ISSN: 1309 – 6575
**Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi**
Journal of Measurement and Evaluation in Education and Psychology
2016; 7(1);59-72

# A Note on the Use of Item Parceling in Structural Equation Modeling with Missing Data*

# Yapısal Eşitlik Modellemesinde Kayıp Verilerin Madde Parsellemenin Kullanımı Üzerine Bir Not

Fatih Orçan**     Yanyun Yang***

**Abstract**
Item parceling procedure may be applied to alleviate some difficulties in analysis with missing data and/or nonnormal data in structural equation modeling. A simulation study was conducted to investigate how item parceling behaves under various conditions in structural equation model with missing and nonnormal distributed data. Design factors included missing mechanism, percentage of missingness, distribution of item data, and sample size. Results showed that analysis conducted at the parcel level yielded lower model rejection rates than analysis based on the individual items, and the patterns were consistent across missing mechanism, percentage of missing, and distribution of item data. In addition, parcel-level analyses resulted in comparable parameter estimates to item-level analyses.

*Keywords:* Item parceling, missing data, nonnormal distribution, structural equation model, SEM

**Öz**
Yapısal eşitlik modellerinde madde parselleme prosedürü kayıp very ve/veya normal dağılım olmadığı durumlarda karşılaşılan zorlukları aztmakta kullanılabilir. Madde parsellemenin kayıp ve normal dağılmayan verilerin olması durumunda yapısal eşitlik modellinde nasıl davranacağı bir simülasyon çalışması ile araştırılmıştır. Dizayn faktörleri kayıp mekanizması, kayıp veri yüzdesi, madde veri dağılımları ve örneklem büyüklüğünü içermektedir. Sonçlara göre, parsel seviyesinde yapılan analizler madde seviyesinde yapılan analizlere oranla daha az model reddedilmesine sebep olmuş ve kayıp mekanizması, kayıp veri yüzdesi ve madde veri dağılımlarında benzer davranmıştır. Ayrıca, parsel seviyesindeki analizler madde seviyesindeki analizlere karşılaştırılabilir parametre tahminleriyle sonuçlanmıştır.

*Anahtar Kelimeler:* Madde parselleme, kayıp veriler, normal olmayan dağılım, yapısal eşitlik modellemesi, YEM

## INTRODUCTION AND LITERATURE REVIEW

Structural equation modeling (SEM) has been frequently used in empirical data analysis to examine hypothesized relationships among a set of variables. A commonly used estimation method in SEM, maximum likelihood (ML), requires the sample size be sufficiently large and observed variables be multivariate normally distributed. Violation of these assumptions results in inaccurate model chi-square statistic, fit indices, parameter estimates, and standard errors associated with parameter estimates (e.g., Bollen, 1989; Curran, West, & Finch, 1996; Chou, Bentler, &Satorra, 1991), and the degree of bias tends to increase as the model complexity increases. Alternatively, estimation methods that do not require assumptions as restrictive as ML may be applied. One alternative is robust maximum likelihood (MLR) estimation method. MLR corrects for the positive bias in model chi-square statistic and the negative bias of standard errors associated with parameter estimates. Another alternative is weighted least square (WLS). WLS does not require variables be multivariate normally distributed. However, WLS requires very large sample sizes even when the data are

* This study was conducted as part the first author's dissertation

** Asst. Prof., Karadeniz Technical University, College of Education, Trabzon – Turkey, fatihorcan@ktu.edu.tr
*** Assoc. Prof. Florida State University, College of Education, Tallahassee – FL, yyang3@admin.fsu.edu


*Geliş Tarihi: 18.11.2015*
DOI: 10.21031/epod.88204
*Kabul Tarihi: 09.03.2016*

multivariate normally distributed (Curran, West, & Finch, 1996; Hu, Bentler, & Kano, 1992) and the performance of WLS becomes worse as the complexity of the model increases (Muthén& Kaplan, 1992). When the purpose of the SEM analysis is not to examine the psychometric properties of individual items, instead, it is to investigate the relationship among latent factors, a parceling technique may be applied to reduce the model complexity, particularly when the analysis involves small sample size and the number of indicators per factor is large (e.g., Sterba&MacCallum, 2010).

Parceling is referred to as a procedure for computing sums or average scores across multiple items. The variables based on the sum or average (called parcels) instead of the individual items are then used as indicators of latent factors in the SEM analysis (Bandalos 2002, 2008; Little, Cunningham, Shahar, &Widaman, 2002; Sass & Smith, 2006; Sterba, 2011; Sterba&MacCallum, 2010; Yang, Nay, & Hoyle, 2010). Bandalos and Finney (2001) reviewed 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 kinds of item parceling techniques; and the percentage varied from 9% in Journal of Marketing Research to 60% in Journal of Educational Measurement. A much higher percentage was found in the review of three psychology journals published in 1996-1999 (Plummer, 2000); among 102 articles used structural equation modeling, about 50% of which involved some kinds of parceling techniques in the analysis. Use of parcels is appealing in that it reduces model complexity, reduces the requirements on sample sizes, reduces influences of individual items' systematic errors on the model estimation, helps reach optimal reliability, increases model convergence rate, and increases the model fit when the dimensionality of the items is known (Bandalos, 2002; Little et al, 2002; Matsunaga, 2008; Meade &Kroustalis, 2005; Nasser & Takahashi, 2003; Nasser &Wisenbaker, 2003; Plummer, 2000; Sass & Smith, 2006; Yang et al., 2010). Sass and Smith (2006) showed analytically that use of parcels does not lead to a bias estimate of the structural relationship among latent factors when model assumptions are met and items are unidimensional. On the other hand, there are some arguments against the use of parceling techniques. For example, using parcels in the analysis may blur the dimensionality of original measures and produce biased estimates of model parameters (Bandalos, 2002; Matsunaga, 2008). It has been recommended that parceling be used thoughtfully so that the drawback of such a use is minimized. Parceling is most beneficial when the analysis is conducted based on a small sample size and the relationship between items and the underlying latent factor is not strong (Sterba&MacCallum, 2010), on the other hand, when the dimensionality of the items is not clear, parceling should not be used (Bandalos& Finney, 2001; Little et al., 2002; Meade &Kroustalis, 2006).

