Pertanika J. Sci. & Technol. 28 (2): 627 - 647 (2020)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Effects on Parameter Estimates and Goodness of Fit Measures: Comparing Item-Level and Item-Parcel Models in Structural Equation Modeling

Ainur Amira Kamaruddin[1]*, Bee Wah Yap[2] and Sayang Mohd Deni[2]

[1]Centre of Statistical and Decision Science Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[2]Advanced Analytics Engineering Centre, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

ABSTRACT

The assessment of model fit is important in Structural Equation Modeling (SEM). Several goodness-of-fit (GoF) measures are affected by sample size and the number of parameters to be estimated. A large sample size is needed to test a complex model involving a large number of parameters to be estimated. One of the solutions to reduce the number of parameters to be estimated in a given model is by considering item parceling. The effects of item parceling on parameter estimates and GoF measures in a structural equation model was investigated via a simulation study. The simulation results indicate that the parameter estimates are closer to the true parameter values for the IL model whenever the distribution of data is normal but biased when the data is highly skewed. The parameter estimates for the IP model were found to be underestimated for both normal and non-normal data. The GoF measures were higher for the IP model. Additionally, the RMSEA was lower for the IP model when data were skewed. This shows that item parceling may improve GoF measures but the effect of exogenous on endogenous variable is underestimated. Application to a real data set confirmed the results of the simulation study.

Keywords: Goodness of fit, item parceling, parameter estimates, simulation, structural equation modeling

ARTICLE INFO

Article history:
Received: 05 November 2019
Accepted: 04 February 2020
Published: 15 April 2020

E-mail addresses:
memiera87@yahoo.com (Ainur Amira Kamaruddin)
beewah@tmsk.uitm.edu.my (Bee Wah Yap)
sayang@tmsk.uitm.edu.my (and Sayang Mohd Deni)
* Corresponding author

INTRODUCTION

Structural equation modeling (SEM) is a multivariate statistical analysis technique used to identify the association between more than one variable. SEM is a unique combination of multivariate techniques such as factor analysis and multiple regression

ISSN: 0128-7680
e-ISSN: 2231-8526

analysis. However, in SEM, the independent variable in a regression equation can be the dependent variable in another regression equation (Hair et al., 2015). SEM is widely used in psychology (Marsh et al., 1988; Mulaik et al., 1989; Bentler, 1990; Curran et al., 1996; Iacobucci, 2010; Little et al., 2013), marketing (Jöreskog & Sörbom, 1982; Bearden et al., 1982; Bagozzi & Yi, 1988), management (Shah & Goldstein, 2006; Rocha & Chelladurai, 2012; Hair et al., 2015), organizational behavior (Landis et al., 2000; Ryu, 2011), and business research (Sharma et al., 2005).

One of the most intensive parts in SEM is estimating the parameter of the model. Maximum Likelihood (ML) is the most popular method of estimation chosen by researchers (Curran et al., 1996; Hall et al., 1999; Nasser & Wisenbaker, 2006; Ory & Mokhtarian, 2010; Sterba & MacCallum, 2010; Ryu, 2011; Hamzah et al., 2017). The estimation process in ML aims to yield the minimum discrepancy between the sample covariance matrix and the estimated covariance matrix (Byrne, 2010). ML has several advantages such as being more stable, produces reliable results and is more accurate compared to other estimation methods (Olsson et al., 2000; Olsson et al., 2004). Besides, it is robust to moderate departures from normality assumption (Hair et al., 2015).

Another important aspect of SEM is the assessment of model fit for the structural model. It is used to test the consistency of a proposed theoretical model with the data. The model is clarified as good when the estimated covariance matrix is sufficiently close to the observed covariance matrix (Hair et al., 2015). The most popular goodness-of-fit (GoF) measures used to examine model fit are chi-squared (x^2), root mean square error of approximation (RMSEA), goodness-of-fit index (GFI), Adjusted goodness-of-fit index (AGFI), Normed Fit Index (NFI), Non-Normed Fit Index (NNFI) or also known as Tucker-Lewis index (TLI) and comparative fit index (CFI).

Several GoF measures are influenced by sample size and number of estimated parameters (Marsh et al., 1988; Ding et al., 1995; Hooper et al., 2008; Rocha & Chelladurai, 2012). Parameter estimates may be biased when a construct consists of a small number of indicators. Thus, a minimum of three indicators per construct is recommended to produce unbiased parameter estimates (Ding et al., 1995; Hair et al., 2015; Iacobucci, 2010). However, the analysis of SEM is problematic for a model that consists of a large number of indicators with a small sample size (Deng et al., 2018). Hence, Hair et al. (2015) suggested that minimum sample sizes were defined with a complement of model complexity and basic measurement model characteristics. The required minimum sample size can be easily achieved for a simple model with a small number of indicators. Nonetheless, a larger sample size is required for a complex model with a large number of indicators (Rocha & Chelladurai, 2012; Deng et al., 2018; Hair et al., 2015). To overcome this problem, some studies (Hall et al., 1999; Landis et al., 2000; Bandalos, 2002; Kim & Hagtvet, 2003; Nasser & Wisenbaker, 2006; Rocha & Chelladurai, 2012) used item parceling when the model consisted of a large number of indicators. The purpose of item parceling is to reduce the model complexity by summing or averaging together two or more indicators as a parcel rather than using individual items as indicators. A new model created by the item parceling technique reduces the nuisance parameters and sampling variability and increases reliability (Little et al., 2013), and improves model fit (Hall et al., 1999; Landis et al., 2000; Sterba & MacCallum, 2010).

