produces estimates that are consistent, asymptotically efficient, and asymptotically normal when the data are MAR" (Allison, 2002, p. 27). Since the raw data are available in the MI, any statistical method can be used (Patrician, 2002) while in the EM, the analysis is restricted to variance/covariance matrix. Despite of these advantages, the MI may cause confusion as "the great flexibility of the MI principle has allowed for a proliferation of different approaches and algorithms. This can lead to considerable uncertainty and confusion about how best to implement MI in any particular application" (Allison, 2003, p. 555). Moreover, data generated by the MI (e.g., PROC MI) method are more normally distributed than the original data (Allison, 2003). Because nonnormality is one of the design factors in my study, the MI methods should not be used. Considering these factors, the FIML was adopted as an estimation method in this study. FIML is one of the well-studied estimation procedures and it is commonly used in empirical

studies in social sciences. It also adequately handles missing data (Enders & Bandalos, 2001) and researchers could save a lot of time and energy by applying this estimation technique.

Purpose of This Study

There have been many studies focusing on the use of parceling techniques as well as missing data handling techniques in SEM. However, no study has examined the missing data problem with respect to item parceling, particularly when the observed data present nonnormality. Some researchers have been using item parceling to handle missing data without realizing it is item parceling (Schafer & Graham, 2002) but others have used these techniques to explicitly handle the missing data. For example, parceling techniques have been used to handle missing data in empirical studies in the area of pharmacology (e.g., Veldstra et al., 2011), psychology (e.g., Signorella, 2011), and social works (e.g., Fakunmoju, 2010). Schafer et al. (2002) referred to this method as "*ipsative mean imputation*" or "*case-by-case item deletion*" but did not connect the idea to item parceling. No simulation study has been conducted to examine if using parceling leads to better (or at least not worse) results than individual items when there are missing values and nonnormality issues in the dataset. The purpose of this simulation study is to investigate how item parceling behaves under various simulation conditions in structural equation modeling with missing and nonnormal distributed data in the context of full structural equation model.

Based on the literature, I expect that:

1. Results from analysis based on parcels perform better or equally well in terms of parameter estimation bias compared to item level analysis under any type of missing mechanisms.

2. The rejection rates of the chi-square test for model-data fit based on parcels are lower than those based on individual items, under any type of missing mechanisms.
3. The percentage of missingness on variables may have less effect on the results (e.g., parameter estimation bias) for parceled level analysis compared to item level analysis under all missing mechanisms.
4. Parcel level analysis encounter fewer problems in model convergence than item level analysis, particularly for data with nonnormal distribution and missing values.
5. As the degree of data nonnormality increases, the superiority of parcel level analysis to the item level analysis is more obvious.
6. As the sample size increases, the superiority of parcel level analyses to the item level analysis might be reduced under any type of missing mechanisms.

CHAPTER 3

METHODS

This chapter details data-design factors for the simulation study, as well as the data generation procedure. In addition, data analysis and assessment criteria for evaluating model results are provided.

Data were generated based on a full structural equation model (i.e., SR model) as shown in Figure 2. The model consisted of three latent factors. The first factor, labeled as F1, was measured by 15 items; both the second factor and the third factor, F2 and F3, were measured by three items. Therefore, the total number of items in this model was 21. As shown in Figure 2, all factor loadings associated with F2 and F3 were fixed at .7. However, the loadings associated with F1 varied across items and were .15, .4, .6, or .8 (see details in the next section). I aimed to simulate a situation in which some items were highly loaded on the factor whereas others were not. The variance of uniqueness for each item was fixed as one minus the value of its squared loading. For example, the variance of the uniqueness for an item with a loading of .8 was .36 (i.e., $1 - .8^2$). The path coefficients from F1 to F2 and F1 to F3 were .4 and .6, respectively. The variance of F1 was 1. The variance of the disturbances associated with F2 and F3 were fixed at .84 and .64, respectively. The covariance between the disturbances of F2 and F3 was .5.

A total of five design factors were considered for data generation as detailed in the next section: sample sizes, missing data patterns, percentage of missingness, levels of nonnormality of the item scores, and factor loadings. For each condition, 2000 data sets were generated. For each generated dataset, analysis was conducted based on both the parcel level and the item level. For analyses based on parceled data, parcel scores were calculated as the average across a subset of

the individual items (details later). Three parcels were created for the items associated with F1 based on the factorial algorithm. The items associated with F2 and F3 (i.e., item16- item21) were not parceled.

Design Factors for Data Generation

Sample size. Three different sample sizes were considered: 100, 300, and 1000. These sample sizes were chosen to represent a range of small to large sample sizes in SEM analysis.

Missing data pattern. Four missing patterns were used: Non-missing, MCAR, MAR, and MNAR.

Percentage of missing data. Three levels of missingness were utilized: 10%, 20%, and 40%. Previous studies have showed that percent of missingness had an effect on parameter estimates but almost no effect on convergence rate for models based on the individual items (Davey et al., 2005; Enders, 2001).

Level of nonnormality. Three types of distribution for the item data were considered: (1) normal distribution; (2) moderate skewness and low kurtosis (*Sk* =1, and *K*=1.5); and (3) high skewness and high kurtosis (*Sk* =1.75, and *K*=3.75). Nonnormality was only applied to items 1 through 15 with the same skewness and kurtosis across these 15 items. Items 16 through 21 were distributed normally in all generation conditions.

Factor loadings. Two different sets of factor loadings were used. Table 5 presents the values of the loading for each item. The loadings were reasonably large in the first set of the loadings, while a few items had small loadings of .15 in the second set. I labeled arbitrarily the first set of loadings as "large" and the second set of loadings as "small" when reporting results in the next chapter.

Table 5: *Factor Loadings in Data Generation Model*

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15
Set 1	.4	.6	.8	.4	.6	.8	.4	.6	.8	.4	.6	.8	.4	.6	.8
Set 2	.15	.6	.8	.15	.6	.8	.4	.6	.8	.4	.6	.8	.15	.6	.8

A total of 180 conditions were used to generate the data obtained from the above five design factors. Among them, 162 were formed from the conditions with missing data (3 sample sizes x 3 missing patterns x 3 percentages of missingness x 3 types of distributions x 2 sets of factor loadings), and 18 were conditions with no missing data (3 sample sizes x 1 missing patterns [no missing data] x 1 percentages of missingness [0% missing] x 3 types of distributions x 2 sets of factor loadings). For data generated from each of the conditions, analyses were conducted at both item and parcel levels. All analysis models were correctly specified.

Data Generation Procedure

Data were generated in R (version 2.13.2) based on the model specified in Figure 2. An example of R code for the data generation is presented in Appendix A. First, the correlations among factors were obtained based on the model specification. Appendix B shows the details about how to obtain the correlation matrix among factors:

$$\Sigma_{factor} = \begin{bmatrix} 1 & .4 & .6 \\ .4 & 1 & .74 \\ .6 & .74 & 1 \end{bmatrix}.$$

To generate multivariate normally distributed factor scores with the above inter-factor correlations, three random and un-correlated variables were first generated with a mean of zero and standard deviation of one. The uncorrelated random factor scores were then converted to multivariate data with the inter-factor correlations using the Cholesky decomposition method. Next, independent random error scores were generated for each item with mean of zero and standard deviation of one. Finally, the observed scores for each item were obtained as a weighted

linear combination of factor and error scores using the following formula (e.g., Bernstein & Teng, 1989):

$$X_{ij} = \lambda_i * F_j + \left(\sqrt{1 - {\lambda_i}^2}\right) * E_{ij}$$

where X_{ij} is an observed score on item *i* for individual *j*, F_j is the factor score for person *j*, λ_i indicates the factor loading for the item *i*, and E_{ij} indicates random error.

The above procedure generated multivariate normally distributed data with predefined inter-item correlations. To generate nonnormal data, Fleishman's power transformation method (Fleishman, 1978) was applied to normally distributed data to obtain new scores with the predefined skewness and kurtosis:

$$X_{non-normal} = a + b \cdot X_{normal} + c \cdot X_{normal}^2 + d \cdot X_{normal}^3$$

where X_{normal} is a normally distributed variable as generated previously. The coefficients "*a*", "*b*", "*c*", and "*d*" are constants corresponding to a specific degree of skewness and kurtosis.

After the complete data were generated, a certain criterion was applied to create missing data. Not all variables contained missing values; instead, only six items have missing values; X1, X2, X3, X6, X7, and X8. To create data that are MCAR, randomly selected cases on the specified items were removed. To create data that are MAR, missingness was associated with some variables but not itself. Table 6 indicates how missingness on one variable was related to other variables. For example, missingness on X1 was related to the values of X4. That is, X4 were sorted from smallest to largest, and then the cases with the lowest values of X4 (e.g., 10%) were assigned as missing on X1 with a probability of .9. The rest of the values (highest 90%) were assigned as missing with a probability of .1. Missingness on the variables X2 and X3 were conditioned on the values of X5 and X16, respectively. In the case of MNAR, missingness on variables was related to the variables themselves. Table 6 also indicates the missingness pattern

under MNAR. For example, missingness on X2 was related to values on X2. Values on X2 were sorted from largest to smallest such that missing values were assigned to larger values on X2 with a probability of .9, and the rest were assigned with a probability of .1.

Table 6: *Criterion for Creating Missing Data*

	MCAR	MAR		MNAR	
Missingness	**References**	**References**	**Sorting types**	**References**	**Sorting types**
X1	Randomly	X4	Increasing	X1	Increasing
X2	Randomly	X5	Increasing	X2	Decreasing
X3	Randomly	X16	Increasing	X3	Increasing
X6	Randomly	X11	Decreasing	X6	Increasing
X7	Randomly	X12	Increasing	X7	Increasing
X8	Randomly	X13	Increasing	X8	Increasing

Data Analysis

M*plus* 6.1 (Muthén & Muthén, 1998-2008) was used for the data analyses using the FIML estimation method. For each dataset, two different models were considered; one was based on the item level and the other was based on the parcel level.

As described above, data were generated at the item level. Parcel scores for items associated with F1were then calculated based on the factorial algorithm by averaging the available individual items within a particular subset. To implement the factorial algorithm procedure, a CFA model was conducted on items 1 to 15. The items were sorted based on the magnitudes of the loadings (labeled as 1^{st} to 15^{th} from the largest to smallest). Then, items were assigned to one of the three parcels. For example the first parcel will have the items with loadings ranked 1^{st}, 6^{th}, 12^{th}, and 13^{th}. See Table 7 for details. The items associated with F2 and F3 were not parceled. Appendix C shows the part of the R code that is relevant to this procedure. For both the item and the parcel level analyses including the CFA model used to obtain the rank of factor loadings, FIML estimation method was used. As described in the previous chapter,

FIML behaves similar to ML estimation method for data with no missing values (Enders, 2001). When missing data are present, FIML computes log likelihood for each case based on the available items and then the log likelihoods across all cases are summed to obtain the log likelihood for the sample. In other words, for analysis with missing data, no data imputation was executed. Note that for parcel level analysis, a particular parcel score for a case was missing only if the values of all items forming that parcel were missing.

Table 7: *Assignment of Items to Parcels*

	Parcel 1	**Parcel 2**	**Parcel 3**
Items*	1^{st}, 6^{th}, 7^{th}, 12^{th}, 13^{th}	2^{nd}, 5^{th}, 8^{th}, 11^{th}, 14^{th}	3^{rd}, 4^{th}, 9^{th}, 10^{th}, 15^{th}

* Values indicate the order of factor loadings from the CFA model.

Assessment of the Models and the Parameters

For each condition, overall model-data fit was evaluated based on the chi-square test, CFI, RMSEA, and SRMR values. For the chi-square test, rejection rate based on the alpha level of .05 was reported for each condition. Fit indices were compared to the values suggested by Hu and Bentler (1999). CFI values larger than .95, RMSEA and SRMR values smaller than .06 and .08 were considered reasonable model fits, respectively (Hu & Bentler, 1999).