For unidimensional measures, items can be assigned to parcels either randomly or purposively (Little et al., 2002; Matsunaga, 2008; Sterba&MacCallum, 2010). In this study, we use the following procedure to allocate items to parcels although results from current simulation study may be also valid for other item allocation methods (Sterba&MaCallum, 2010): First, a factor analysis is to conduct on the items to be parceled. Second, the obtained factor loadings are used to allocate items such that the sums of loadings are as equivalent as possible across parcels. Third, an SEM analysis is conducted based on parcel scores. From the classical test score theory perspective, the approach is preferred because parcels tend to be essentially tau-equivalent and thus maximize reliability of the scale based on the parcels, consequently, lead to the least bias estimates of structural coefficients among latent factors. This approach has been adopted in previous studies, however, in these studies the parcels are created based on the magnitude of loadings in population. In other words, the assignment of items to parcels is the same across all samples. Sterba and MacCallum (2010) argued that assigning the same items to parcels across all datasets do not take sampling error into consideration because the estimates of loadings may vary across samples. In our study, we conducted factor analysis on the items to be parceled for each sample, sorted items based on the magnitude of the loadings, and then assigned items to parcels. Consequently, the assignment of items to parcels varies from sample to sample while the sums of loadings are as equivalent as possible across parcels for each sample.

When the distributional assumptions underlying ML are violated, an applied researcher might face the decision of analyzing the model based on individual items using alternative estimation methods (e.g., MLR, WLS) or creating parcels and then analyzing the model based on the parcel scores. If item-level analysis using alternative estimation methods yields no worse results than those from parcel-level analysis, then adopting a parceling technique becomes an unnecessary work, not mentioning that there have been many debates regarding the use of parceling (Little, et al., 2002). As we are aware of, no previous study has compared model results (e.g., model rejection rates based on chi-square statistic, accuracy of estimated structural coefficients among latent factors) between item-level analysis and parcel-level analysis using alternative estimation methods when the distributional assumptions underlying ML are violated. This is one of the purposes of current study.

The presence of missing data creates a potential problem in SEM analysis. Although some other types of missing mechanisms have been discussed in the literature, most widely discussed missing mechanisms are missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Among numerous techniques proposed to handle missing data (see Peugh& Enders, 2004), maximum likelihood is the most commonly used in empirical studies in social sciences and is considered the best approach in general (Enders &Bandalos, 2001); when the missing data are present, it is named as full information maximum likelihood (FIML). This method uses all the available information to estimate the model (Acock, 2005; Chen &Astebro; 2003). The likelihood function based on the available observed variables is first computed for each observation in the sample. The individual likelihoods are then summed to give the likelihood of the whole sample (Enders &Bandalos, 2001). The multivariate normality assumption plays an essential role in the FIML estimation under MAR (Allison, 2002; Chen &Astebro, 2003; Enders, 2004). FIML produces accurate model-data fits when data are multivariate normal (Enders, 2001). Researchers also found that FIML showed unbiased parameter estimates under MCAR and MAR (Enders &Bandalos, 2001).

In an empirical study involving relatively small sample size, large number of measurement indicators, and the presence of missing data, a parceling technique may be applied to obviate some of these difficulties in SEM analysis, when the purpose of the study is to examine the hypothesized relationship among latent factors, but not the psychometric properties of individual items. Schafer and Graham (2002) raised a similar idea but labeled it in an ambiguous manner. 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). The parcels are then used as indicators of latent factors in SEM analysis. They labeled this method as “case-by-case item deletion” or “ipsative mean imputation”. This method has been applied in empirical studies (e.g., Achenbach, Bernstein, &Dumenci,2005; Signorella, & Cooper, 2011; Yoder, Snell, & Tobias, 2012), however, “its properties remained largely unstudied” (Schafer & Graham, 2002, p. 158).

Examination of the effects of item parceling on model fit indices and parameter estimates has mainly been conducted under the conditions when the model assumptions are met. Specifically, the analysis model is consistent with the data generation model (Finney & DiStefano, 2006), item scores are continuously and multivariate normally distributed, and no missing data are present. A few studies focused on categorical item scores and misspecified model (Bandalos, 2002; Bandalos, 2008). However, as we are aware of, none of the studies examined item parceling techniques for data with missing values, and compared the performance of parcel-level analysis to item-level analysis using estimation methods other than ML. The purpose of this simulation study is to investigate how item parceling behaves under various conditions in SEM with missing and nonnormal distributed data via a simulation study. Results based on the parcels are compared with those based on the individual items. For both parcel-level and item-level analysis, both maximum likelihood and robust maximum likelihood estimation methods are applied.

## METHODS

Data were generated based on a structural equation model as shown in Figure 1. Specifically, the model consisted of three latent factors measured by 21 items. The first factor, F1, was measured by 15 items (item1 - item15); both the second and the third factors, F2 and F3, were measured by three items (item16 – item18 and item19 - item21, respectively). The factor loading was .70 from items 16-21. The factor loadings associated with F1 varied across items and were .40, .60, and .80. The variance of uniqueness for each item was fixed as one minus the squared loading. The path coefficient from F1 to F2 (F1→F2) and F1 to F3 (F1→F3) was .40 and .60, respectively. The covariance between the disturbances of F2 and F3 (F2 ↔F3) was .50. Based on this model, we evaluated the performance of parcel level analysis by manipulating four design factors in the simulation study: missing mechanism, percentage of missingness, degree of nonnormality of item scores, and sample size.