Although the application of item parceling has received attention among researchers in SEM, the debates on the use of item parceling have continued and it remains a controversial issue. Bandalos (2002) pointed out the disadvantages of using the parceling technique such as the dimensionality of original measures being doubtful, resulted in biased parameter estimates and improvement in model fit without taking into account the model misspecification issue. The item parceling technique may also make the multidimensional constructs reflect as unidimensional when the indicators are not well defined in the constructs (Bandalos, 2002). Therefore, Little et al. (2002) emphasized the need to understand the original structure and dimension before assigning items to parcels. Little et al. (2013) suggested that a simulation study should be carried out to investigate whether the item parcel (IP) model was affected by the parameter estimates in SEM by comparing the results of the item-level (IL) model. Thus, this study aims to examine whether the parameter estimates and model fit are affected by item parceling in a structural equation model and compare these results with the IL model. The statistical analyses of this simulation study were performed using IBM SPSS Statistic 20, AMOS and R programming language software.

MATERIALS AND METHODS

Review of Theoretical Framework

Sterba and MacCallum (2010) presented the theoretical developments by MacCallum and Tucker (1991) on population data. Let an i subscript denotes item-level, x_i the vector deviation score on items in the population, and E denotes expectation operator. The population covariance structure of the items is given by Equation 1

$$E\left(x_i, x_i'\right) = \mathbf{\Sigma}_i = \mathbf{\Lambda}_i \mathbf{\Phi}_i \Lambda_i' + \mathbf{\Psi}_i^2 \qquad [1]$$

where $\mathbf{\Lambda}_i$ is the common factor loading, $\mathbf{\Phi}_i$ is the covariance of common factor loading and $\mathbf{\Psi}_i$ is the diagonal of unique variances. The unique factors are assumed to be independent from each other and from common factors in the population.

To illustrate, suppose we want to construct n parcels from a set of m items representing q factors (n = 4, m = 12 and q = 2). Let a p subscript denotes parcel-level and $\mathbf{A}$

be an $m \times n$ selection matrix that allocates items to parcels. Given the vector of parcels, $x_p = \mathbf{A}x_i$ can be presented in the matrix form as

$$\begin{bmatrix} x_{p1} \\ x_{p2} \\ x_{p3} \\ x_{p4} \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \end{bmatrix} \begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ x_{i4} \\ x_{i5} \\ x_{i6} \\ x_{i7} \\ x_{i8} \\ x_{i9} \\ x_{i10} \\ x_{i11} \\ x_{i12} \end{bmatrix}$$

$$\begin{bmatrix} x_{p1} \\ x_{p2} \\ x_{p3} \\ x_{p4} \end{bmatrix} = \begin{bmatrix} (x_{i1} + x_{i2} + x_{i3})/3 \\ (x_{i4} + x_{i5} + x_{i6})/3 \\ (x_{i7} + x_{i8} + x_{i9})/3 \\ (x_{i10} + x_{i11} + x_{i12})/3 \end{bmatrix}$$

where, x_{p1} and x_{p2} belong to the first factor and x_{p3} and x_{p4} belong to the second factor. The population covariance structure of the parcels can be derived as Equation 2

$$\mathbf{\Sigma}_p = E(x_p, x_p') \qquad [2]$$

and can be rewritten as Equation 3

$$\mathbf{\Sigma}_p = E(\mathbf{A}x_i x_i' \mathbf{A}') = \mathbf{A}\Sigma_i \mathbf{A}' \qquad [3]$$

which implies Equation 4

$$\mathbf{\Sigma}_p = \mathbf{A}\mathbf{\Lambda}_i \mathbf{\Phi}_i \mathbf{\Lambda}_i' \mathbf{A}_i' + \mathbf{A}\mathbf{\Psi}_i^2 \mathbf{A}' \qquad [4]$$

Let $\mathbf{\Lambda}_p = \mathbf{A\Lambda}_i$ and $\mathbf{\Psi}_p^2 = \mathbf{A\Psi}_i^2\mathbf{A}'$, then (Equation 5)

$$\mathbf{\Sigma}_p = \mathbf{\Lambda}_p\mathbf{\Phi}_i\mathbf{\Lambda}'_p + \mathbf{\Psi}_p^2 \quad [5]$$

The factor loading for a parcel will equal the average of factor loading at the item-level when parcels are formed by averaging items. However, these conditions will hold in the population when the factor model fits perfectly at the item-level.