For parameter estimates, I focused on two direct effects among factors, that is, from F1 to F2 and from F1 to F3, and the covariance between the disturbances of F2 and F3. Relative bias was computed for both point estimates and standard errors associated with parameter estimates using the formula

$$\text{Bias} = \frac{\hat{\theta} - \theta}{\theta},$$

where $\hat{\theta}$ and θ indicate the mean of estimates and the population parameter, respectively. For standard errors, the population value was approximated by using the standard deviation of parameter estimates based on complete data for both the item level analysis and the parcel level

analysis. Few different rule-of-thumbs were suggested for parameter estimation bias. Muthén, Kaplan, and Hollis (1987) suggested using .10 (in absolute value) as the cutoff. Hoogland and Boomsma (1998) suggested biases smaller than .05 (in absolute value) for point estimates and .10 for standard errors be acceptable. I followed Hoogland and Boomsma's criteria because it is more conservative.

In order to explore the effects of the five design factors (sample size, percentage of missingness, missing mechanisms, nonnormality of the data, and factor loadings) on model-data fit indices (i.e., CFI, RMSEA, and SRMR) and the parameter estimates from both parcel and item level analyses, a series of 5-way ANOVAs were conducted. ANOVA test has three assumptions: independency of error scores, normality of error scores, and homogeneity of error variances across groups. Because I randomly drew 2000 datasets for each condition, dependency of error scores is not a concern. Homogeneity of error variances was examined by using Levene's test. Because the sample size across groups was the same (the sample size for the majority of the combined groups was 2000; the sample size ranged from 1995 to 1999 for other groups; see Table 8 and Table 9), Levene's test is robust to the violation of homogeneity of error variances. In addition, the statistical testing for ANOVA is likely to be significant due to the large sample size involved in the analysis as that is in my study (Alhija & Wisenbaker, 2006). Therefore, partial eta squares were considered to evaluate the importance of each design factor. Eta square is the proportion of the variance in the outcome variable that is explained by a specific factor while controlling for other independent variables.

CHAPTER 4

RESULTS

As described in the previous chapter, 180 conditions were created by combining five factors for data generation: 2 sets of factor loadings (small vs. large); sample sizes (100, 300, and 1000); missing data pattern (No missing, MCAR, MAR, and MNAR); percent of missingness (10%, 20%, and 40%); and distribution of items (normal distribution, nonnormal distribution with skewness of 1.0 and kurtosis of 1.5, and nonnormal distribution with skewness of 1.75 and kurtosis of 3.75). For each condition, 2000 datasets were generated. For each generated dataset, analysis was conducted at both parcel and item levels using full information likelihood estimation method. For parcel level analysis, a CFA model was conducted for the 15 items used to form parcels. Three parcels were then formed based on the ranking order of the factor loadings. In this chapter, I summarized the results of the analyses. Specifically, for each condition, model convergence, rejection rate based on the chi-square test, and fit indices were reported. Parameter estimates and their standard errors were examined by computing relative biases. ANOVAs were run to examine the main and interaction effects of the five design factors on model data-fit indices (e.g., CFI), parameter estimates (e.g., the direct effect from factor F1 to F2) and standard errors associated with parameter estimates.

Convergence

Frequencies of model non-convergences were reported in Table 8 for conditions with the first set of factor loadings and in Table 9 for conditions with the second set of factor loadings. Overall, only a few replications (or none) encountered model non-convergence. The frequencies did not differ across two sets of factor loadings. Specifically, all replications converged to a solution for

all conditions except for a few conditions associated with the sample size of 100 and 40% missing data when analyses were conducted based on the item level. The occurrence of non-convergence did not seem to vary across missing data mechanisms and the degree of nonnormality of data. Parcel level analysis did not encounter any non-convergence issues. Replications with non-convergence problems were excluded from further analyses.

Rejection Rate based on Chi-square Test

Table 10 and Table 11 report the rejection rate based on chi-square test at the alpha level of .05 for conditions with the first set of factor loadings and the second set of factor loadings, respectively. Because the models were considered correctly specified, I expected the rejection rate (i.e., empirical alpha) to be around 5%. For alpha level of .05, the standard error of the alpha was around .01. Thus, the 95% of the confidence interval (CI) for the alpha was .04-.06. The rejection rate was lower for analyses based on the parcel level than that based on the item level. In addition, the rejection rates from parcel level analysis were within the CI for majority of the conditions regardless of missing mechanisms and percentages of missingness, while rejection rates from the item level analysis fell outside of the CI when the sample size was less than 1000 and when the data were nonnormally distributed. It can be observed from both tables that the empirical alpha tended to be larger (i.e., more likely to reject a correctly specified model) for analyses based on the item level compared to the analysis based on the parcel level. For example, for the conditions with normally distributed data, sample size of 100, and the percentage of missingness of 10%, the rejection rates were around 30% and around 7% for the item level analysis and parcel level analysis, respectively. Overall, for both the item and the parcel level analyses the rejection rates based on complete data were slightly lower than those based on incomplete data but did not seem to vary across missing mechanisms.

Table 8: *Frequency of Model Non-convergence for Each Condition for the First Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	0	0	0	1	0	0	4	0	0	1
		Parcel	0	0	0	0	0	0	0	0	0	0
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
Sk=1, K=1.5	100	Item	0	0	0	5	0	0	1	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
Sk=1.75, K=3.75	100	Item	0	0	0	5	0	0	2	0	0	1
		Parcel	0	0	0	0	0	0	0	0	0	0
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0

Table 9: *Frequency of Model Non-convergence for Each Condition for the Second Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	0	0	0	2	0	0	2	0	0	1
		Parcel	0	0	0	0	0	0	0	0	0	0
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
Sk=1, K=1.5	100	Item	0	0	0	1	0	0	1	0	0	2
		Parcel	0	0	0	0	0	0	0	0	0	0
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
Sk=1.75, K=3.75	100	Item	0	0	0	3	0	0	3	0	0	1
		Parcel	0	0	0	0	0	0	0	0	0	0
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0

Table 10: *Chi-square Rejection Rate (%) Based on Alpha Level of .05 for the First Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	27	30	35	81	30	35	77	31	37	81
		Parcel	8	7	7	7	7	7	7	7	7	7
	300	Item	11	11	11	14	10	11	13	11	11	12
		Parcel	6	6	6	7	6	6	6	6	6	6
	1000	Item	5	6	6	6	5	6	7	6	6	7
		Parcel	6	6	5	6	6	5	6	6	6	6
Sk=1, K=1.5	100	Item	48	52	56	89	51	51	84	52	52	87
		Parcel	8	8	9	8	8	8	8	8	8	8
	300	Item	26	26	24	28	24	21	26	26	23	28
		Parcel	5	5	5	6	5	6	5	5	6	5
	1000	Item	18	19	18	18	17	16	18	17	17	19
		Parcel	6	5	5	6	5	6	6	5	6	5
Sk=1.75, K=3.75	100	Item	77	79	83	96	77	80	95	75	78	95
		Parcel	7	7	7	7	6	7	6	6	7	7
	300	Item	60	60	61	65	59	59	60	55	56	62
		Parcel	7	6	6	6	6	6	6	6	6	6
	1000	Item	57	56	56	57	54	53	51	51	53	57
		Parcel	6	6	6	6	6	6	6	6	5	6

Table 11: *Chi-square Rejection Rate (%) Based on Alpha Level of .05 for the Second Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	26	31	36	79	30	35	79	30	35	82
		Parcel	8	7	8	8	7	8	8	8	8	9
	300	Item	10	10	11	13	11	10	13	10	11	13
		Parcel	6	6	6	6	6	6	6	6	6	6
	1000	Item	6	6	6	7	6	6	7	6	7	7
		Parcel	4	5	5	5	5	5	5	5	5	5
Sk=1, K=1.5	100	Item	43	48	53	89	47	51	85	47	52	86
		Parcel	9	8	9	8	9	8	7	8	8	8
	300	Item	22	24	24	28	21	22	26	22	23	27
		Parcel	6	6	6	6	5	6	5	5	6	6
	1000	Item	17	18	18	18	17	16	17	16	16	19
		Parcel	5	5	6	5	5	6	6	6	5	5
Sk=1.75, K=3.75	100	Item	70	73	78	96	72	75	93	70	73	93
		Parcel	8	8	8	8	8	8	8	8	8	8
	300	Item	47	47	49	53	45	46	48	44	45	51
		Parcel	6	6	6	6	6	6	6	6	6	6
	1000	Item	45	44	46	46	44	43	43	40	41	46
		Parcel	6	6	6	6	6	6	6	6	6	6

When the sample size was 100 and the analysis was conducted based on the item level, the rejection rate tended to increase as the percentage of missingness increased; whereas when the sample size was greater than 100, the rejection rate was relatively stable across different percentage of missingness. In addition, for item level analysis, the empirical alpha in general tended to be smaller as the sample size increased and the data demonstrated normality. For example, when the analysis was conducted based on the item level with sample size of 100, and missingness of 10% for MCAR, the rejection rates were about 79%, 52% and 30% for highly skewed, moderately skewed, and normally distributed data, respectively. For the parcel level analysis, the rejection rates were stable across levels of percentage of missingness, sample sizes, the degree of nonnormality of the data, and the missing data mechanisms. Specifically, the rejection rates were in the range of 4% to 9% across all conditions.

Fit Indices

The mean of CFIs, RMSEAs, and SRMRs for each condition are reported in this section. For each condition, the percentage of models rejected based on the recommended cutoff was also reported. I adopted the recommended cutoffs of .95, .06, and .08 for CFI, RMSEA, and SRMR, respectively (Hu & Bentler, 1999). In addition, a 5-way ANOVA was conducted to examine the impact of the design factors on each fit index.

CFI

Table 12 and Table 13 present the mean of the CFI values by condition with the first set of factor loadings and the second set of factor loadings, respectively. The mean CFIs were almost identical across the two sets of factor loadings for the both item level analysis and the parcel level analysis. When the analysis was the based on item level, several findings were observed: First, the mean CFI decreased (indicating more likely to reject the model) as the

degree of nonnormality of the data increased when the sample size was 100. For example under MCAR, the mean CFI was .97, .95, and .92 for data with normal distributions, moderate skewed distributions, and highly skewed distributions, respectively. The pattern was similar across missing mechanisms and percentage of missingness. Second, the mean CFI tended to decrease with larger percentage of missingness when the sample size was 100. For example, under MAR, the mean CFI was .97, .96, and .89 for normally distributed data with 10%, 20% and 40% of missingness. Third, the mean CFI was higher than .97 for all conditions when sample size was 300 or more. Fourth, the mean CFI was comparable across different missing mechanisms. When the analysis was conducted at the parcel level, the mean CFIs were higher than .99 across all conditions.

Tables 14 and 15 report percentages of replications with CFI values smaller than .95 for each condition. Since the models were correctly specified, values smaller than .95 suggested that the model and the data did not fit reasonably well. Consistent with the results reported above, when the analysis was conducted at the item level, models were more likely to be rejected for smaller sample size (e.g., 23% for sample size of 100 under MCAR with 10% missing and normal distribution), larger percentages of missingness (e.g., 23%, 30%, and 80% as the percentage of missingness increased from 10% to 40%), and when data deviated from normality (e.g., 23%, 44%, and 82% under MCAR for sample size of 100 with 10% missing). Missing mechanisms did not appear to impact the rejection rate. When the analysis was conducted at the parcel level, the rejection rate based on CFI was less than 4% across all conditions and was 0% when sample size reached to 300, regardless of the degree of nonnormality, missing mechanism, and percentage of missingness.

To further understand the impact of the missing mechanism, percentage of missingness, level of analysis (i.e., item level vs. parcel level), degree of nonnormality, and sample size on the model CFI, a 5-way ANOVA was conducted on sample CFIs. The column "CFI" in Table 16 shows the eta squares for the main and the interaction effects for sample CFIs. The eta squares indicate the proportion of the variation in the outcome variable (i.e., CFI) that was explained by a specific variable, controlling for other variables in the model. Among all main and interaction effects, the largest eta square, .48, was due to sample size, which indicates that 48% of the variation in sample CFIs was explained by the sample size, holding all other factors constant. The next largest eta squares were due to the interaction between level of analysis and sample size (.36), main effect of level of analysis (.32), the interaction between percentage of missingness and sample size (.14), and the three-way interaction effect among level of analysis, percentage of missingness, and sample size (.14). A small proportion of variation in sample CFIs was explained by percentage of missingness (.08), the degree of nonnormality (.07), the interaction between level of analysis and percentage of missingness (.08), between level of analysis and the degree of nonnormality (.06), the degree of nonnormality and sample size (.05), and the 3-way interaction among level of analysis, sample size and the degree of nonnormality (.04). All other effects had zero eta squares. Table 16 reports only two of the 3-way interaction effects and no 4-way and 5-way interaction effects because the corresponding eta squares were all zero (not only for CFI, but also all other dependent variables I examined in this study).