### *Design Factors for Data Generation*

- *Missing mechanisms*. Three missing mechanisms were considered: MCAR, MAR, and MNAR). The results from these conditions were compared to the corresponding conditions with no missing data.
- *Percentage of missingness.* Three levels of percentage of missingness were considered: 10%, 20%, and 40%. Previous studies have showed that percent of missingness had an effect on parameter estimates for analysis based on individual items (Davey &Savla, 2005; Enders, 2001).
- *Distribution of item scores.* Three types of distribution of item scores 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 1-15 with the same population skewness and kurtosis. Items 16 through 21 were distributed normally in all generation conditions.
- *Sample sizes*. 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.

These four design factors created a total of 90 conditions for data generation. Among them, 81 were formed from the conditions with missing data (3 patterns of missingness × 3 percentages of missingness × 3 types of distributions × 3 sample sizes), and 9 were formed from the conditions with no missing data (3 types of distributions × 3 sample sizes). For each condition, 2000 data sets were generated.

### *Data Generation Procedure*

Data were generated in R (version 2.13.2) based on the model shown in Figure 1. First, the correlation matrix among factors was obtained based on the parameters specified in the model. Second, 24 random variables were generated each with a mean of zero and standard deviation of one, three of which represented factor scores and the other 21 represented item uniqueness. Factors were then converted to have multivariate normal distribution with the intended correlations using the Cholesky decomposition method. Third, the observed item scores were obtained as a weighted linear combination of the factor score and item uniqueness such that (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*, $\lambda_i$ indicates the factor loading for the item *i*, and $E_{ij}$ indicates item uniqueness. For conditions with nonnormally distributed data, Fleishman's power transformation method (Fleishman, 1978) was applied to normally distributed data to obtain observed scores with the predefined skewness and kurtosis.

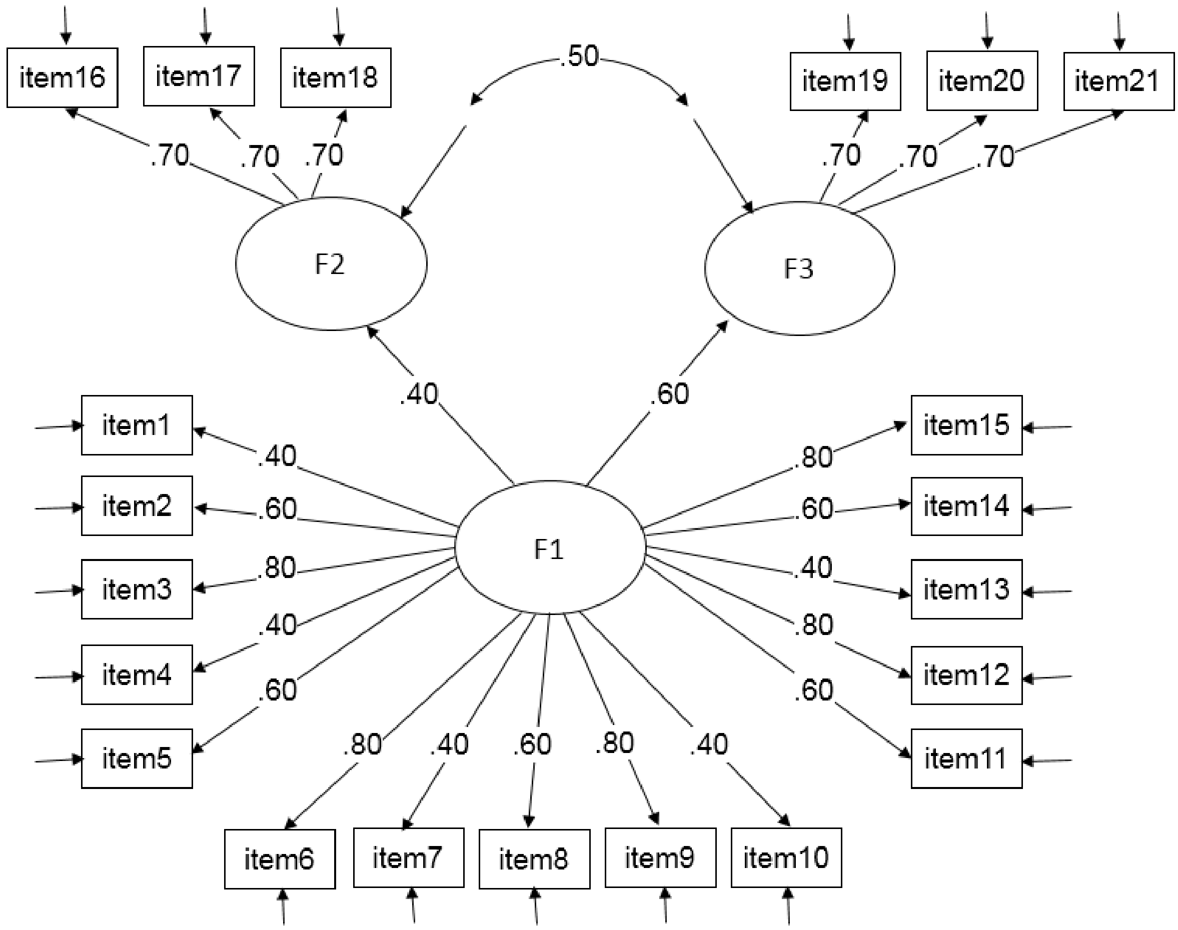


Figure 1. Models for Data Generation

The following rules were applied to create missing data. Only six items have missing values: items 1, 2, 3, 6, 7, and 8. To create data that are MCAR, randomly selected cases on the specified items were removed. To create data that are MAR, missingness on item1 was related to the values of item4. That is, item4 were sorted from the smallest to the largest, and then the cases with the lowest values on the item4 (e.g., 10%) were assigned being missing on item1 with a probability of .90. The rest of the values (highest 90%) were assigned being missing with a probability of .10. The same procedure was applied to the other five items; specifically, the missingness on the item2, 3, 6, 7, and 8 was related to the values on the item5, 16, 11, 12, and 13, respectively. For conditions with MNAR, missingness on variables was related to the variables themselves.