Sterba and MacCallum (2010) considered the structure of item-level sample covariance matrix developed by MacCallum and Tucker (1991). The assumptions that the unique factors are independent from each other and also with common factors cannot hold due to sampling variability. Thus, the item-level sample covariance structure is denoted by Equation 6

$$\mathbf{C}_i = \mathbf{\Lambda}_i\mathbf{C}_{cc_i}\mathbf{\Lambda}'_i + \mathbf{\Lambda}_i\mathbf{C}_{cu_i}\mathbf{\Psi}'_i + \mathbf{\Psi}_i\mathbf{C}_{uc_i}\mathbf{\Lambda}'_i + \mathbf{\Psi}_i\mathbf{C}_{uu_i}\mathbf{\Psi}'_i \quad [6]$$

Where $\mathbf{C}_{cc_i}$ is the sample covariance matrix of common factors, $\mathbf{C}_{cu_i}$ is the sample covariance matrix of common and unique factors, $\mathbf{C}_{uc_i}$ is the sample covariance matrix of unique and common factors and $\mathbf{C}_{uu_i}$ is the sample covariance matrix of unique factors. Let $\mathbf{\Lambda}_p = \mathbf{A\Lambda}_i$ and $\mathbf{\Psi}_p = \mathbf{\Psi}'_i\mathbf{A}' = \mathbf{A\Psi}_i$, then (Equation 7)

$$\mathbf{C}_p = \mathbf{\Lambda}_p\mathbf{C}_{cc_i}\mathbf{\Lambda}'_p + \mathbf{\Lambda}_p\mathbf{C}_{cu_i}\mathbf{\Psi}_p + \mathbf{\Psi}_p\mathbf{C}_{uc_i}\mathbf{\Lambda}'_p + \mathbf{\Psi}_p\mathbf{C}_{uu_i}\mathbf{\Psi}_p \quad [7]$$

The item-level sample covariance structure that represents lack of fit due to the sampling error, $\mathbf{\Delta}_{SE_i}$ can be simplified as in Equation 8

$$\mathbf{C}_i = \mathbf{\Lambda}_i\mathbf{C}_{cc_i}\mathbf{\Lambda}'_i + \mathbf{\Psi}_i^2 + \mathbf{\Delta}_{SE_i} \quad [8]$$

Hence, the parcel-level sample covariance structure can be derived as Equation 9

$$\mathbf{C}_p = \mathbf{A\Lambda}_i\mathbf{C}_{cc_i}\mathbf{\Lambda}'_i\mathbf{A}' + \mathbf{A\Psi}_i^2\mathbf{A}' + \mathbf{A\Delta}_{SE_i}\mathbf{A}' \quad [9]$$

Let $\mathbf{\Lambda}_p = \mathbf{A\Lambda}_i$, $\mathbf{\Psi}_p^2 = \mathbf{A\Psi}_i^2\mathbf{A}'$ and $\mathbf{\Delta}_{SE_p} = \mathbf{A\Delta}_{SE_i}\mathbf{A}'$, then (Equation 10)

$$\mathbf{C}_p = \mathbf{\Lambda}_p\mathbf{C}_{cc_i}\mathbf{\Lambda}'_p + \mathbf{\Psi}_p^2 + \mathbf{\Delta}_{SE_p} \quad [10]$$

The effect of sampling error can be reduced when the sample size is large, item communalities are high (where the unique loadings $\mathbf{\Psi}_i$ are low, hence $\mathbf{C}_{uui}$, $\mathbf{C}_{cui}$ and $\mathbf{C}_{uci}$

matrices have little weight), and with smaller dimensions of $\mathbf{C}_{uui}$, $\mathbf{C}_{cui}$ and $\mathbf{C}_{uci}$ matrices (since more items are located in each parcel) (Bandalos, 2002; Sterba & MacCallum, 2010).

METHODS

Study 1: Simulation

To provide empirical evidence on the usefulness of item parceling, a simulation study was used to identify the effects of item parceling on parameter estimates and model fit based on different sample sizes and distribution of data by comparing the IL model and IP model. The simulations were carried out using the function of *cfa* from the *lavaan package* built in the R programming software. This study generated data for a structural equation model for the IL model and IP model. Sample sizes of 100, 150, 200, 250, 300, 500, 1000, 1500 and 2000 were considered with different distribution of data which are normal (skewness = 0, kurtosis = 3), non-normal (skewness = 1, kurtosis = 1.5) and non-normal (skewness = 1.75, kurtosis = 3.75).

The structural model was adapted from a study by Goodhue et al. (2012) which had twelve items for exogenous variable and three items for endogenous variable. However, this study only focused on the single path coefficient between exogenous and endogenous variables. The smallest sample size (n = 100) was considered as recommended by Hair et al. (2015). The method for generating non-normal data followed the work of Goodhue et al. (2012) by using the Fleishman method (Fleishman, 1978) which is commonly used in the generation of non-normal data (Orcan, 2013; Goodhue et al., 2012; Morgan et al., 2016). Two non-normal distributions were considered based on Orcan (2013) study to represent moderate skewness with low kurtosis (skewness = 1, kurtosis = 1.5) and high skewness with high kurtosis (skewness = 1.75, kurtosis = 3.75).