Table 12: *Mean of CFIs for Each Condition for the First Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	.97	.97	.96	.90	.97	.96	.90	.97	.96	.89
		Parcel	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
	300	Item	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
		Parcel	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	Item	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Parcel	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Sk=1, K=1.5	100	Item	.96	.95	.95	.88	.95	.95	.88	.95	.94	.87
		Parcel	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
	300	Item	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
		Parcel	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	Item	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Parcel	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Sk=1.75, K=3.75	100	Item	.92	.92	.91	.83	.92	.91	.84	.92	.91	.83
		Parcel	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
	300	Item	.98	.98	.98	.97	.98	.98	.97	.98	.98	.97
		Parcel	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	Item	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
		Parcel	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note: although not reported in the table, the standard error (SE) of the CFIs from the parcel level analysis was smaller than those from the item level analysis across all design factors. For example, under small sample size and MCAR, SE for CFIs was .025, .029 and .038 for the item level analysis with normal, moderately skewed, and highly skewed distributions, respectively. However, the corresponding SE from parcel level analysis was .013, .015, and .015, respectively. In addition, the SE was larger as the percentages of missingness increased, sample size decreased, and missing mechanisms changed from MCAR to MNAR.

Table 13: *Mean of CFIs for Each Condition for the Second Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	.97	.97	.96	.90	.97	.96	.89	.96	.96	.89
		Parcel	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
	300	Item	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
		Parcel	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	Item	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Parcel	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Sk=1, K=1.5	100	Item	.96	.95	.95	.88	.95	.95	.88	.95	.94	.87
		Parcel	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
	300	Item	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
		Parcel	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	Item	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Parcel	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Sk=1.75, K=3.75	100	Item	.93	.92	.91	.83	.92	.91	.83	.92	.91	.83
		Parcel	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
	300	Item	.98	.98	.98	.98	.98	.98	.98	.98	.98	.98
		Parcel	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	Item	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
		Parcel	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 14: *Percentage (%) of Replications with CFI Smaller than .95 for Each Condition with the First Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	18	23	30	80	24	32	78	28	36	83
		Parcel	1	1	2	2	2	2	2	2	2	2
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
Sk=1, K=1.5	100	Item	38	44	54	90	45	52	86	47	55	89
		Parcel	2	2	4	4	2	4	4	2	3	4
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
Sk=1.75, K=3.75	100	Item	78	82	86	98	81	84	97	79	84	97
		Parcel	3	3	3	3	3	3	3	3	3	3
	300	Item	1	2	3	7	2	3	6	2	4	9
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0

Table 15: *Percentage (%) of Replications with CFI Smaller than .95 for Each Condition with the Second Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	20	26	33	80	27	34	81	29	38	84
		Parcel	3	3	3	3	3	3	3	2	3	3
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
Sk=1, K=1.5	100	Item	37	45	52	90	45	52	86	47	55	89
		Parcel	3	4	4	4	4	4	4	3	4	5
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
Sk=1.75, K=3.75	100	Item	74	78	83	98	78	81	96	77	82	96
		Parcel	4	4	4	6	5	5	5	5	5	5
	300	Item	1	2	2	5	1	2	4	2	3	7
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0

Table 16: *Partial Eta Squares Based on ANOVA*

	CFI	RMSEA	SRMR	L12	L13	L23	SE_L12	SE_L13	SE_L23
Level of Analysis* (L)	.32	.08	.81	.00	.00	.00	.01	.02	.00
Missing Mechanism (MM)	.00	.00	.00	.00	.00	.00	.00	.00	.00
Missing Percentage (MP)	.08	.01	.23	.00	.00	.00	.00	.00	.00
Normality (N)	.07	.02	.07	.04	.07	.03	.00	.00	.03
Sample Size (SS)	.48	.38	.93	.00	.00	.00	.95	.96	.86
L * MM	.00	.00	.00	.00	.00	.00	.00	.00	.00
L * MP	.08	.01	.20	.00	.00	.00	.00	.00	.00
L * N	.06	.02	.03	.00	.00	.00	.00	.00	.00
L * SS	.36	.08	.54	.00	.00	.00	.00	.00	.00
MM * MP	.00	.00	.00	.00	.00	.00	.00	.00	.00
MM * N	.00	.00	.00	.00	.00	.00	.00	.00	.00
MP * SS	.00	.00	.00	.00	.00	.00	.00	.00	.00
MP * N	.00	.00	.00	.00	.00	.00	.00	.00	.00
MP * SS	.14	.02	.14	.00	.00	.00	.00	.00	.00
N *SS	.05	.00	.01	.00	.00	.00	.00	.00	.01
L * MP * SS	.14	.02	.13	.00	.00	.00	.00	.00	.00
L * N * SS	.04	.00	.01	.00	.00	.00	.00	.00	.00

* A dummy variable indicating 0 for the parcel level analysis and 1 for the item level analysis.

Note: L12 stands for the direct effect from factor F1 to F2, L13 is for the direct effect from F1 to F3, and L23 is for the covariance between disturbances of F2 with F3, and SE stands for standard error.

RMSEA

RMSEA values smaller than .06 indicate good model-data fit (Hu & Bentler, 1999). Table 17 and Table 18 summarize the mean of RMSEAs for conditions with the first set of factor loadings and the second set of factor loadings, respectively. With a few exceptions, the mean RMSEAs were smaller than .06 for all conditions. The exceptions were associated with sample size 100 and 40% missingness in the dataset when the analyses were conducted at the item level. For these conditions, the mean RMSEAs were in the range of .06 to .08. Unexpectedly, the standard error of RMSEAs from the item level analysis was smaller than those from the parcel level analysis. The column "RMSEA" in Table 16 shows the ANOVA result for RMSEAs. Consistent with the findings based on the mean RMSEAs, the largest eta square was associated with the main effect of sample size (eta square of .38), followed by the main effect of the level of analysis (.08) and the interaction effect between level of analysis and sample size (.08). All other main effects or interaction effects had small eta squares (all $\leq$.02; many were zero).

The percentages of replications with RMSEA larger than .06 were reported in Tables 19 and 20. A value larger than .06 indicates that the model is rejected based on RMSEA. As can be seen from both tables, the rejection rates based on RMSEA were all zero for conditions with sample sizes of 300 and 1000. However, when the sample size was 100, the rejection rates were greater than zero for both item level analysis and parcel level analysis. The percentages increased as percentage of missingness increased for item level analysis under all missing mechanisms. The increment accelerated as percentages of missingness reached 40%. For example under MCAR, the rejection rate increased only by 2% for 20% missingness whereas the rejection rate reached 55% for 40% missing data (see Table 20). In contrast, the percentages were stable for the parcel level analysis regardless of missing mechanisms and percentages of missingness.

Table 17: *Mean of RMSEA for Conditions with the First Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	.03	.03	.03	.06	.03	.03	.06	.03	.03	.06
		Parcel	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02
	300	Item	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
	1000	Item	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
Sk=1, K=1.5	100	Item	.04	.04	.04	.06	.04	.04	.06	.04	.04	.06
		Parcel	.03	.03	.03	.03	.03	.03	.03	.03	.03	.03
	300	Item	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
	1000	Item	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
Sk=1.75, K=3.75	100	Item	.05	.05	.05	.08	.05	.05	.07	.05	.05	.07
		Parcel	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02
	300	Item	.02	.03	.03	.03	.02	.02	.03	.02	.02	.03
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
	1000	Item	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01

Table 18: *Mean of RMSEA for Conditions with the Second Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	.03	.03	.03	.06	.03	.03	.06	.03	.03	.06
		Parcel	.03	.03	.03	.03	.03	.03	.03	.03	.03	.03
	300	Item	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
	1000	Item	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
Sk=1, K=1.5	100	Item	.04	.04	.04	.06	.04	.04	.06	.04	.04	.06
		Parcel	.03	.03	.03	.03	.03	.03	.03	.03	.03	.03
	300	Item	.02	.02	.02	.02	.01	.02	.02	.02	.02	.02
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
	1000	Item	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
Sk=1.75, K=3.75	100	Item	.05	.05	.05	.07	.05	.05	.07	.05	.05	.07
		Parcel	.03	.03	.03	.03	.03	.03	.03	.03	.03	.03
	300	Item	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
	1000	Item	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01

Table 19: *Percentage (%) of Replications with RMSEA Larger than .06 for Each Condition with the First Set of Factor*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	2	3	3	52	3	4	51	3	4	55
		Parcel	15	15	14	15	15	14	15	14	15	14
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
Sk=1, K=1.5	100	Item	7	8	9	64	8	9	58	8	9	60
		Parcel	16	17	16	17	17	16	16	17	16	17
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
Sk=1.75, K=3.75	100	Item	28	33	35	83	29	32	75	26	30	76
		Parcel	14	15	15	15	14	14	14	14	14	15
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0

Table 20: *Percentage (%) of Replications with RMSEA Larger than .06 for Each Condition with the Second Set of Factor*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	2	2	4	55	3	4	51	2	4	52
		Parcel	16	16	16	16	15	16	16	16	17	16
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
Sk=1, K=1.5	100	Item	5	7	9	65	7	9	59	7	9	61
		Parcel	17	16	16	16	16	16	16	17	16	17
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
Sk=1.75, K=3.75	100	Item	20	22	27	81	21	25	74	19	22	73
		Parcel	16	15	16	16	16	16	16	16	16	16
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0

The rejection rates for the parcel level analysis ranged between 14% and 17% for all missing percentages. Unexpectedly, the parcel level analysis under small sample size resulted in higher rejection rates compared to item level analysis when 10% or 20% of the data were missing with normally distributed or moderately skewed data. In contrast, when the percentage of missingness was 40%, the parcel level analysis resulted in smaller rejection rates in comparison to the item level analysis under any condition.

SRMR

A SRMR value smaller than .08 indicates good model-data fit (Hu & Bentler, 1999). The mean of SRMR values is reported in Table 21 for the conditions with the first set of factor loadings and Table 22 for the conditions with the second set of factor loadings. The values were almost identical across two sets of factor loadings under any condition. Consistent with the results based on RMSEA, the mean SRMRs were below the cutoff for most conditions. The mean of the SRMRs increased as percentages of missingness increased under item level analysis with small sample size. For example the mean SRMR values ranged between .06 and .07 for 10% of missingness; whereas they ranged between .09 and .10 for 40% of missingness, under all missing mechanisms. Compared to the item level analysis, the SE from the parcel level analysis was smaller for some conditions but larger for other conditions.

Results from the ANOVA on SRMR values are shown in the column "SRMR" in Table 16. Similar to CFI and RMSEA, the largest eta square was associated with the main effect of sample size, indicating that 93% of the variation in sample SRMRs was explained by sample size controlling for all other independent variables. The next largest eta squares were due to the main effect of level of analysis (eta square of .81), the interaction effect between level of analysis and sample size (.53), the main effect of percentage of missingness (.23), the interaction effect

between level of analysis and percentage of missingness (.20), the interaction effect between percentage of missingness and sample size (.14), and the 3-way interaction among level of analysis, percentage of missingness, and sample size.