### *Data Analysis*

Mplus 6.1 (Muthén&Muthén, 1998-2008) was used for the data analysis. For each dataset, two different models were considered. One was based on item level and the other was based on parcel level. All analysis models were considered correctly specified given that the analysis model was consistent with the data generation model. To create parcel factorial parceling technique was used. The factorial algorithm technique decomposes “item specific components” and combines them within different parcels (Matsunaga, 2008). For the analysis based on parcel scores, a CFA model was first conducted on items 1 to 15. The items were sorted based on the magnitudes of the loadings (labeled as 1st to 15th from the largest to the smallest). Three parcels were then created such that the 1st parcel contained items with the order of 1st, 6th, 7th, 12th, and 13th, the 2nd parcel consisted of items with the order of 2nd, 5th, 8th, 11th, and 14th, and the 3rd parcel comprised the rest of the five items. The mean across the five (or available variables if there were missing values) items was computed as the parcel score. The items associated with F2 and F3 were not parceled. For both the item and the parcel level analyses, ML and MLR were used for model estimation.

### *Analysis of Outcome Variables*

For each condition, overall model-data fit was evaluated based on the chi-square test, Comparative Fit Index (CFI; Bentler, 1990) andRoot Mean Square Error of Approximation (RMSEA; Steiger& Lind, 1980). For the chi-square test, rejection rate based on the nominal level of .05 was reported. CFI values larger than .95 and RMSEA smaller than .08 were considered reasonable model fits (Hu &Bentler, 1999). For parameter estimates, direct effects among factors, that is, from F1 to F2 (F1→ F2) and from F1 to F3 (F1→ F3), and the covariance between the disturbances of F2 and F3 (F2 ↔ F3) were evaluated. Relative bias was computed for both point estimates and standard errors associated with the parameter estimates as:

$$\text{Relative Bias} = \frac{\hat{\theta} - \theta}{\theta} \times 100\%,$$

where $\hat{\theta}$ and $\theta$ indicate the mean of estimates and the population parameter, respectively. The true standard errors were approximated by using the standard deviations of parameter estimates based on the corresponding conditions without missing data (Bandalos, 2006). Hoogland and Boomsma (1998) suggested biases (in absolute value) smaller than 5% for point estimates and 10% for standard errors to be acceptable.

## RESULTS

For each condition, model convergence, rejection rates based on the chi-square test, and rejection rates based on fit indices were reported. Parameter estimates and their standard errors were examined by computing relative biases. There were no inadmissible solutions across all conditions except those conditions with a sample size of 100 and 40% of missing data when analyses were conducted on the item level. For these conditions, 53% to 72% of replications encountered inadmissible solutions and the rates of inadmissible solutions from ML estimation method were similar to those from MLR estimation method. Replications with inadmissible solutions were excluded from further analyses. Parcel-level analysis did not encounter any inadmissible solutions. In the following section, selected results were reported.

### *Model Rejection Rates Based on CFI and RMSEA*

CFI greater than .95 and RMSEA smaller than .08 indicate good model-data fit (Hu &Bentler, 1999). Table 2 report percentages of replications with CFI smaller than .95 and RMSEA greater than .08 for conditions with sample size of 100. Results for conditions with sample size of 300 and 1000 were not provided because the percentage was zero or closed to zero for these conditions.

Table 2.Percentage of Replications with CFI Smaller Than .95 and RMSEA Greater than .08 for Conditions with Sample Size of 100

| Sample Size | Skewness/Kurtosis | No Missing | MCAR | | | MAR | | | MNAR | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 10% | 20% | 40% | 10% | 20% | 40% | 10% | 20% | 40% |
| | | | | | | CFI | | | | | |
| Item | 0/0 | 18 (23) | 23 (30) | 30 (37) | 80 (52) | 24 (31) | 32 (40) | 78 (58) | 28 (34) | 36 (45) | 83 (60) |
| | 1/1.5 | 38 (43) | 44 (50) | 54 (63) | 90 (73) | 45 (51) | 52 (61) | 86 (75) | 86 (53) | 47 (65) | 89 (77) |
| | 1.75/3.75 | 78 (77) | 82 (83) | 86 (87) | 98 (91) | 81 (81) | 84 (86) | 97 (93) | 79 (81) | 84 (88) | 97 (93) |
| Parcel | 0/0 | 1 (2) | 1 (2) | 2 (3) | 2 (3) | 2 (3) | 2 (3) | 2 (3) | 2 (2) | 2 (3) | 2 (3) |
| | 1/1.5 | 2 (3) | 2 (3) | 4 (5) | 4 (6) | 2 (3) | 4 (5) | 4 (5) | 2 (4) | 3 (5) | 4 (5) |
| | 1.75/3.75 | 3 (4) | 3 (5) | 3 (5) | 3 (5) | 3 (4) | 3 (5) | 3 (5) | 3 (5) | 3 (5) | 3 (5) |
| | | | | | | RMSEA | | | | | |
| Item | 0/0 | 0 (0) | 0 (0) | 0 (0) | 12 (0) | 0 (0) | 0 (0) | 14 (0) | 0 (0) | 0 (0) | 11 (0) |
| | 1/1.5 | 0 (0) | 0 (0) | 0 (0) | 19 (1) | 0 (0) | 0 (0) | 18 (1) | 0 (0) | 0 (0) | 17 (0) |
| | 1.75/3.75 | 0 (1) | 0 (1) | 1 (2) | 43 (5) | 0 (1) | 1 (2) | 36 (5) | 0 (1) | 1 (2) | 33 (5) |
| Parcel | 0/0 | 4 (5) | 3 (5) | 4 (5) | 3 (5) | 4 (5) | 4 (6) | 4 (6) | 3 (5) | 4 (5) | 4 (6) |
| | 1/1.5 | 5 (7) | 4 (6) | 5 (7) | 5 (8) | 4 (7) | 5 (7) | 5 (7) | 4 (6) | 5 (7) | 4 (7) |
| | 1.75/3.75 | 4 (6) | 3 (6) | 3 (6) | 3 (6) | 3 (6) | 3 (6) | 3 (6) | 3 (6) | 3 (6) | 3 (6) |