For the IL model, twelve items for exogenous variable and three items for endogenous variable were generated with standardized loadings repeatedly fixed to 0.70, 0.80 and 0.90. The higher loadings (≥0.7) were selected to reach convergence and model stability as suggested by Hair et al. (2015). For the IP model, the twelve items for exogenous variable generated in the IL model were assigned randomly to four parcels. The parcel scores were computed by averaging three items for each parcel. Due to identification reasons, one factor loading for each factor was fixed to 1 (Lyhagen & Kraus, 2013). The simulation process was repeated for 5000 replications for each combination (9 sample sizes x 3 distributions x 2 models). The mean of parameter estimates and model fit were calculated for each combination.

Study 2: Empirical Example

In Study 2, we used an empirical example to observe the effect of item parceling on parameter estimation and model fit for a real dataset. This empirical dataset was from a study on self-efficacy and quality of life (QoL) of mothers with autistic children conducted by Nasir (2015). This study involved 181 respondents. The original (IL) model consisted of three latent variables with two endogenous variables: health related quality of life (HRQOL) and parenting sense of competence scale (PSOC) and one exogenous variable: parenting stress index (PSI). The HRQOL was a second-order construct measured by two factors (*physical* and *mental*) with five and four items, respectively. The PSOC was a second-order construct measured by two factors (*value comforting* and *skill knowledge*) with four items for each factor. Finally, the PSI was a second-order construct measured by three factors (*parental distress*, *parent child* and *difficult child*) with ten, eight and six items, respectively. Figure 1 shows the path diagram of the empirical example.

As mentioned before, to study the effectiveness of item parceling on parameter estimation and model fit, three different models of item parceling were tested. For Model 1, the items for *parental distress*, *parent child* and *difficult child* were parceled into five, four and three parcels of two items each (Figure 2). The items for *value comforting*, *skill knowledge*, *physical* and *mental* remained unchanged. Model 2 was similar to Model 1 except that the items for *value comforting*, *skill knowledge* and *mental* were parceled to form two parcels of two items each (Figure 3). Meanwhile, the four items for *physical* were parceled to form two parcels of two items each and the remaining item remained as a single indicator. Model 3 represented the simplest model where all the items in each factor were parceled together. The structures of these models are summarized in Table 1. The data analysis was carried out using SPSS AMOS Version 22.