Percentages of replications with SRMR larger than .08 are reported in Tables 23 and 24 for the first and the second sets of factor loadings. SRMR larger than .08 indicates that the model does not fit the data. As can be seen from these two tables, models based on parcel level data were never rejected for all of the conditions (i.e., 0% rejection rate). However, up to 99% of models were rejected when the analysis was conducted based on the item level when the sample size was small (i.e., 100). For these conditions, rejection rates increased as percentage of missingness increased. For example under MCAR, rejection rates were 0%, 1% and 71% for normally distributed data with 10%, 20% and 40% of missingness, respectively. Moreover, the rejection rates increased as the data deviated more from normality. For example, for 20% missingness under MCAR, the percentages of replications with SRMR larger than .08 were 1% for normally distributed data (i.e., $Sk = 0$), 4% for moderately skewed data (i.e., $Sk = 1$), and 26% for highly skewed data (i.e., $Sk = 1.75$). In addition, the rejection rate was the largest for the data under MNAR and the smallest for MCAR. For example rejection rates were 79%, 87% and 92% for MCAR, MAR, and MNAR, respectively (see Table 24). The rejection rate was zero for all conditions with sample sizes of 300 and 1000.

Parameter Estimates and Standard Errors

For parameter estimates and the associated standard errors, I focused on the parameters of the direct effects among the factors (i.e., F1→F2 and F1→F3) and the covariance between the disturbances associated with F2 and F3 (i.e., the structural part of the model).

Table 21: *Mean of SRMR for Conditions with the First Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	.06	.06	.07	.09	.06	.07	.09	.06	.07	.09
		Parcel	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04
	300	Item	.03	.04	.04	.04	.04	.04	.04	.04	.04	.04
		Parcel	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02
	1000	Item	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
Sk=1, K=1.5	100	Item	.06	.07	.07	.09	.07	.07	.09	.07	.07	.09
		Parcel	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04
	300	Item	.04	.04	.04	.04	.04	.04	.05	.04	.04	.05
		Parcel	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02
	1000	Item	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
Sk=1.75, K=3.75	100	Item	.07	.07	.08	.10	.07	.08	.10	.07	.08	.10
		Parcel	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04
	300	Item	.04	.04	.04	.05	.04	.04	.05	.04	.04	.05
		Parcel	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02
	1000	Item	.02	.02	.02	.03	.02	.02	.03	.02	.02	.03
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01

Table 22: *Mean of SRMR for Conditions with the Second Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	.06	.06	.07	.09	.07	.07	.09	.07	.07	.09
		Parcel	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04
	300	Item	.04	.04	.04	.04	.04	.04	.04	.04	.04	.05
		Parcel	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02
	1000	Item	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
Sk=1, K=1.5	100	Item	.06	.07	.07	.09	.07	.07	.09	.07	.07	.09
		Parcel	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04
	300	Item	.04	.04	.04	.04	.04	.04	.05	.04	.04	.05
		Parcel	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02
	1000	Item	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
Sk=1.75, K=3.75	100	Item	.07	.07	.08	.10	.07	.08	.10	.07	.08	.10
		Parcel	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04
	300	Item	.04	.04	.04	.05	.04	.04	.05	.04	.04	.05
		Parcel	.02	.02	.02	.03	.02	.03	.03	.03	.03	.03
	1000	Item	.02	.02	.02	.03	.02	.02	.03	.02	.02	.03
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01

Table 23: *Percentage (%) of Replications with SRMR Larger than .08 for Each Condition with the First Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	0	0	1	71	0	2	78	0	2	87
		Parcel	0	0	0	0	0	0	0	0	0	0
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
Sk=1, K=1.5	100	Item	0	0	4	88	0	6	87	0	7	92
		Parcel	0	0	0	0	0	0	0	0	0	0
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
Sk=1.75, K=3.75	100	Item	2	8	26	98	6	25	95	7	29	98
		Parcel	0	0	0	0	0	0	0	0	0	0
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0

Table 24: *Percentage (%) of Replications with SRMR Larger than .08 for Each Condition with the Second Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	0	0	2	79	1	3	87	0	4	92
		Parcel	0	0	0	0	0	0	0	0	0	0
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
Sk=1, K=1.5	100	Item	0	1	5	88	1	5	87	1	7	92
		Parcel	0	0	0	0	0	0	0	0	0	0
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
Sk=1.75, K=3.75	100	Item	2	7	29	99	7	29	98	8	30	99
		Parcel	0	0	0	0	0	0	0	0	0	0
	300	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0
	1000	Item	0	0	0	0	0	0	0	0	0	0
		Parcel	0	0	0	0	0	0	0	0	0	0

These parameter estimates and their standard errors were assessed by computing the relative bias using the formula I provided in the previous chapter:

$$\text{Bias} = \frac{\hat{\theta} - \theta}{\theta},$$

where $\hat{\theta}$ is the mean of the estimates and θ is the population value. In the following sections, I report the relative bias of estimates and standard errors for each parameter as well as the ANOVA results for the biases.

Point Estimates

Relative biases of the parameter estimates are reported in Tables 24 and 26 for F1→ F2, in Tables 27 and 28 for F1 → F3, and in Tables 29 and 30 for the covariance between the disturbances F2 and F3. The population values were .4, .6, and .5 for F1→F2, F1→F3, and the covariance between the two disturbances, respectively.

First, similar findings were obtained across all three parameters in terms of bias. Second, the relative bias did not vary across two sets of factor loadings. Third, missing mechanisms, percentage of missingness, and sample size did not seem to influence the bias of the parameter estimates. Fourth, bias increased as the degree of the nonnormality of the data increased, regardless of sample size, missing mechanism, and percentages of missingness. Specifically, when the data were normally distributed (i.e., *Sk* = 0), there was no indication of bias under any conditions. As the nonnormality increased (i.e., skewness increased), estimates of the direct effects showed negative bias indicating that the obtained estimates tended to be too small. The relative biases were less than .05 for conditions with moderately nonnormal distribution but up to .13 for conditions with highly nonnormal distribution. Fifth, parcel level analysis yielded slightly smaller bias than the item level analysis under all conditions. For example, for the item

level analysis with 40% MNAR and highly skewed data (i.e., *Sk* = 1.75), the biases ranged from -.11 to -.13 for F1 → F2, whereas the corresponding parcel level analysis resulted in biases between -.08 and -.09 (see Table 26). Overall, the absolute values of the biases were smaller than .05 for normal and moderately nonnormal data and greater than .05 for highly skewed data for both parcel level and item level analyses. Unlike the two direct effects, estimates of the covariance between the disturbances were mostly positively biased when the data were highly skewed, indicating the estimated parameter values on average were larger than the population values of the parameter. Last, when there was bias, item level analysis and parcel level analysis always resulted in bias in the same direction (i.e., either both positively biased or both negatively biased).

Results from ANOVA were consistent with the above results. The eta squares were zero for all main and interaction effects with only one exception, that is, the main effect of data nonnormality (see the columns of "L12", "L13", and "L23" in Table 16). The eta squares for this main effect were .04 for the estimates of F1→F2, .07 for the estimates of F1→F3, and .03 for the estimates of covariance between the disturbances, indicating that 3% to 7% of variance in these parameter estimates was explained by the main effect of nonnormality, controlling for all other main and interaction effects.

Standard Errors

Relative biases of standard errors (SEs) for parameter estimates were reported in Tables 31 to 36 for two direct effects (F1→ F2 and F1→ F3) and the covariance (F2 ↔ F3) for conditions with the first and the second sets of factor loadings. It has been suggested that relative bias of SEs smaller than .10 (in absolute value) are acceptable (Hoogland & Boomsma, 1998).

Table 25: *Bias of the Direct Effect from F1 to F2 for Each Condition with the First Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
		Parcel	.00	.00	.00	.00	.00	.00	.01	.01	.01	.00
	300	Item	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
		Parcel	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
	1000	Item	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
		Parcel	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Sk=1, K=1.5	100	Item	-.03	-.03	-.02	-.02	-.03	-.02	-.02	-.03	-.02	-.02
		Parcel	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
	300	Item	-.02	-.02	-.01	-.01	-.02	-.01	-.01	-.02	-.01	-.02
		Parcel	-.02	-.02	-.01	-.01	-.02	-.01	-.01	-.02	-.01	-.01
	1000	Item	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
		Parcel	-.01	-.01	-.02	-.02	-.01	-.01	-.01	-.01	-.02	-.01
Sk=1.75, K=3.75	100	Item	-.11	-.11	-.12	-.12	-.11	-.12	-.12	-.11	-.11	-.12
		Parcel	-.10	-.10	-.10	-.10	-.10	-.09	-.09	-.09	-.10	-.09
	300	Item	-.11	-.11	-.11	-.11	-.11	-.11	-.11	-.11	-.11	-.11
		Parcel	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.09
	1000	Item	-.11	-.11	-.11	-.11	-.11	-.11	-.11	-.10	-.11	-.11
		Parcel	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.08

Table 26: *Bias of the Direct Effect from F1 to F2 for Each Condition with the Second Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
		Parcel	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
	300	Item	.00	.00	.00	.00	.00	.00	.00	.00	.00	-.01
		Parcel	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
	1000	Item	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
		Parcel	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Sk=1, K=1.5	100	Item	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
		Parcel	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
	300	Item	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.02
		Parcel	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01
	1000	Item	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
		Parcel	-.02	-.02	-.02	-.02	-.01	-.01	-.01	-.02	-.02	-.01
Sk=1.75, K=3.75	100	Item	-.12	-.12	-.12	-.12	-.12	-.12	-.12	-.12	-.12	-.13
		Parcel	-.10	-.09	-.09	-.10	-.09	-.09	-.09	-.09	-.09	-.09
	300	Item	-.12	-.12	-.12	-.12	-.12	-.12	-.12	-.11	-.12	-.12
		Parcel	-.10	-.10	-.10	-.09	-.09	-.09	-.10	-.09	-.09	-.09
	1000	Item	-.11	-.11	-.11	-.11	-.11	-.11	-.11	-.10	-.11	-.11
		Parcel	-.09	-.09	-.09	-.09	-.08	-.08	-.09	-.08	-.08	-.08

Table 27: *Bias of the Direct Effect from F1 to F3 for Each Condition with the First Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	.00	.00	.00	-.01	.00	.00	.00	.00	.00	-.01
		Parcel	.00	.00	.00	-.01	.00	.00	.00	.00	.00	-.01
	300	Item	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
		Parcel	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
	1000	Item	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
		Parcel	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Sk=1, K=1.5	100	Item	-.02	-.02	-.03	-.03	-.02	-.03	-.03	-.02	-.03	-.04
		Parcel	-.02	-.02	-.03	-.03	-.01	-.03	-.03	-.02	-.03	-.03
	300	Item	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
		Parcel	-.02	-.02	-.01	-.01	-.02	-.01	-.01	-.02	-.01	-.01
	1000	Item	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
		Parcel	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01
Sk=1.75, K=3.75	100	Item	-.12	-.12	-.12	-.12	-.12	-.12	-.12	-.12	-.12	-.13
		Parcel	-.10	-.10	-.10	-.10	-.10	-.10	-.10	-.10	-.10	-.09
	300	Item	-.11	-.11	-.11	-.11	-.11	-.11	-.11	-.11	-.11	-.11
		Parcel	-.09	-.09	-.09	-.08	-.09	-.09	-.09	-.09	-.09	-.08
	1000	Item	-.11	-.11	-.11	-.11	-.11	-.11	-.11	-.10	-.11	-.11
		Parcel	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.08

Table 28: *Bias of the Direct Effect from F1 to F3 for Each Condition with the Second Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01
		Parcel	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01
	300	Item	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
		Parcel	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
	1000	Item	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
		Parcel	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Sk=1, K=1.5	100	Item	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.04
		Parcel	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03
	300	Item	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
		Parcel	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01
	1000	Item	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
		Parcel	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01
Sk=1.75, K=3.75	100	Item	-.11	-.11	-.11	-.12	-.11	-.12	-.12	-.11	-.11	-.12
		Parcel	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.09
	300	Item	-.11	-.11	-.11	-.11	-.11	-.11	-.11	-.11	-.11	-.11
		Parcel	-.09	-.09	-.09	-.08	-.09	-.09	-.09	-.08	-.09	-.08
	1000	Item	-.11	-.11	-.11	-.11	-.11	-.11	-.11	-.11	-.11	-.11
		Parcel	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.08