*Note*. Percentages from robust maximum likelihood estimation method are in parentheses.

Consistent with the findings based on the model chi-square test, when the analysis was conducted at the item level, CFI tended to demonstrate misfit when the sample size was 100 (18%-97% of models had CFI<.95). Models were more likely to demonstrate misfit when larger percentage of data were missing and the distribution of data became more skewed. As sample size reached to 300, less than 9% of models had CFI<.95 and the percentage was nearly zero for most of the conditions. The pattern was similar across missing mechanisms. When the analysis was conducted at the parcel level, less than 6% of replications had CFI<.95 for all conditions and the percentages were nearly zero when sample size reached to 300, regardless of the degree of nonnormality, missing mechanism, and percentage of missingness.

The findings from RMSEA were slightly different. When the analysis was conducted at the item level data, less than 2% of models showed RMSEA>.08 when the sample size was 100, unless 40%

of data in the sample were missing; for these conditions, 11%-43% of models yielded RMSEA>.08. However, the percentage was less than 5% when MLR was applied. As sample size reached to 300, nearly 0% of models showed RMSEA>.08. When the analysis was conducted at parcel level, 3%-8% of models yielded RMSEA>.08 and the percentages were relatively stable across missing mechanisms, percentage of data being missing and the degree of nonnormality. In addition, results from MLR and ML were comparable.

### *Parameter Estimates and Standard Errors*

Table 3 reports relative bias of estimates for the parameter F1→ F2 from MLR. Parameter estimates from ML were identical to those from MLR and thus were not provided.

Table 3.Relative Bias (%) of the Direct Effect from F1 to F2

| Sample Size | Level | No Missing | MCAR | | | MAR | | | MNAR | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 10% | 20% | 40% | 10% | 20% | 40% | 10% | 20% | 40% |
| Skewness=0 & Kurtosis=0 | | | | | | | | | | | |
| 100 | Item | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| | Parcel | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 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 |
| Skewness=1 & Kurtosis=1.5 | | | | | | | | | | | |
| 100 | Item | -3 | -3 | -2 | -3 | -3 | -2 | -1 | -3 | -2 | -2 |
| | Parcel | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 |
| 300 | Item | -2 | -2 | -1 | -1 | -2 | -1 | -1 | -2 | -1 | -2 |
| | Parcel | -2 | -2 | -1 | -1 | -2 | -1 | -1 | -2 | -1 | -1 |
| 1000 | Item | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 |
| | Parcel | -1 | -1 | -2 | -2 | -1 | -1 | -1 | -1 | -2 | -1 |
| Skewness=1.75 & Kurtosis=3.75 | | | | | | | | | | | |
| 100 | Item | -11 | -11 | -12 | -13 | -11 | -12 | -11 | -11 | -11 | -11 |
| | Parcel | -10 | -10 | -10 | -10 | -10 | -9 | -9 | -9 | -10 | -9 |
| 300 | Item | -11 | -11 | -11 | -11 | -11 | -11 | -11 | -11 | -11 | -11 |
| | Parcel | -9 | -9 | -9 | -9 | -9 | -9 | -9 | -9 | -9 | -9 |
| 1000 | Item | -11 | -11 | -11 | -11 | -11 | -11 | -11 | -11 | -11 | -11 |
| | Parcel | -9 | -9 | -9 | -9 | -9 | -9 | -9 | -9 | -9 | -8 |

Results showed that relative bias of parameter estimates was mainly attributable to the degree of data nonnormality, while missing mechanisms, percentage of missingness, and sample size did not appear to influence the bias of the parameter estimates. When the data were normally distributed or with skewness of 1 and kurtosis of 1.5, the relative bias was smaller than 5% for all conditions. When the skewness and kurtosis increased to 1.75 and 3.75, parameter estimates tended to be negatively biased with the relative bias be greater than 5% (in the range of -9% to -11%). Parcel-level analysis and item-level analysis yielded very similar degrees of relative bias. Although not shown in the table, similar findings were obtained for the other two parameters F1→ F3 and F2 ↔ F3, except that the relative bias for F2 ↔ F3 tended to be positive when the data had skewness of 1.75 and kurtosis of 3.75.

Standard errors of parameter estimates were also examined. The detailed results from standard errors are available upon requests. In summary, relative bias of standard errors were similar across different missing mechanism and percentage of missingness. The absolute value of relative bias was smaller than 5% for all conditions, which was smaller than the suggested cutoff of 10% for being nontrivial (Hoogland&Boomsma, 1998). In addition, parcel-level analysis and item-level analysis yielded very similar degree of relative bias.