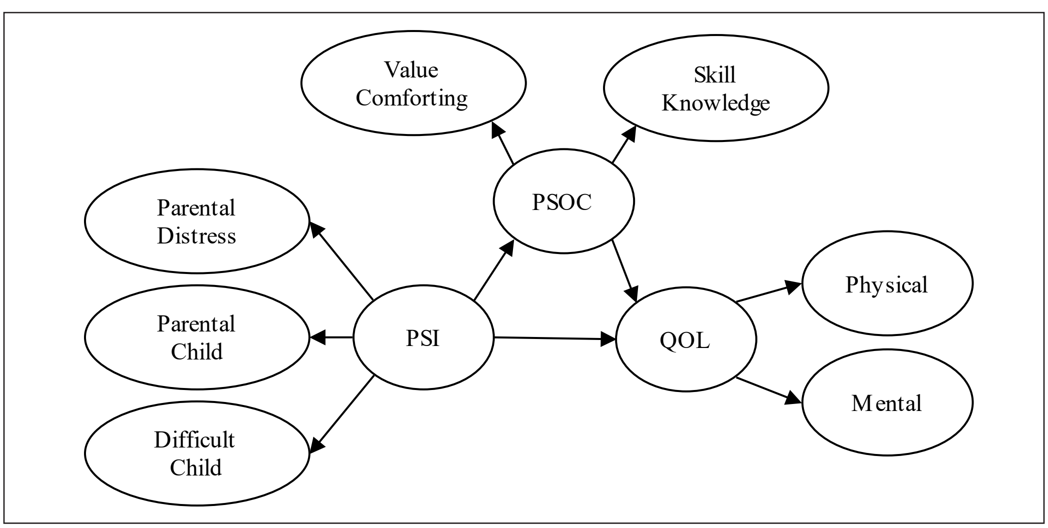

Figure 1. The path diagram of empirical example

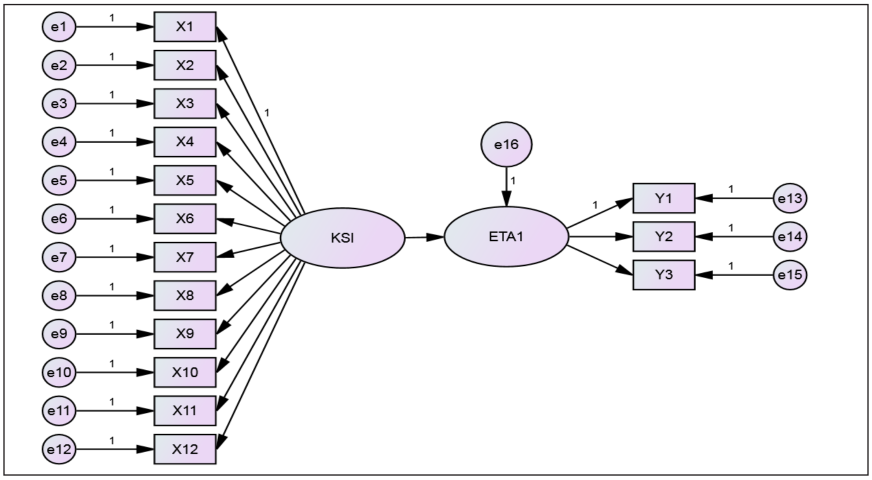

Figure 2. Model 1: Item-level model

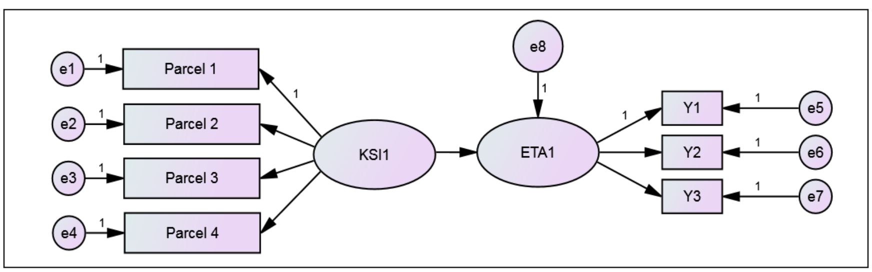

Figure 3. Model 2: Item-parcel model

Table 1
Summary of model structures

		IL	IP		
			Model 1	Model 2	Model 3
parenting stress index (PSI)	parental distress	B1, B3, B4, B5, B6, B7, B8, B9, B10, B12	B1B3, B4B5, B6B7, B8B9, B10B12	B1B3, B4B5, B6B7, B8B9, B10B12	Averaging all items
	parent-child	B14, B15, B16, B17, B18, B19, B20, B22	B14B15, B16B17, B18B19, B20B22	B14B15, B16B17, B18B19, B20B22	Averaging all items
	difficult child	B25, B26, B27, B28, B29, B30	B25B26, B27B28, B29B30	B25B26, B27B28, B29B30	Averaging all items
arenting sense of competence (PSOC)	skill knowledge	C10, C11, C13, C15	C10, C11, C13, C15	C10C11, C13C15	Averaging all items
	value comforting	CC9, CC12, CC16, C17	CC9, CC12, CC16, C17	CC9CC12, CC16C17	Averaging all items
health related quality of life (HRQoL)	physical	D2, D3, D4, D5, DD8	D2, D3, D4, D5, DD8	D2D3, D4D5, DD8	Averaging all items
	mental	D6, D7, D11, D12	D6, D7, D11, D12	D6D7, D11D12	Averaging all items
Total items		41	29	21	7

RESULTS AND DISCUSSION

Study 1: Results of Simulation Study

This section presents the results of parameter estimates, mean square error and model fit for the IL model and IP model for different distributions of data via a simulation in Study 1. Table 2 displays the performance of parameter estimates for the IL model and IP model under different sample sizes and distributions of data. The results indicate that the parameter

Table 2
Parameter estimates, standard deviation and mean square error

n	$\beta = 0.585$					
	Model 1: IL			Model 2: IP		
	$\hat{\beta}$	SD	MSE	$\hat{\beta}$	SD	MSE
[a]100	0.588	0.132	0.017	0.510	0.102	0.016
[b]100	0.583	0.152	0.023	0.501	0.115	0.020
[c]100	0.539	0.201	0.043	0.445	0.143	0.040
[a]150	0.592	0.109	0.012	0.514	0.085	0.012
[b]150	0.586	0.126	0.016	0.505	0.096	0.015
[c]150	0.540	0.164	0.029	0.449	0.118	0.032
[a]200	0.589	0.092	0.008	0.513	0.071	0.010
[b]200	0.582	0.105	0.011	0.504	0.080	0.013
[c]200	0.533	0.135	0.021	0.448	0.100	0.029
[a]250	0.585	0.080	0.006	0.511	0.063	0.010
[b]250	0.577	0.093	0.009	0.501	0.072	0.012
[c]250	0.529	0.120	0.018	0.445	0.090	0.028
[a]300	0.588	0.074	0.005	0.513	0.057	0.008
[b]300	0.580	0.086	0.007	0.503	0.066	0.011
[c]300	0.531	0.112	0.015	0.447	0.083	0.026
[a]500	0.585	0.056	0.003	0.512	0.045	0.007
[b]500	0.577	0.064	0.004	0.502	0.050	0.009
[c]500	0.526	0.082	0.010	0.445	0.062	0.024
[a]1000	0.585	0.040	0.002	0.512	0.032	0.006
[b]1000	0.577	0.046	0.002	0.502	0.036	0.008
[c]1000	0.526	0.059	0.007	0.445	0.045	0.022
[a]1500	0.585	0.033	0.001	0.512	0.026	0.006
[b]1500	0.576	0.038	0.001	0.502	0.030	0.008
[c]1500	0.525	0.048	0.006	0.445	0.037	0.021
[a]2000	0.586	0.028	0.001	0.512	0.022	0.006
[b]2000	0.577	0.032	0.001	0.502	0.025	0.007
[c]2000	0.525	0.041	0.005	0.445	0.031	0.020