Table 29: *Bias of the Covariance between Disturbances of F2 with F3 for Each Condition with the First Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
		Parcel	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
	300	Item	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
		Parcel	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
	1000	Item	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
		Parcel	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Sk=1, K=1.5	100	Item	-.01	-.01	-.02	-.02	-.01	-.02	-.02	-.01	-.02	-.01
		Parcel	-.01	-.01	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
	300	Item	.01	.00	.01	.01	.00	.01	.01	.00	.01	.01
		Parcel	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
	1000	Item	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
Sk=1.75, K=3.75	100	Item	.07	.07	.08	.08	.07	.08	.08	.07	.07	.08
		Parcel	.06	.06	.06	.06	.06	.06	.05	.06	.06	.05
	300	Item	.09	.09	.09	.09	.09	.09	.09	.08	.09	.09
		Parcel	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07
	1000	Item	.09	.09	.09	.09	.09	.09	.09	.09	.09	.10
		Parcel	.08	.07	.07	.07	.07	.07	.07	.07	.07	.07

Table 30: *Bias of the Covariance between Disturbances of F2 with F3 for Each Condition with the Second Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.02	-.02	-.02
		Parcel	-.01	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
	300	Item	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01
		Parcel	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01
	1000	Item	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
		Parcel	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Sk=1, K=1.5	100	Item	-.02	-.02	-.02	-.01	-.02	-.02	-.02	-.02	-.02	-.01
		Parcel	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
	300	Item	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
		Parcel	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
	1000	Item	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
		Parcel	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
Sk=1.75, K=3.75	100	Item	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08
		Parcel	.06	.06	.06	.06	.05	.05	.05	.05	.05	.05
	300	Item	.09	.09	.09	.09	.09	.09	.09	.08	.09	.09
		Parcel	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07
	1000	Item	.10	.10	.10	.10	.10	.10	.10	.09	.10	.10
		Parcel	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08

As can be observed from these tables, the absolute value of the relative bias was smaller than .05 for all conditions, much smaller than the suggested cutoff of .10. This indicates that the estimates of standard errors were similar to those based on data with no missing values and across different missing mechanism and percentage of missingness (Recall that the population standard errors were approximated based on the complete data with given sample size and skewness/kurtosis). The patterns were similar for both item level analysis and parcel level analysis and across different sets of factor loadings.

ANOVAs for the SEs of the parameter estimates were conducted to further understand the impact of the design factors (such as sample size and missing mechanisms) on standard errors estimates. Although not reported in the results tables, the estimated standard errors were smaller as sample size increased. As shown under columns "SE_L12", "SE_L13", and "SE_L23" in Table 16, for all three parameters (F1→ F2, F1→ F3, and F2 ↔ F3), the largest eta square was associated with sample size, with eta square of .95, .96, and 86, respectively. This means, for example, 95% of the variation in the SEs of estimates of the direct effect from factor F1 to F2 was explained by sample size, controlling for all other main and interaction effects. For both direct effects, a very small portion of variance in the SEs was due to the main effect of level of analysis (eta square of .01 for F1→ F2 and .02 for F1→ F3). For the covariance between the two disturbances, a small portion of the variance of SEs was attributed to the distribution of variables (eta square of .03). All other main and the interaction effects were zero.

Table 31: *Bias of SE of the Direct Effect from F1 to F2 for Each Condition with First Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	.01	.01	.01	.02	.01	.01	.02	.01	.01	.02
		Parcel	.01	.02	.02	.03	.02	.02	.03	.02	.02	.03
	300	Item	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01
		Parcel	-.02	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01
	1000	Item	.02	.02	.02	.03	.02	.02	.03	.02	.03	.03
		Parcel	.02	.02	.03	.03	.02	.03	.03	.02	.03	.03
Sk=1, K=1.5	100	Item	.00	.00	.01	.01	.00	.01	.01	.00	.01	.02
		Parcel	.00	.00	.01	.01	.00	.01	.01	.00	.01	.01
	300	Item	-.01	-.01	.00	.01	-.01	.00	.01	-.01	.01	.01
		Parcel	-.01	-.01	.01	.01	-.01	.01	.01	.00	.01	.01
	1000	Item	.03	.03	.02	.03	.03	.02	.03	.03	.03	.03
		Parcel	.02	.03	.03	.04	.03	.03	.04	.03	.04	.04
Sk=1.75, K=3.75	100	Item	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	.00
		Parcel	-.02	-.02	-.02	-.01	-.02	-.01	-.01	-.02	-.01	-.01
	300	Item	.00	.00	.00	.01	.00	.00	.01	.00	.01	.01
		Parcel	.00	.00	.00	.01	.00	.00	.01	.00	.01	.01
	1000	Item	.03	.03	.04	.04	.03	.04	.04	.04	.04	.05
		Parcel	.03	.03	.04	.04	.03	.04	.04	.04	.04	.05

Table 32: *Bias of SE of the Direct Effect from F1 to F2 for Each Condition with the Second Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	.00
		Parcel	-.01	-.01	-.01	.00	-.01	.00	.00	-.01	.00	.00
	300	Item	.00	.00	.00	.01	.00	.00	.01	.00	.01	.01
		Parcel	.00	.01	.01	.02	.01	.01	.02	.01	.01	.01
	1000	Item	-.01	-.01	-.01	.00	-.01	-.01	.00	.00	.00	.00
		Parcel	-.01	.00	.00	.01	.00	.00	.01	.00	.00	.01
Sk=1, K=1.5	100	Item	.01	.01	.01	.01	.01	.01	.01	.01	.01	.02
		Parcel	.00	.00	.01	.01	.00	.01	.01	.01	.01	.01
	300	Item	.00	.00	.00	.01	.00	.00	.01	.00	.01	.01
		Parcel	.00	.00	.01	.01	.00	.01	.01	.01	.01	.01
	1000	Item	.02	.02	.02	.03	.02	.02	.03	.02	.03	.03
		Parcel	.02	.03	.03	.04	.03	.03	.04	.03	.04	.04
Sk=1.75, K=3.75	100	Item	-.01	-.01	-.01	.00	-.01	-.01	.00	.00	.00	.00
		Parcel	-.02	-.02	-.01	.00	-.02	-.01	.00	-.01	-.01	.00
	300	Item	.00	.01	.01	.01	.01	.01	.01	.01	.01	.02
		Parcel	.00	.00	.01	.02	.01	.01	.02	.01	.01	.02
	1000	Item	.02	.02	.02	.03	.02	.02	.03	.02	.03	.03
		Parcel	.01	.02	.03	.03	.02	.03	.03	.03	.03	.04

Table 33: *Bias of SE of the Direct Effect from F1 to F3 for Each Condition with the First Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	-.02	-.02	-.02	-.01	-.02	-.02	-.01	-.02	-.01	-.01
		Parcel	-.02	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01
	300	Item	-.01	-.01	.00	.00	-.01	.00	.00	.00	.00	.00
		Parcel	-.01	-.01	-.01	.00	-.01	-.01	.00	-.01	.00	.00
	1000	Item	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
		Parcel	.01	.01	.01	.02	.01	.01	.02	.01	.02	.02
Sk=1, K=1.5	100	Item	-.03	-.03	-.03	-.02	-.03	-.03	-.02	-.03	-.02	-.02
		Parcel	-.03	-.03	-.02	-.02	-.03	-.02	-.02	-.03	-.02	-.02
	300	Item	-.04	-.04	-.02	-.02	-.04	-.02	-.02	-.04	-.02	-.02
		Parcel	-.04	-.04	-.02	-.01	-.04	-.02	-.01	-.04	-.02	-.01
	1000	Item	.01	.01	.00	.01	.01	.00	.01	.01	.00	.01
		Parcel	.01	.02	.01	.01	.02	.01	.01	.02	.01	.01
Sk=1.75, K=3.75	100	Item	.02	.02	.03	.03	.02	.03	.03	.03	.03	.03
		Parcel	.02	.02	.02	.03	.02	.02	.03	.02	.03	.03
	300	Item	.03	.03	.03	.04	.03	.03	.04	.03	.03	.04
		Parcel	.03	.04	.04	.05	.04	.04	.05	.04	.04	.05
	1000	Item	.01	.01	.01	.02	.01	.01	.02	.01	.02	.02
		Parcel	.01	.01	.02	.02	.01	.02	.02	.02	.02	.03

Table 34: *Bias of SE of the Direct Effect from F1 to F3 for Each Condition with the Second Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	-.01	-.01	.00	.00	-.01	.00	.00	.00	.00	.00
		Parcel	-.01	-.01	-.01	.00	-.01	.00	.00	-.01	.00	.00
	300	Item	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01
		Parcel	-.01	.00	.00	.01	.00	.00	.01	.00	.00	.01
	1000	Item	.00	.00	.00	.01	.00	.00	.01	.00	.01	.01
		Parcel	.00	.01	.01	.01	.01	.01	.01	.01	.01	.01
Sk=1, K=1.5	100	Item	-.03	-.03	-.03	-.02	-.03	-.03	-.02	-.03	-.02	-.02
		Parcel	-.03	-.03	-.02	-.02	-.03	-.02	-.02	-.02	-.02	-.02
	300	Item	-.03	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
		Parcel	-.03	-.02	-.02	-.01	-.02	-.02	-.01	-.02	-.02	-.01
	1000	Item	.00	.00	.00	.01	.00	.00	.01	.00	.00	.01
		Parcel	.00	.00	.01	.01	.00	.01	.01	.01	.01	.01
Sk=1.75, K=3.75	100	Item	.00	.01	.01	.01	.01	.01	.01	.01	.01	.01
		Parcel	.01	.01	.01	.02	.01	.02	.02	.01	.02	.03
	300	Item	.02	.02	.02	.03	.02	.02	.03	.02	.03	.03
		Parcel	.01	.02	.02	.03	.02	.02	.03	.02	.02	.03
	1000	Item	.01	.01	.01	.01	.01	.01	.02	.01	.01	.02
		Parcel	.01	.01	.02	.02	.01	.02	.02	.02	.02	.03

Table 35: *Bias of SE for the Covariance between Disturbances of F2 with F3 for Each Condition with the First Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
		Parcel	-.03	-.03	-.02	-.02	-.03	-.02	-.02	-.03	-.02	-.02
	300	Item	.02	.02	.02	.02	.02	.02	.02	.02	.02	.03
		Parcel	.02	.02	.02	.03	.02	.02	.03	.02	.02	.03
	1000	Item	-.02	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01
		Parcel	-.02	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01
Sk=1, K=1.5	100	Item	-.01	-.01	-.03	-.03	-.01	-.03	-.03	-.01	-.03	-.03
		Parcel	-.01	-.01	-.03	-.03	-.01	-.03	-.03	-.01	-.03	-.03
	300	Item	-.03	-.02	-.01	.00	-.02	-.01	.00	-.02	.00	.00
		Parcel	-.03	-.03	.00	.00	-.03	.00	.00	-.03	.00	.00
	1000	Item	.00	.00	-.02	-.02	.00	-.02	-.02	.00	-.02	-.02
		Parcel	.00	.00	-.02	-.02	.00	-.02	-.02	.01	-.02	-.02
Sk=1.75, K=3.75	100	Item	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.02
		Parcel	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03
	300	Item	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
		Parcel	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
	1000	Item	-.04	-.04	-.03	-.03	-.04	-.03	-.03	-.04	-.03	-.03
		Parcel	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03

Table 36: *Bias of SE for the Covariance between Disturbances of F2 with F3 for Each Condition with the Second Set of Factor Loadings*

	Sample Size	Level	No Missing	MCAR			MAR			MNAR		
				10%	20%	40%	10%	20%	40%	10%	20%	40%
Sk=0, K=0	100	Item	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	.00
		Parcel	-.02	-.01	-.01	-.01	-.02	-.01	-.01	-.01	-.01	-.01
	300	Item	-.04	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03
		Parcel	-.04	-.04	-.03	-.03	-.04	-.03	-.03	-.04	-.03	-.03
	1000	Item	.01	.01	.01	.01	.01	.01	.01	.01	.01	.02
		Parcel	.01	.02	.02	.02	.02	.02	.02	.02	.02	.02
Sk=1, K=1.5	100	Item	-.04	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03
		Parcel	-.04	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03	-.03
	300	Item	-.01	-.01	-.01	.00	-.01	-.01	.00	-.01	.00	.00
		Parcel	-.01	-.01	.00	.00	-.01	.00	.00	.00	.00	.00
	1000	Item	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
		Parcel	-.03	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
Sk=1.75, K=3.75	100	Item	-.04	-.04	-.04	-.03	-.04	-.04	-.04	-.04	-.04	-.03
		Parcel	-.04	-.04	-.03	-.03	-.04	-.03	-.03	-.03	-.03	-.03
	300	Item	-.01	-.01	-.01	.00	-.01	-.01	.00	-.01	.00	.00
		Parcel	.00	.01	.01	.01	.00	.01	.01	.01	.01	.01
	1000	Item	-.01	-.01	-.01	.00	-.01	-.01	.00	-.01	.00	.00
		Parcel	-.01	-.01	.00	.00	-.01	.00	.00	-.01	.00	.00