## DISCUSSION AND CONCLUSION

Applied 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, how parcel-level analysis behaves when the analysis involves with missing data and/or nonnormally distributed data has not been examined.In this study, we examined the performance of parcel-level analysis under various conditions of missing and nonnormally distributed data. Results from parcel-level analysis were compared to those from item-level analysis. Robust maximum likelihood (MLR) estimation method has been recommended when data demonstrate nonnormality. However, the use of MLR has not been discussed in the literature of parceling techniques, we thus analyzed data using both ML and MLR estimation methods.

Based on the results from this simulation study, we offered four reasons for advocating the use of parceling in SEM when the analysis involves a small sample size. The smallest sample size manipulated in this study was 100, which yielded a ratio of sample size to the number of observed variables being slightly less than 5:1. This ratio can be viewed as small in SEM analysis. First, because parceling reduces model complexity, parcel-level analysis is less likely than item-level analysis to encounter estimation difficulties when the sample size is small. Second, we found that model rejection rates based on RMSEA was around the idea of 5% when the analysis was conducted at the parcel level. The chi-square test and CFI were low when the analysis was conducted at the parcel level, while the rejection rates were too high for most of the conditions when the analysis was conducted at the item level, unless sample size was large (300 or 1000). Because the analysis model was considered correctly specified in our study, parcel-level analysis yields more reasonable empirical Type I error rates for chi-square test. These two findings are consistent with previous studies (e.g., Little et al., 2002). The current study adds to this existing literature that such an advantage becomes more obvious as the percentage of data being missing increases under all three missing mechanisms: MCAR, MAR, and MNAR. Third, although MLR corrected to certain degree for the inflated model chi-square when the data were nonnormally distributed, MLR from the item-level analysis still resulted in very high Type I error rates. On the contrary, the model rejection rates based on chi-square and fit indices were reasonable when the analysis was conducted at the parcel level. Four, parcel-level analysis and item-level analysis yielded similar estimates and standard errors for structural coefficients among latent factors. This finding itself does not support the use of parceling. This was consistent with the literature where it was indicated that parceling may not be appealing under optimal conditions (Matsunaga, 2008). However, parameter estimates tend not to be interpreted when the model and data show misfit. Instead, additional parameters will be added to the model in an attempt to improve model-data fit. In other words, item-level analysis is more likely to result in an over-parameterized model, particularly when the sample size is small.

Similar to other simulation studies, one should be cautious when generalizing these conclusions to other situations. First, all the generated data are continuous. However, in practice, data are often categorical. Future research may consider categorical incomplete data. Second, the analysis model was consistent with the data generation model. The performance of fit indices may be different and the superiority of parceling might not hold when the model is misspecified. Future research may consider misspecified models to investigate the performance of item parceling with missing data. Finally, only a limited number of levels were considered for each design factor. 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%. Consequently, only 11% (= 29% × 40%), at maximum, of the data were missing. Some other simulation studies have considered much higher percentage of missingness. For example, Davey and Savla (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 this study had is worth considering.

## 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. doi:10.1207/s15327752jpa8401_10

Acock, A. C. (2005).Working with missing values.*Journal of Marriage and Family, 67,* 1012-1028. doi:10.1111/j.1741-3737.2005.00191.x

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. doi:10.1037/0021-843X.112.4.545

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. doi:10.1207/S15328007SEM0901_5

Bandalos, D. L. (2006).*The Use of Monte Carlo Studies in Structural Equation Modeling Research.* In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (pp. 385-426). Greenwich, Connecticut: Information Age Publishing.

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. doi:10.1080/10705510801922340

Bandalos, D.L., & Finney, S. J. (2001).*Item parceling issues in structural equation modeling*. In G.A. Marcoulides& R.E. Schumacker (Eds.).*New developments and techniques in structural equation modeling* (pp.269-293). Mahwah, NJ: Lawrence Erlbaum Associates.

Bentler, P.M. (1990), Comparative Fit Indexes in Structural Models. *Psychological Bulletin, 107,* 238-246. doi:10.1037/0033-2909.107.2.238

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. doi:10.1037/0033-2909.105.3.467

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

Chen, G., &Astebro, T. (2003). How to deal with missing categorical data: Test of a simple Bayesian method. *Organizational Research Methods, 6,* 309-327. doi:10.1177/1094428103254672

Chou, C.P., Bentler, P.M., &Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: a Monte Carlo study. *British Journal of Mathematical and Statistical Psychology, 44,* 347–357. doi:10.1111/j.2044-8317.1991.tb00966.x

Curran, P.J., West, S.G., & Finch, J.F. (1996).The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1,* 16–29. doi:10.1037/1082-989X.1.1.16

Davey, A., &Savla, J. (2005). Issues in evaluating model fit with missing data. *Structural Equation Modeling, 12*(4), 578-597. doi:10.1207/s15328007sem1204_4

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. doi:10.1037/1082-989X.6.4.352

Enders, C. K. (2004). The impact of missing data on sample reliability estimates: Implications for reliability reporting practices. *Educational and Psychological Measurement, 64,* 419-436. doi:10.1177/0013164403261050

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. doi:10.1207/S15328007SEM0803_5

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 Modeling: A Second Course (pp. 269-314). Greenwich, Connecticut: Information Age Publishing.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika, 43*, 521-532. doi:10.1007/BF02293811

Hoogland, J. J. &Boomsma, A. (1998). Robustness Studies in Covariance Structure Modeling: An Overview and a Meta-Analysis. *Sociological Methods & Research, 26,* 329-367. doi:10.1177/0049124198026003003

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. doi:10.1080/10705519909540118

Hu, L., Bentler, P.M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112,* 351-362.

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. doi:10.1207/S15328007SEM0902_1

Matsunaga, M. (2008). Item parceling in structural equation modeling: A primer. *Communication Methods and Measures, 2,* 260-293. doi:10.1080/19312450802458935

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.

Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor-analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology, 45,* 19 –30

Muthén, B., &Muthén, L. (1998-2008*). Mplus user's guide (6th eedition).* 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. doi:10.1207/S15324818AME1601_4

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,* 729–757. doi:10.1177/0013164403258228

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. doi:10.3102/00346543074004525

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.

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. doi:10.1207/s15328007sem1304_4

Schafer, J. L., & Graham, J. W. (2002).Missing data : Our view of the state of the art. *Psychological Methods, 7*, 147-177. doi:10.1037/1082-989X.7.2.147

Signorella, M. L., & Cooper, J. E. (2011). Relationship suggestions from self-help books: *Gender stereotyping, preferences, and context effects. Sex Roles, 65,* 371-382. doi:10.1007/s11199-011-0023-4

Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Sterba, S.K. (2011). Implications of parcel-allocation variability for comparing fit of item-solutions and parcelsolutions. *Structural Equation Modeling, 18,* 554-577.

Sterba, S.K. &MacCallum, R.C. (2010). *Variability in parameter estimates and model fit across repeated allocations of items to parcels.* Multivariate Behavioral Research, 45, 322-358.

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. doi:10.1177/0146621609338592

Yoder, J. D., Snell, A. F., & Tobias, A. (2012). Balancing multicultural competence with social justice: Feminist beliefs and optimal psychological functioning. *The Counseling Psychologist, 40,* 1101-1132. doi:10.1177/001100001142629

## UZUN ÖZET

### *Giriş*

Yapısal Eşitlik Modellemesi (YEM) bir grup değişken arasındaki varsayılan ilişkilerin ampirik veriler kullanılarak test etilmesi için sıklıkla kullanılır. YEM'de yoğun olarak kullanılan maksimum olabilirlik (ML) tahmin metodu örneklem büyüklüğünün yüksek ve gözlenen değişkenlerin normal dağılmasını varsaymaktadır. Bu varsayımların sağlanmaması ki-kare istatistiği, uyum ideksleri, parameter tahmini ve parameter tahmininin standart hatalarının hatalı olmasına sebep olmaktadır (Bollen, 1989; Curran, West, & Finch, 1996; Chou, Bentler, &Satorra, 1991). Ayrica bu yanlılık modelin karmaşıklığı arttıkça artma eğilimi göstermektedir. Varsayımların sağlanmaması durumunda ML'ye alternatif olarak sağlam maksimum olabilirlik (MLR) tahmin metodu kullanılabilir. MLR ki-kare değerlerindeki pozitif yanlılığı ve standart hatalardaki negatif yanlılığı düzeltmektedir.

YEM'in amacının tekil maddelerin özelliklerini test etmekden ziyade gizil değişkenler arasındaki ilişkiyi incelemek olduğunda, özellikle örneklem büyüklüğünün yetersiz olduğu ve herbir gizil değişkenin gösterge sayısının fazla olması durumunda, parsel tekniklerinden biri kullanılabilir (Sterba & MacCallum, 2010). Parselleme birçok maddenin toplam veya ortalama puanlarının hesaplanması olarak tanımlanan bir prosedür olarak tanımlanır. Parsel olarak tanımlanan bu yeni değişkenler bireysel maddeler yerine YEM analizinde gizil değişkenlerin göstergesi olarak kullanılır (Bandalos 2002, 2008; Little, Cunningham, Shahar, &Widaman, 2002; Sass & Smith, 2006; Sterba, 2011; Sterba&MacCallum, 2010; Yang, Nay, & Hoyle, 2010). Parsel kullanmak model karmaşıklığını azalttığı, örneklem gereksinimin azalttığı, bireysel maddelerden kaynaklanan sistematik hataların model üzerindeki etkisini azalttığı, yüksek güvenilirliğe yardımcı oluğundan, modellerin yakınsama oranlarını arttırdığından ve maddelerin boyutları bilindiğinde model veri uyumunu arttırdığından dolayı caziptir. Sass ve Smith (2006) model varsayımlarının sağlanması ve maddelerin tek boyutlu olduğu durumlarda parsellemenin faktörler arasındaki yapısal ilişki parametrelerinin tahmininde yanlılığa sebep olmadığını göstermişlerdir.

Kayıp verilerin varlığı YEM analizlerinde muhtemel problemler oluşturmaktadır. Literatürde en yaygın tartışılan kayıp veri türleri *tamamen rastgel kayıp* (MCAR), *rastgel kayıp* (MAR) ve *rastgel olmayan kayıp* (MNAR) olarak tanımlanabilir. Kayıp verileri ele alan bir çok teknikten sosyal bilimlerdeki ampirik analizlerde en çok kullanılan maksimum olabilirlik metodudur (Enders & Bandalos, 2001). Kayıp verilerin olması durumunda bu metod tam bilgi maksimum olabilirlik (FIML) olarak adlandırılır. Araştımacılar FIML metodunun verilerin MCAR ve MAR olması durumunda yansız parametre tahmini yaptığını bulmuslardır (Enders & Bandalos, 2001).

Örneklem büyüklüğünün düşük olması, modelin karmaşık olması ve verilerde kayıpların olması ve araştırmanın amacının maddeler arasındaki ilişkiden ziyade faktörler arasındaki ilişkiyi incelemek olduğunda, ampirik çalışmalarda parcelleme tekniği, YEM analizlerindeki bazı zorlukları gidermek için kullanılablir. Schafer and Graham (2002) buna benzer fikirleri farklı şekilde adlandırarak (örn. "*case-by-case item deletion*" veya "*ipsative mean imputation*") ileri sürmüştür.