Note. n is sample size. IL is item-level model. IP is item-parcel model. $\hat{\beta}$ is the estimated parameter. SD is the standard deviation. MSE is the mean square error. [a]Normal distribution (skewness=0, kurtosis=3).[b]Non-normal distribution (skewness=1, kurtosis=1.5).[c]Non-normal distribution (skewness=1.75, kurtosis=3.75).

estimates approach the true parameter value for the IL model under normal condition but biased when the data is highly skewed. The simulation results also show that the parameter estimates for the IP model are underestimated for both normal and non-normal data. The mean square error (MSE) measures the closeness of fitted $\hat{\beta}$ to the true value β . The MSE for both the IL model and IP model decreases when sample size increases. The MSE is the lowest for IL model and IP model when data is normal. The MSE is higher for both models when data is non-normal.

The box-plots in Figure 4 show that the dispersion (standard deviation) of parameter estimates is large for a small sample size. As expected, the dispersion of the parameter

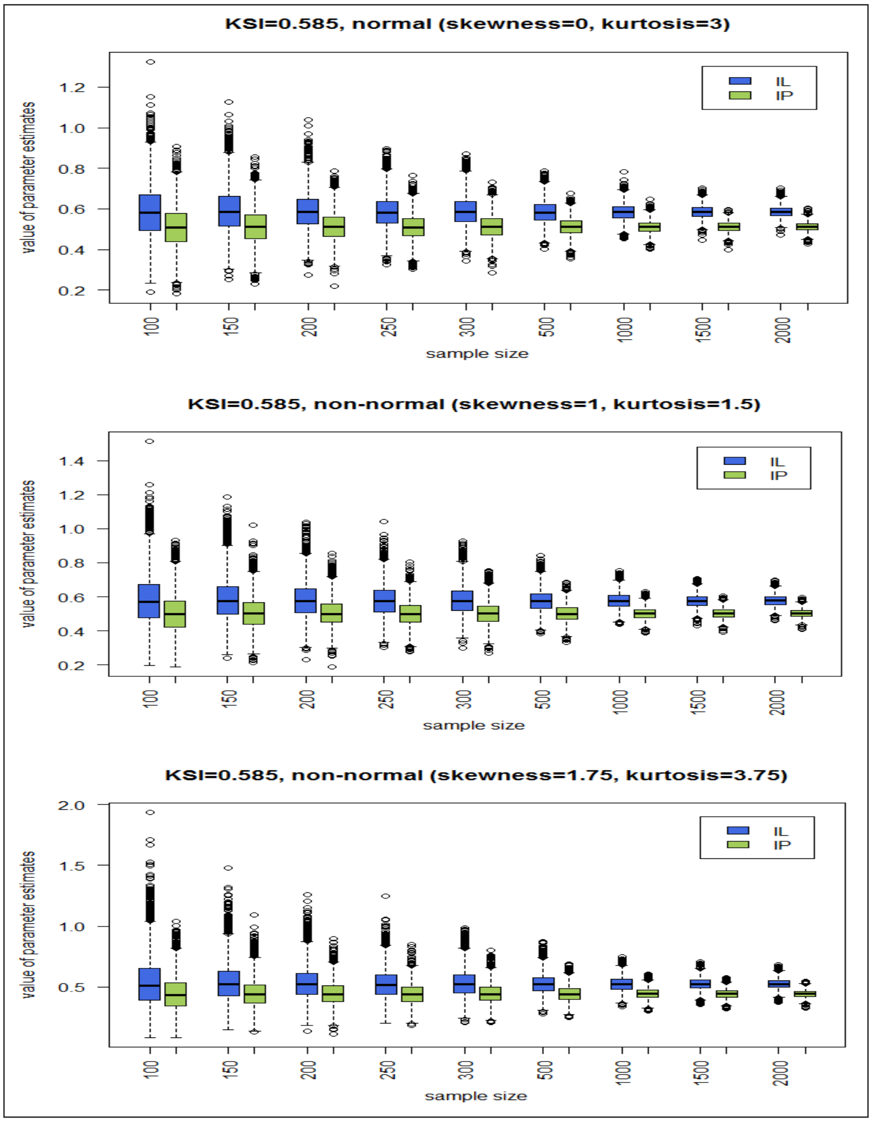

Figure 4. Box-plots for parameter estimates, $\hat{\beta}$ for item-level (IL) model and item-parcel (IP) model

estimates decreases as the sample size increases. The box-plots also show that the median of the parameter estimates is lower for the IP model across all sample sizes and types of distribution.

Table 3 shows the value of GoF measures for the IL model and IP model for different distributions of data. The results indicate that the model fit for all GoF measures are likely to improve with a higher sample size. The IP model has higher GoF measures for both normal and non-normal data across all sample sizes. When the data is normal, the RMSEA is lower for the IL model. However, the RMSEA is slightly lower for the IP model when data is non-normal. Since the AGFI and NFI are formed in a similar pattern with GFI and due to space constraint, the box-plots for AGFI and NFI are not presented.