CHAPTER 5

DISCUSSIONS AND CONCLUSIONS

Researchers have been using item parceling techniques in their empirical data analysis, particularly when the analysis involves relatively small sample sizes. It has been shown that item parceling helps reduce model complexity, avoid violation of normality assumptions, and obtain better model-data fit, among other benefits (e.g., Little et al., 2002; Yang et al., 2010). It may also help obviate some difficulties in analysis when missing data are present. However, this procedure has not been well studied in the literature. In my dissertation, I investigated how item parceling techniques behave in SEM analysis with missing and nonnormally distributed data. I expected that analysis based on the parcel level would yield results at least as good as those from the item level analysis particularly in the presence of missing data and nonnormal distributions. A Monte Carlo simulation study was conducted using a full structural model (see Figure 2) by manipulating five design factors: (1) sample size, (2) missing data mechanisms, (3) percentage of missingness, (4) degrees of nonnormality of data, and (5) magnitude of factor loadings. In total, 180 conditions were created to generate data. For each generated dataset, a SEM analysis was conducted based on both the item level and the parcel level. Full information maximum likelihood (FIML) estimation methods were applied in all model analyses because this method has been studied extensively in the SEM literature and has been considered the most appropriate for SEM analysis with missing data (e.g., Acock, 2002; Chen et al., 2003; Enders, 2006; Enders at al., 2001). Results from each of the 180 conditions were evaluated based on overall model-data fit and parameter estimates. In this chapter, the key findings are summarized and discussed

in line with the research expectations I listed at the end of the chapter 2. Then, limitations of the study and directions for future research are given.

Major Findings from Simulation Study

I discuss below the major findings based on the five expectations I listed at the end of the chapter 2.

First, I expected that the parcel level analysis would perform better or equally well in terms of parameter estimation bias compared to item level analysis under any type of missing mechanism. In the simulation study, I focused on three parameters: two direct effects (F1→F2 and F1→F3) and the covariance between the disturbances of F2 and F3 (F2↔F3). I computed relative bias of parameter estimates and standard errors for all conditions, and then conducted ANOVAs to examine the impact of the five design factors on parameter estimates and standard errors. In general, the results are consistent with my expectations. The relative bias of the parameter estimates was similar for the parcel level analysis and for the item level analysis when the data were distributed normally or moderately nonnormally. When the data were highly skewed, the parcel level analysis yielded slightly smaller biases than item level analysis for all three parameters estimated in the models. The patterns were similar across missing mechanisms.

The ANOVA results also supported this finding in that the eta square was .03-.07 for the main effect of nonnormality, but the eta square was zero for the main effects of missing mechanisms and level of analysis (see Table 16). Similar results were also found for the standard errors of parameter estimates; missing mechanisms and level of analysis did not impact the accuracy of the SE estimates (eta square values were less than .02). The results were also consistent with the existing literature (e.g., Matsunaga, 2008) showing that the item level analysis with FIML estimation showed biased parameter estimates with nonnormal data, but that

using item parceling mitigates some of the nonnormality effects even for data with missing values. In debating whether parceling techniques should or should not be used, one of the major concerns is that parceling may lead to biased estimates of parameter estimates (Hall et. al., 1999; Stephenson et. al., 2003). Using a factorial algorithm to create parcels, the results from my study showed that parceling did not yield greater bias in parameter estimates and standard errors than item level analysis, and under some conditions, parceling performed better. Because I considered only unidimensional scales for creating parcels, the conclusion may not generalize to multidimensional scales.

Second, I expected the rejection rates of the chi-square test based on the parcel level analysis would be lower than those based on the item level analysis under any type of missing mechanisms. The results support my expectation. As can be summarized from Tables 10 to 11, the rejection rates based on the chi-square test for the parcel based models ranged from 5% to 8% but were much higher for the item level analysis (ranging from 5% to 96%). The largest rejection rates were associated with the item level analysis for small sample size and highly skewed distributions. Analyses based on parceled data showed lower rejection rates because the model became much simpler when a subset of the items was grouped together. For the model based on items, the degrees of freedom are 186, but the degrees of freedom are reduced to 24 when the 15 items associated with F1 are assigned to form 3 parcels.

In addition, evaluation of fit indices (CFI, RMSEA, and SRMR) also led to similar conclusion that models were less likely to be rejected for the parcel level analysis than the item level analysis. The main effects of "level of analysis" were .32, .08, and .81 for CFIs, RMSEAs, and SRMRs, respectively. However, I should note that the analysis models considered in this study are consistent with the data generation model. In other words, the analysis models are

correctly specified, thus the rejection rate based on chi-square test (or fit indices) should be the empirical type I error rate. If the models are misspecified, lower rejection rates based on parceled data indicate lack of power for detecting model misspecification, which would result in another serious issue in SEM analysis.

Although I did not expect it, I found inconsistent results between the means of RMSEAs and the rejection rates based on RMSEA. For example, in Table 17, the mean of RMSEAs was .03 for the item level analysis and .02 for the parcel level analysis, when the sample size was 100 and data were normally distributed with no missing value. Because a higher RMSEA indicates worse model-data fit, I would expect the rejection rate based on RMSEA would be higher for the item level analysis than for the parcel level analysis. However, the rejection rates were reversed with 2% and 15% for item level analysis and parcel level analysis, respectively.

To further understand the performance of RMSEA, I examined the sampling distribution of the RMSEAs for these two conditions. The distributions of RMSEA were quite different between these two conditions. The majority of replications resulted in RMSEA values smaller than .04 in parcel level analysis but a small proportion of replications resulted in large RMSEAs. However, under item level analysis most of the RMSEA values were between .04 and .06. When using .06 as the cutoff, parcel level analysis yielded a higher rejection rate than item level analysis.

In addition, I found that RMSEA was relatively stable across levels of analysis compared to CFI and SRMR. The eta square for the main effect of level of analysis was .08 for RMSEAs but much higher for CFIs and SRMRs (.32 and .81). The difference between parcel level analysis and item level analysis in the model rejection rate based on RMSEAs was smaller than those based on CFIs and SRMRs. This indicates that RMSEA and CFI/SRMR may perform differently

in the evaluation of model-data fit. We want fit indices to help us make the correct decision about the model specification (whether correctly specified or misspecified), no matter whether analysis is conducted based on the parcel level or item level. From this point of view, RMSEA may be more appropriate than CFI and SRMR when using parceled data. But further research is needed to understand the performance of RMSEAs in comparison to CFIs and SRMRs for parceled data, particularly when the model is misspecified.

Third, I expected that the percentage of missingness on variables would have less effect on the results (both overall model-data fit and parameter estimates) for the parcel level analysis compared to the item level analysis under all missing mechanisms. Part of my research expectation is supported by the simulation study. When analysis was based on the item level, as the percentage of missingness increased, models demonstrated worse fit according to the chi-square test, CFI, RMSEA, and SRMR. Compared to the increase in percentage of missingness from 10% to 20%, the model-data fit was substantially worse when the percentage increased from 20% to 40%. However, the performance of the chi-square test, CFI, RMSEA, and SRMR was very similar across different percentage of missingness in the data for parcel level analysis. This finding provides evidence favoring parcel analysis when missing data are present in an empirical study.

Unexpectedly, this finding was not observed when examining parameter estimates and standard errors. The bias of parameter estimates and standard errors was similar across different percentages of missingness and was also similar to that from analysis based on complete data. The associated eta squares were all zero for the parameter estimates and standard errors, indicating that the estimates of the parameters (e.g., the direct effect from F1 to F2) did not vary across different percentages of missingness. This might be due to the percentages of missingness

considered in the study. The highest percentage of missingness considered in the study was 40%, but only six out of 21 variables had missing values on them. This constitutes a small proportion of missingness in the data as a whole. Further study may include higher proportions of missingness on larger numbers of variables.

Four, I expected that the parcel level analysis would encounter fewer problems in model convergence than the item level, particularly for nonnormal data with missing values. As evidenced from Tables 8 and 9, this expectation is not supported. The model non-convergence rate was either zero or nearly zero for all conditions, regardless of sample size, percentage of missingness, the degree of nonnormality, and missing mechanism. This may due to the level of conditions for each design factor considered in the simulation study. Models may encounter convergence problems if the sample size is small, higher percentages of data are missing, and if the distribution deviates greatly from the normality. Further research may consider including these more extreme conditions in the study. Findings from the current study suggest that model non-convergence is not a concern in the choice of parcel vs. item level analysis as long as the data do not seriously deviate from normality and the percentage of missingness is not too large (considering only 6 out of 21 items have missing data). Information about the normality of the data and the percentage of missingness can be fairly easily obtained after data are collected.

Fifth, I expected that as the degree of data nonnormality increase, the superiority of parcel level analysis to the item level analysis would be more obvious. The expectation is supported in terms of model-data fit. For the item level analysis, the model-data was fit better as the data approached to normality, holding other conditions constant. However for the parcel level analysis, model-data fit was equally well regardless of the degree of nonnormality. This is also evidenced by the ANOVAs where the eta square of the interaction effect between level of

analysis and the degree of nonnormality was not zero for CFIs (.06), RMSEAs (.02), and SRMR (.03). This finding is consistent with arguments made by several authors (Bandalos, 2002; Little et al., 2002; Matsunaga, 2008). That is, parcel scores tend to be more normally distributed compared to individual item scores, thus should be impacted less by the degree of nonnormality of the individual items.

Last, I expected that as the sample size increased, the superiority of the parcel level analyses to the item level analysis would be reduced compared to item level analysis under any type of missing mechanism. This expectation is supported in terms of overall model-data fit. Parcel level analysis resulted in better model-data fit than item level analysis (as summarized previously). As sample size increased, the chi-square test, CFI, RMSEA, and SRMR suggested better fit for both parcel and item level analyses, holding other conditions constant. However, the superiority of parcel level analysis to item level analysis became less obvious as sample size increased (See Tables 10, 14, and 23). In other words, the impact of sample size in model-data fit was greater for item level analysis than for parcel level analysis. This is also evidenced by the ANOVAs where the interaction effect between sample size and level of analysis was not trivial for CFIs, RMSEAs and SRMRs (the eta squares were .36, .08, and .54 for CFIs, RMSEAs, and SRMRs, respectively). When the analysis was conducted based on the item level, the number of model parameters was 45, resulting in ratios of sample size to the number of model parameters of 2.2, 6.7, and 22.2 for sample size of 100, 300, and 1000, respectively. Model fit increased substantially as the ratio increased from small to medium. But little gain was obtained by increasing the sample size up to a certain level. This is also the reason that the increase in fit is less obvious in parcel level analysis.

Limitations and Future Research

Based on the current simulation study, parcel level analysis yielded SEM results as good as or better than those from item level analysis under three types of missing mechanisms, across different degrees of nonnormality of data, and varying percentages of missingness. Similar to other simulation studies, one should be cautious when generalizing these conclusions to other situations. Below I list several major limitations in my study, some of which will direct my future research.

First, all the generated data are continuous. However, in practice, data are often categorical. The results may be different for analysis of categorical variables. Full information maximum likelihood may not be appropriate for categorical variables. Future research can be conducted to consider categorical incomplete data with alternative estimation methods.

Second, the analysis model was consistent with the data generation model. Misspecified models are not considered in this study. As I discussed previously, the performance of fit indices may be different and the superiority of parceling might not hold in detecting model misspecification in such cases. In the future, models with misspecification can be considered to investigate the performance of item parceling with missing data.