Buna bağli olarak, çalışmanın amacı, parcellemenin kayıp ve normal olmayan verilerle YEM analizlerinde nasıl davranacağını simülasyon aracılığıyla araştırmaktır. Parcellemeye dayalı sonuçlar bireysel maddelerle oluşturalan modellerle karşılaştırılmıştır.

### *Metod*

Veriler şekil 1'de gösterilen modele göre üretilmiştir. Faktör yükleri şekilde gösterildiği gibidir. Her bir maddenin hata terimi ise bir eksi yükün karesi olarak sabitlenmiştir. Veri üretimini için kullanılan dizayn faktörleri şu şekildedir:

- Kayıp mekanizması: Üç farklı mekanizma dikkate alınmştır; MCAR, MAR ve MNAR.
- Kayıp yüzdesi: Üç farklı yüzde seviyesi dikkate alınmştır; 10%, 20% ve 40%.
- Madde puanlarının dağılımı: Üç farklı dağılım dikkate alınmştır; (1) normal dağılım; (2) orta seviye çarpık ve duşük basıklık (Sk =1, ve K=1.5); (3) yüksek çarpıklık ve yüksek basıklık (Sk =1.75 ve K=3.75). Sadece 1-15 arasındaki maddelerde non-normallik kullanılmıştır.
- Örneklem büyüklüğü: Üç farklı örneklem büyüklüğü (SS) dikkate alınmştır; 100, 300, 1000.

Bu dizayn faktörler kullanılarak, toplam 90 farklı durum oluşturulmustur. Herbir durum için 2000 veri seti, R (versiyon 2.13.2) programında üretilmiştir. İlk olarak faktörler arasındaki korelasyon elde edilmiştir. Daha sonra ortalama sıfır ve standart sapma bir olacak şekilde rastgele 24 değişken üretilmiştir. Cholesky ayrıştırma metodu kullanılarak faktör puanları çoklu normal dağılacak şekilde dönüştrülmüştür. Üçüncü olarak, gözlenen madde puanları faktörlerin ve hata terimlerinin doğrusal bir kombinasyonu olarak elde edilmiştir (örn. Bernstein & Teng, 1989). Sabit çarpıklık ve basıklık değerleri için Fleishman'ın tekniği kullanılarak gözlenen değerler oluşturulmuştur (Fleishman, 1978). Verilerin MCAR olması durumunda rastgele seçilen değerler değişkenlerden silinmiştir. MAR olması durumunda ise madde 1, 2, 3, 6, 7 ve 8'in değerleri sırasıyla 4, 5, 16, 11, 12 ve 13'ün değerlerine göre silinmiştir. MNAR olması durumunda ise değişkenlerin değerleri kendi değerlerine göre silinmiştir.

Mplus 6.1 (Muthén & Muthén, 1998-2008) programı kullanılarak herbir veri seti madde ve parsel seviyelerinde test dilmiştir. Parcel oluşturmak için factorial parselleme tekniği kullanılmış (Matsunaga, 2008). Herbir model ise ML ve MLR tekniği altında test edilmiştir. Her bir durum için genel model veri uyumu ki-kare testi, karşılastırmalı uyum indeksi (CFI) ve kök ortalama kare yaklaşım hatasına (RMSEA) dayanarak Hu ve Bentler'in (1999) kriterleriyle karşılaştırılmıştır. Ayrıca nokta tahmini ve standart hatası için göreli yanlılık değerleri hesaplanmıştır.

### *Sonuçlar ve Tartışma*

Her bir durum için model yakınsaması ve model reddetme oranları rapor edilmiştir. Tablo 1 ki-kare ye dayalı model reddetme oranlarını göstermektedir. Model doğru tanımlanmış olduğudan red oranının 5% olması beklenir. Madde seviyesindeki ML'ye dayalı oranlar 5% 'den oldukça yüksektir. Bu oran SS=1000 olması durumunda azalmaktadır. Bu reddetme oranı kayıp veri yüzdesi ile artmaktadır. Aynı şartlar altında parsel seviyesindeki modellerin reddetme oranları daha düşüktür ve dağılıma kayıp mekanizmasına ve kayıp yüzdesine göre göreceli olarak sabittir.

Sonuçlara göre, tahmin yanlılığı çoğunlukla verilerin normal dağılmamasından kaynaklanmaktadır. Kayıp oranı ve mekanizması ve örneklem büyüklüğünün yanlılığı etkilediği söylenemez. Parsel ve madde analizleri benzer parametre yanlılıkları ve parametre tahmininin standart hata yanlılığı göstermiştir.

Parselleme teknikleri ampirik çalışmalada kullanılmaktadır, özellikle örneklem büyüklüğünün düşük olması durumunda. Parsellemenin model karmaşıklığını azalttığı ve daha iyi bir model-veri uyumu sağladığı gibi bazı yararları vardır. Bu çalışmada parselleme tekniğinin kayıp ve normal olmayan verilerdeki performansı incelenmiştir.

Sonuçlara göre, parsel seviyesindeki modellerde daha az tahmin zorluğu olmuştur, özellikle küçük örneklem büyüklüğünde. Örnekleme büyüklüğünün yüksek olması durumlarindan baska, parsel seviyesinde ki-kare ve CFI'ya dayalı model reddetme oranı yüksek fakat RMSEA değerleri .05

civarındadır. Ayrıca, madde seviyesinde model reddetme oranları MLR'nin kullanılması durumunda da parsel seviyesindeki değerlerden yüksektir. Son olarak, madde ve parsel seviyesinde modeller YEM parametereleri ve bunların standart hataları bakımından benzer sonuçlar vermiştir. Bu çalışmada sürekli değişkenler üzerinde ve doğru tanımlanmış modeller üzerinden simülasyon yapılmıştır. Daha sonraki çalışmalarda doğru tanımlanmamış maddeler veya kategorik değişkenler için bir simülasyon yapılabilir.