Table 3
GoF measure values

n	Model	Distribution	Chi-sq	GFI	AGFI	RMSEA	NFI	TLI	CFI
100	IL	a	96.303	0.892	0.855	0.026	0.924	0.993	0.992
		b	114.962	0.874	0.830	0.050	0.908	0.974	0.977
		c	141.432	0.849	0.797	0.075	0.869	0.937	0.947
	IP	a	13.638	0.964	0.922	0.028	0.978	0.998	0.996
		b	15.643	0.959	0.911	0.040	0.974	0.993	0.993
		c	18.555	0.952	0.896	0.056	0.964	0.982	0.987
150	IL	a	93.981	0.926	0.900	0.018	0.949	0.997	0.995
		b	113.168	0.912	0.881	0.039	0.938	0.984	0.986
		c	139.329	0.894	0.857	0.060	0.909	0.959	0.965
	IP	a	13.510	0.976	0.947	0.022	0.985	0.999	0.997
		b	15.534	0.972	0.940	0.031	0.983	0.995	0.996
		c	18.401	0.967	0.929	0.045	0.975	0.988	0.992
200	IL	a	92.526	0.944	0.924	0.014	0.962	0.998	0.997
		b	112.095	0.932	0.909	0.033	0.953	0.988	0.990
		c	138.322	0.918	0.889	0.051	0.931	0.969	0.974
	IP	a	13.204	0.982	0.961	0.018	0.989	0.999	0.998
		b	15.341	0.979	0.955	0.027	0.987	0.997	0.997
		c	18.363	0.975	0.946	0.038	0.981	0.991	0.994
250	IL	a	91.776	0.954	0.939	0.012	0.970	0.999	0.998
		b	111.263	0.945	0.926	0.029	0.962	0.991	0.992
		c	137.066	0.933	0.910	0.045	0.944	0.976	0.980
	IP	a	13.357	0.985	0.968	0.016	0.991	0.999	0.999
		b	15.548	0.983	0.963	0.024	0.989	0.997	0.997
		c	18.385	0.980	0.956	0.035	0.985	0.993	0.995
300	IL	a	91.500	0.962	0.948	0.011	0.975	0.999	0.998
		b	111.178	0.954	0.938	0.026	0.968	0.992	0.993
		c	137.070	0.944	0.924	0.041	0.953	0.980	0.983
	IP	a	13.097	0.988	0.974	0.014	0.993	0.999	0.999
		b	15.338	0.986	0.969	0.022	0.991	0.998	0.998
		c	18.317	0.980	0.964	0.031	0.988	0.994	0.997
500	IL	a	90.113	0.977	0.969	0.008	0.985	0.999	0.999
		b	110.218	0.972	0.962	0.020	0.981	0.996	0.996
		c	136.344	0.965	0.953	0.032	0.972	0.988	0.990

Table 3 *(continue)*

n	Model	Distribution	Chi-sq	GFI	AGFI	RMSEA	NFI	TLI	CFI
500	IP	a	13.101	0.993	0.984	0.011	0.996	0.999	0.999
		b	15.338	0.991	0.981	0.017	0.995	0.999	0.999
		c	18.350	0.990	0.978	0.024	0.992	0.996	0.997
1000	IL	a	89.763	0.988	0.984	0.005	0.992	0.999	0.999
		b	109.711	0.986	0.981	0.014	0.990	0.998	0.998
		c	136.011	0.982	0.976	0.022	0.985	0.994	0.995
	IP	a	13.045	0.996	0.992	0.008	0.998	0.999	0.999
		b	15.290	0.996	0.991	0.012	0.997	0.999	0.999
		c	18.347	0.995	0.989	0.017	0.996	0.998	0.999
1500	IL	a	89.398	0.992	0.989	0.004	0.995	0.999	0.999
		b	109.525	0.990	0.987	0.011	0.994	0.999	0.999
		c	136.485	0.988	0.984	0.018	0.990	0.996	0.997
	IP	a	12.961	0.998	0.995	0.006	0.999	1.000	0.999
		b	15.211	0.997	0.994	0.010	0.998	0.999	0.999
		c	18.296	0.997	0.993	0.014	0.997	0.999	0.999
2000	IL	a	89.347	0.994	0.992	0.004	0.996	0.999	0.999
		b	109.559	0.993	0.990	0.010	0.995	0.999	0.999
		c	136.773	0.991	0.988	0.016	0.993	0.997	0.997
	IP	a	13.006	0.998	0.996	0.005	0.999	0.999	0.999
		b	15.269	0.998	0.995	0.008	0.999	0.999	0.999
		c	18.391	0.997	0.994	0.012	0.998	0.999	0.999

Note. n is sample size. IL is item-level model. IP is item-parcel model. [a]Normal (skewness=0, kurtosis=3). [b]Non-normal (skewness=1, kurtosis=1.5). [c]Non-normal (skewness=1.75, kurtosis=3.75). Chi-sq is the chi-squared. GFI is the goodness-of-fit Index. AGFI is the Adjusted goodness-of-fit index. RMSEA is the root mean square error of approximation. NFI is the Normed Fit Index. TLI is the Tucker-Lewis index. CFI is the comparative fit index.