Finally, only a limited number of levels were considered for each design factor, most particularly only one model was used to generate data. Based on the findings from the current study, it may be worth considering increasing the number of variables with missing values and/or the percentage of missingness. The ratio of the number of variables with missing values to the total number of variables was only .29 (= 6/21) in this study. The largest percentage of missingness was 40%. Therefore, at maximum, only 11% (= 29% * 40%) of the data were missing. Some other simulation studies have considered much higher percentage of missingness. For example, Davey et al. (2005) and Allison (2003) included conditions with 95% and 90%

missingness, respectively. Although having 95% or 90% of missing data is unlikely to encounter in an empirical study, including conditions with percentages of missingness higher than what I had in the simulation study is worth considering.

APPENDIX A

R-CODES FOR DATA GENERATION

Normally distributed data were generated in R. Then, normally distributed data transformed nonnormal data by Fleishman's method.

```
rm(list=ls())

n.d<-2000                                                    #Number of data set
list_normal<-rep(0,n.d)
list_non_normal<-rep(0,n.d)
for(dd in 1:n.d){
ss<-100                                                      #Sample Size

library(Matrix)
library(psych)

C<-matrix(c(1,.4, .6, .4, 1,.74, .6, .74,1), ncol=3)
      #Factor correlations
U<-chol(C)                                                   #Cholesky decomposition

ran.f1<-rnorm(ss,0,1)                                        #random factor scores
ran.f2<-rnorm(ss,0,1)
ran.f3<-rnorm(ss,0,1)
R<-as.matrix(cbind(ran.f1,ran.f2,ran.f3))                    #Random un-correlated
                                                             Factor scores for all
                                                             factors
corr.f<-R%*%U                                                #Random Correlated
                                                             factor scores for all
                                                             factors

rand.err.1<-rnorm(ss,0,1)
rand.err.2<-rnorm(ss,0,1)
rand.err.3<-rnorm(ss,0,1)
rand.err.4<-rnorm(ss,0,1)

n.f1<-15                                                     #For F1 in the model,
                                                             15 items
l.f1<-c(rep(c(.4,.6,.8),5))                                  #Loadings of the first
                                                             factor
l.f1[1]<-l.f1[4]<-l.f1[13]<-.15
x1<-matrix(0, ncol=n.f1, nrow=ss)

for(i in c(1,3)){
e1<-rnorm(ss,0,1)
for(j in 1:ss){
x1[j,i]<- l.f1[i] * corr.f[j,1] + cond[1]*rand.err.1[j]+(sqrt(1- l.f1[i]^2-
cond[1]^2))*e1[j]
}}
for(i in c(2,4)){
e1<-rnorm(ss,0,1)
for(j in 1:ss){
```

```
x1[j,i]<- l.f1[i] * corr.f[j,1] + cond[2]*rand.err.2[j]+(sqrt(1- l.f1[i]^2-
cond[2]^2))*e1[j]
}}
for(i in c(7,8)){
e1<-rnorm(ss,0,1)
for(j in 1:ss){
x1[j,i]<- l.f1[i] * corr.f[j,1] + cond[3]*rand.err.3[j]+(sqrt(1- l.f1[i]^2-
cond[3]^2))*e1[j]
}}
for(i in c(9,10)){
e1<-rnorm(ss,0,1)
for(j in 1:ss){
x1[j,i]<- l.f1[i] * corr.f[j,1] + cond[4]*rand.err.4[j]+(sqrt(1- l.f1[i]^2-
cond[4]^2))*e1[j]
}}

for(i in c(5,6,11,12,13,14,15)){
e1<-rnorm(ss,0,1)
for(j in 1:ss){
x1[j,i]<- l.f1[i] * corr.f[j,1] + (sqrt(1- l.f1[i]^2))*e1[j]
}}

n.f2<-3                                                    #For F2, 3 items
l.f2<-c(rep(.7, n.f2))
x2<-matrix(0, ncol=n.f2, nrow=ss)
for(i in 1:n.f2){
e2<-rnorm(ss,0,1)
for(j in 1:ss){
x2[j,i]<- l.f2[i] * corr.f[j,2] + (sqrt(1- l.f2[i]^2))*e2[j]
}}

n.f3<-3                                                    #For F3, 3 items
l.f3<-c(rep(.7, n.f3))
x3<-matrix(0, ncol=n.f3, nrow=ss)
for(i in 1:n.f3){
e3<-rnorm(ss,0,1)
for(j in 1:ss){
x3[j,i]<- l.f3[i] * corr.f[j,3] + (sqrt(1- l.f3[i]^2))*e3[j]
}}

final.data.normal<-data.frame(cbind(x1,x2,x3))             #All 21 items
                                                           together
                                                           #These are still
                                                           normal.
name1<-paste("Raw Data/Normal.", dd, ".dat", sep="")
name1_list<-paste("Normal.", dd, ".dat", sep="")
list_normal[dd]<-name1_list


#
#     Nonnormality (Fleishman's Table 1)
#
#           non_x1<-a+b*x1+c*x1^2+d*x1^3  The formula for nonnormality
#
#           sk=0
#           k=0
```

```
#     b=1
#     c=0
#     d=0
#     a=-c
#
#           sk=1.00
#            k=1.5
#
#     b<- .9530769
#     c<- .1631943
#     d<- .0065974
#     a<- -c
#
            sk=1.75
            k=3.75

      b=.9296605
      c=.3994967
      d=-.036467
      a=-c
#
```

```
final.data<-final.data.normal
for(i in 1:5){                                          #Nonnormality added on
                                                        items 1-5.
final.data[,i]<-a+b*final.data[,i]+c*final.data[,i]^2+d*final.data[,i]^3
}
      for(i in 6:10){                                   #Nonnormality added on
                                                        items7-10.
final.data[,i]<-a+b*final.data[,i]+c*final.data[,i]^2+d*final.data[,i]^3
}
for(i in 11:15){                                        #Nonnormality added on
                                                        items11-15.
final.data[,i]<-a+b*final.data[,i]+c*final.data[,i]^2+d*final.data[,i]^3
}
name_file<-paste("sk=", sk, ", k=", k, "ss=", ss, sep="")
dir.create(name_file)
name2<-paste(name_file,"/Non_Normal.", dd, ".dat", sep="")
name2_list<-paste("Non_Normal.", dd, ".dat", sep="")

write.table(final.data, name2, sep="\t", quote=FALSE, col.names=FALSE,
row.names=FALSE)
list_non_normal[dd]<-name2_list
}

write.table(list_non_normal, paste(name_file,"/List_Non_Normal", ".dat",
sep=""),quote=FALSE, col.names=FALSE, row.names=FALSE)
```

APPENDIX B

CORRELATION MATRIX AMONG THE FACTORS

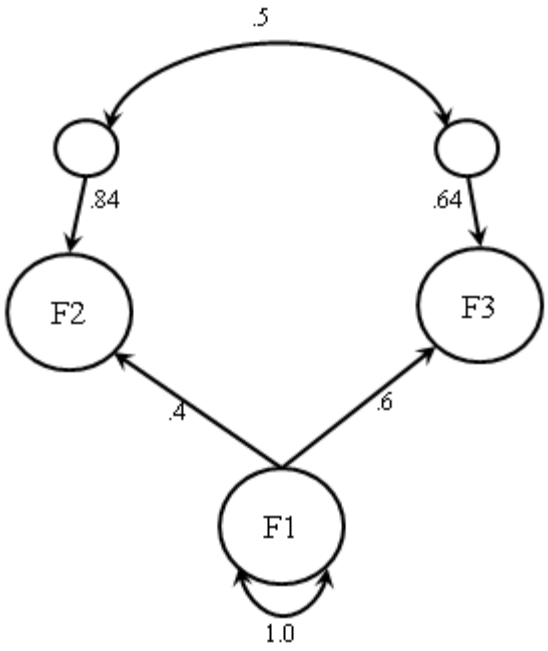

Figure 15: Correlations among the Factors

This appendix shows the calculation of the covariance matrix among the factors for the research model, which is shown in figure above.

Formulas:

1) $var(x) = E(x - E(x))^2$
2) $cov(x, y) = E(x - E(x))\, E(y - E(y))$

Given Information:

1) $var(F1) = 1$
2) $var\ (e2) = 1 - .4^2 = .84$
3) $var\ (e3) = 1 - .6^2 = .64$
4) $cov(e2, e3) = .5$
5) $cov(F, e) = 0$
6) $F2 = .4F1 + e2$

7) $F3 = .6F1 + e3$

Covariance between factor 1 (F1) and factor 2 (F2):

$$cov(F1, F2) = E(F1 - E(F1)) \cdot E(F2 - E(F2) \quad (\#1)$$

$$E\big(F2 - E(F2)\big) = E\big(.4F1 + e2 - .4F(F1) - E(e2)\big)$$

$$= E\left(.4\big(F1 - E(F1)\big) + e2 - E(e2)\right) = .4 \cdot E\big(F1 - E(F1)\big) + E(e2 - E(e2)) \quad (\#2)$$

Bases on (#1) and (#2);

$$cov(F1, F2) = E\big(F1 - E(F1)\big) \cdot [.4 \cdot E\big(F1 - E(F1)\big) + E(e2 - E(e2))]$$

$$= .4\, E\big(F1 - E(F1)\big)^2 + E(F1 - E(F1)) \cdot E(e2 - E(e2))$$

$$= .4 \cdot var(F1) + cov(F1, e2) = .4 \cdot 1 + 0 = .4$$

Thus, $cov(F1, F2) = .4$

Fallowing the same logic, it is easy to say that $cov(F1, F3) = .6$.

Covariance between factor 2 (F2) and factor 3 (F3):

$$cov(F2, F3) = E\big(F2 - E(F2)\big) \cdot E\big(F3 - E(F3)\big) \quad (\#3)$$

$$E\big(F2 - E(F2)\big) = E\big(.4F1 + e2 - .4F(F1) - E(e2)\big) \quad (\#4)$$

$$E\big(F3 - E(F3)\big) = E\big(.6F1 + e2 - .6F(F1) - E(e2)\big) \quad (\#5)$$

Bases on (#3), (#4) and (#5);

$$cov(F2, F3) = \big[.4 \cdot E\big(F1 - E(F1)\big) + E\big(e2 - E(e2)\big)\big] * \big[.6 \cdot E\big(F1 - E(F1)\big) + E\big(e3 - E(e3)\big)\big]$$

$$= .24\, E\big(F1 - E(F1)\big)^2 + E\big(F1 - E(F1)\big) \cdot E\big(e2 - E(e2)\big) + E\big(F1 - E(F1)\big) \cdot E\big(e3 - E(e3)\big) + E\big(e2 + E(e2)\big)E\big(e3 + E(e3)\big)$$

$$= .24\, var(F1) + cov(F1, e2) + cov(F1, e3) + cov(e2, e3)$$

$$= .24var(F1) + 0 + 0 + cov(e2, e3) = .24 + .5 = .74$$

Thus, $cov(F2, F3) = .74$

APPENDIX C

CREATING PARCELS IN R

```
#############################################################################
#Creating Parcels:
#                      parcel1 (mean of 1, 6, 7, 12 and 13)
#                      parcel2 (mean of 2, 5, 8, 11, and 14)
#                      parcel3 (mean of 3, 4, 9, 10, and 15)
#############################################################################

for( p in 1:ss){
parcel1[p] <-mean(c(data.miss[p,n.fl[1]],
data.miss[p,n.fl[6]],data.miss[p,n.fl[7]],data.miss[p,n.fl[12]],
                        data.miss[p,n.fl[13]]),na.rm=TRUE)
parcel2[p] <-mean(c(data.miss[p,n.fl[2]],
data.miss[p,n.fl[5]],data.miss[p,n.fl[8]],data.miss[p,n.fl[11]],
                        data.miss[p,n.fl[14]]),na.rm=TRUE)
parcel3[p] <-mean(c(data.miss[p,n.fl[3]],
data.miss[p,n.fl[4]],data.miss[p,n.fl[9]],data.miss[p,n.fl[10]],
                        data.miss[p,n.fl[15]]),na.rm=TRUE)
}
data.parcel<-round(cbind(parcel1,parcel2, parcel3, data.cont[,16:21]), 4)

#############################################################################
```

APPENDIX D

BIAS OF PARAMETER ESTIMATES AND SE IN R

```
##########################          Bias    ################################

bias.1<-matrix(c(((data.1[,"L12"]-.4)/.4),((data.1[,"L13"]-
.6)/.6),((data.1[,"L23"]-.5)/.5)), ncol=3,byrow=FALSE)
       colnames(bias.1)<-c("bias_12", "bias_13", "bias_23")
bias.2<-matrix(c(((data.2[,"L12"]-.4)/.4),((data.2[,"L13"]-
.6)/.6),((data.2[,"L23"]-.5)/.5)), ncol=3,byrow=FALSE)
       colnames(bias.2)<-c("bias_12", "bias_13", "bias_23")
bias.3<-matrix(c(((data.3[,"L12"]-.4)/.4),((data.3[,"L13"]-
.6)/.6),((data.3[,"L23"]-.5)/.5)), ncol=3,byrow=FALSE)
       colnames(bias.3)<-c("bias_12", "bias_13", "bias_23")
bias.4<-matrix(c(((data.4[,"L12"]-.4)/.4),((data.4[,"L13"]-
.6)/.6),((data.4[,"L23"]-.5)/.5)), ncol=3,byrow=FALSE)
       colnames(bias.4)<-c("bias_12", "bias_13", "bias_23")
bias.5<-matrix(c(((data.5[,"L12"]-.4)/.4),((data.5[,"L13"]-
.6)/.6),((data.5[,"L23"]-.5)/.5)), ncol=3,byrow=FALSE)
       colnames(bias.5)<-c("bias_12", "bias_13", "bias_23")
bias.6<-matrix(c(((data.6[,"L12"]-.4)/.4),((data.6[,"L13"]-
.6)/.6),((data.6[,"L23"]-.5)/.5)), ncol=3,byrow=FALSE)
       colnames(bias.6)<-c("bias_12", "bias_13", "bias_23")
bias.7<-matrix(c(((data.7[,"L12"]-.4)/.4),((data.7[,"L13"]-
.6)/.6),((data.7[,"L23"]-.5)/.5)), ncol=3,byrow=FALSE)
       colnames(bias.7)<-c("bias_12", "bias_13", "bias_23")
bias.8<-matrix(c(((data.8[,"L12"]-.4)/.4),((data.8[,"L13"]-
.6)/.6),((data.8[,"L23"]-.5)/.5)), ncol=3,byrow=FALSE)
       colnames(bias.8)<-c("bias_12", "bias_13", "bias_23")


############################          SD      ################################

sd1<-c(sd(data.1[,"L12"]), sd(data.1[,"L13"]), sd(data.1[,"L23"]))
sd5<-c(sd(data.5[,"L12"]), sd(data.5[,"L13"]), sd(data.5[,"L23"]))


##########################          Bias_SE         ######################

b.se.1<-matrix(c(((data.1[,"S_L12"]-sd1[1])/sd1[1]),((data.1[,"S_L13"]-
sd1[2])/sd1[2]),((data.1[,"S_L23"]-sd1[3])/sd1[3])), ncol=3,byrow=FALSE)
b.se.2<-matrix(c(((data.2[,"S_L12"]-sd1[1])/sd1[1]),((data.2[,"S_L13"]-
sd1[2])/sd1[2]),((data.2[,"S_L23"]-sd1[3])/sd1[3])), ncol=3,byrow=FALSE)
b.se.3<-matrix(c(((data.3[,"S_L12"]-sd1[1])/sd1[1]),((data.3[,"S_L13"]-
sd1[2])/sd1[2]),((data.3[,"S_L23"]-sd1[3])/sd1[3])), ncol=3,byrow=FALSE)
b.se.4<-matrix(c(((data.4[,"S_L12"]-sd1[1])/sd1[1]),((data.4[,"S_L13"]-
sd1[2])/sd1[2]),((data.4[,"S_L23"]-sd1[3])/sd1[3])), ncol=3,byrow=FALSE)
b.se.5<-matrix(c(((data.5[,"S_L12"]-sd5[1])/sd5[1]),((data.5[,"S_L13"]-
sd5[2])/sd5[2]),((data.5[,"S_L23"]-sd5[3])/sd5[3])), ncol=3,byrow=FALSE)
b.se.6<-matrix(c(((data.6[,"S_L12"]-sd5[1])/sd5[1]),((data.6[,"S_L13"]-
sd5[2])/sd5[2]),((data.6[,"S_L23"]-sd5[3])/sd5[3])), ncol=3,byrow=FALSE)
```

```
b.se.7<-matrix(c(((data.7[,"S_L12"]-sd5[1])/sd5[1]),((data.7[,"S_L13"]-
sd5[2])/sd5[2]),((data.7[,"S_L23"]-sd5[3])/sd5[3])), ncol=3,byrow=FALSE)
b.se.8<-matrix(c(((data.8[,"S_L12"]-sd5[1])/sd5[1]),((data.8[,"S_L13"]-
sd5[2])/sd5[2]),((data.8[,"S_L23"]-sd5[3])/sd5[3])), ncol=3,byrow=FALSE)


        colnames(b.se.1)<-colnames(b.se.2)<-colnames(b.se.3)<-
colnames(b.se.4)<-c("B_SE_12", "B_SE_13", "B_SE_23")
        colnames(b.se.5)<-colnames(b.se.6)<-colnames(b.se.7)<-
colnames(b.se.8)<-c("B_SE_12", "B_SE_13", "B_SE_23")

##############################################################################
```

REFERENCES

Achenbach, T. M., Bernstein, A. & Dumenci, L. (2005). DSM-oriented scales and statistically based syndromes for ages 18 to 59: Linking taxonomic paradigms to facilitate multitaxonomic approaches. *Journal of personality assessment*, *84*, 49-63.

Acock, A. C. (2005). Working with missing values. *Journal of Marriage and Family*, *67*, 1012-1028.

Alhija, F. N., & Wisenbaker, J. (2006). A Monte Carlo study investigating the impact of item parceling strategies on parameter estimates and their standard errors in CFA. *Structural Equation Modeling*, 13, 2-4-228.

Allison, P. D. (2002). *Missing data.* California: Sage Publications.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, *112*, 545-557.

Andreassen, T. W., Lorentzen, B. G., & Olsson, U. H. (2006). The impact of non-normality and estimation methods in SEM on satisfaction research in marketing. *Quality & Quantity, 40(1)*, 39–58.

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

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

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

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

Bentler, P.M. (1990), Comparative Fit Indexes in Structural Models. *Psychological Bulletin*, *107*, 238-246.

Bentler, P.M. (1995). EQS structural equations program manual. Encino, CA: Multivariate Software.

Bernstein, I. H., & Teng, G. (1989). Factoring items and factoring scales are different : Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin*, *105*, 467-477.

Carter, R. L. (2006). Solutions for missing data in structural equation modeling. *Research & Practice in Assessment*, *1*, 1-6.

Chen, G., & Astebro, T. (2003). How to deal with missing categorical data : Test of a simple Bayesian method. *Organizational Research Methods*, *6*, 309-327.

Davey, A., & Savla, J. (2005). Issues in evaluating model fit with missing data. *Structural Equation Modeling*, *12*(4), 578-597.

Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, *6*, 352-370.

Enders, C. K. (2004). The impact of missing data on sample reliability estimates: Implications for reliability reporting practices. *Educational and Psychological Measurement, 64*(3), 419-436.

Enders, C., K. (2006). Analyzing structural equation models with missing data. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (p. 313 - 342). USA: Information Age Publishing.

Enders, C. K. (2010). *Applied missing data analysis*. NY: The Guilford Press.

Enders, C. K. (2011). Missing Not at Random Models for Latent Growth Curve Analyses. *Psychological Methods, 16*, 1-16.

Enders, C. K. & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 3*, 430-457.

Fakunmoju, S., Woodruff, K., Kim, H. H., LeFevre, A., & Hong, M. (2010). Intention to leave a job: The role of individual factors, job tension, and supervisory support. *Administration in Social Work*, *34*, 313-328.

Finney, S., J & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation eodeling: A second course* (p. 269 - 314). USA: Information Age Publishing.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*, 521-532.

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*, 233-256.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Hoogland, J. J. & Boomsma, A. (1998). Robustness Studies in Covariance Structure Modeling: An Overview and a Meta-Analysis. *Sociological Methods & Research, 26*, 329-367.

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

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3th ed.). NY: The Guilford Press.

Landis, R. S., Beal, D. J., & Tesluk, P. E. (2002). A comparison of approaches to forming composite measures in structural equation models. *Organizational Behavioral Research, 3*, 186-207.

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

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

Matsunaga, M. (2008). Item parceling in structural equation modeling: A primer. *Communication Methods and Measures*, *2*, 260-293.

Meade, A. W. & Kroustalis, C. M. (2005, April). *Problems with Item parceling for confirmatory factor analysis tests of measurement invariance of factor loadings.* Paper presented at the 20th Annual Conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction.* NY: The Guilford Press.

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

Muthén, B. & Muthén, L. (1998-2008). *Mplus user's guide. Sixth edition*. Los Angeles, CA: Muthén & Muthén.

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

Nasser, F., & Wisenbaker, J. (2003). A Monte Carlo study investigating the impact of item parceling on measures of fit in confirmatory factor analysis. *Educational and Psychological Measurement, 63(5),* 729–757.

Patrician, P. A. (2002). Focus on research methods multiple imputation for missing data. *Research in Nursing & Health*, *25*, 76-84.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74*, 525-556.

Plummer, B. A. (2000). *To parcel or Not to Parcel: The Effects of Item Parceling in Confirmatory Factor Analysis*. Unpublished doctoral dissertation, University of Rhode Island (UMI No. 9989451)

Rogers, W. M., & Schmitt, N. (2004). Parameter Recovery and Model Fit Using Multidimensional Composites: A Comparison of Four Empirical Parceling Algorithms. *Multivariate Behavioral Research, 39*, 379-412

Rubin, D. B. (1976). Inference and missing data. *Biometrica*, *63*, 581-592.

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

Schafer, J. L., & Graham, J. W. (2002). Missing data : Our view of the state of the art. *Psychological Methods*, *7*, 147-177.

Signorella, M. L., & Cooper, J. E. (2011). Relationship suggestions from self-help books: Gender stereotyping, preferences, and context effects. *Sex Roles*, *65*, 371-382.

Stephenson, M. T., & Holbert, R. L. (2003). A Monte Carlo simulation of observable versus latent variable structural equation modeling techniques. *Communication Research*, *30*, 332-354.

Tucker, L. R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. Psychometrika, *38*, 1-10

Veldstra, J. L., Brookhuis, K. A., de Waard, D., Molmans, B. H. W., Verstraete, A. G., Skopp, G., & Jantos, R. (2011). Effects of alcohol (BAC 0.5‰) and ecstasy (MDMA 100 mg) on simulated driving performance and traffic safety. *Psychopharmacology, NA*.

Yang, C., Nay, S., & Hoyle, R. H. (2010). Three approaches to using lengthy ordinal scales in structural equation models: Parceling, latent scoring, and shortening scales. *Applied Psychological Measurement*, *43*, 122-142.

Yoder, J. D., Snell, A. F., & Tobias, A. (2012). Balancing multicultural competence with social justice: Feminist beliefs and optimal psychological functioning. *The Counseling Psychologist*, *NA*, 1-32.

BIOGRAPHICAL SKETCH

Fatih Orcan was born in Kahramanmaras, Turkey. Fatih received his BA and MA degrees in Mathematics Education from Inonu University, Malatya, Turkey in 2006. He worked as a Mathematics teacher for one semester in a private sector high school in 2006. In 2008, he started the Measurement and Statistics program in the College of Education at Florida State University. He earned an MS degree in 2010, and continued to work on his PhD in the same program. During his stay at Florida State University, he worked as a teaching assistant for numerous courses and tutored mathematics to college students at Florida State University and other institutions. Currently, he is working with the Bureau of Postsecondary Assessment at Florida Department of Education. In his dissertation, he focused on the use of item parceling techniques in structural equation modeling when data present missing values and nonnormally distributed.