The box-plots for GFI in Figure 5 show that GFI is affected by sample size, model structure (IL model and IP model), and distribution of data. The GFI improves as the sample size increases for both models and GFI is lower when data is non-normal. This finding is consistent the studies of Fan et al. (1999) and DoĞan and Özdamar (2017) but contradicts with the studies of Jöreskog and Sörbom (1982) and Bagozzi and Yi (1988) who found that sample size does not affect the values of GFI and AGFI. The box-plots also show that the median of GFI is higher for the IP model and the medians of both models are close to each other when the sample size is large. The dispersion is large for small sample sizes (n<500) for both models under normal and non-normal conditions.

The box-plots in Figure 6 show that RMSEA for the IL model and IP model are within the acceptable threshold level (RMSEA<0.06), indicating that both models fit the data well. RMSEA is affected by sample size, model structure (IL model and IP model) and distribution of data. The median of RMSEA is lowest for the IP model and it is close to the median of the IL model for a large sample size. The dispersion of RMSEA declines as the sample size increases under normal and non-normal conditions.

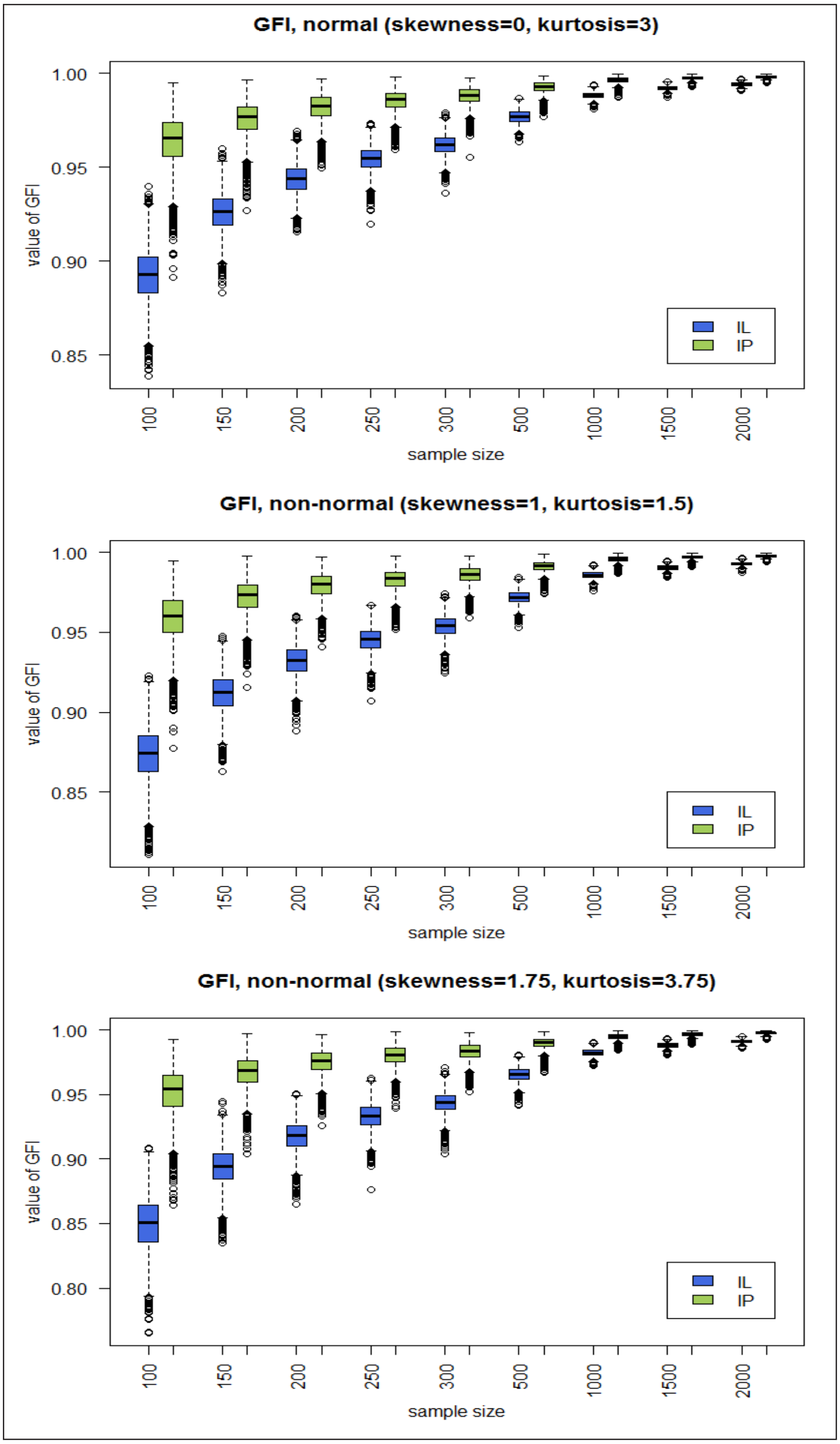

Figure 5. Box-plots for GFI for item-level (IL) model and item-parcel (IP) model

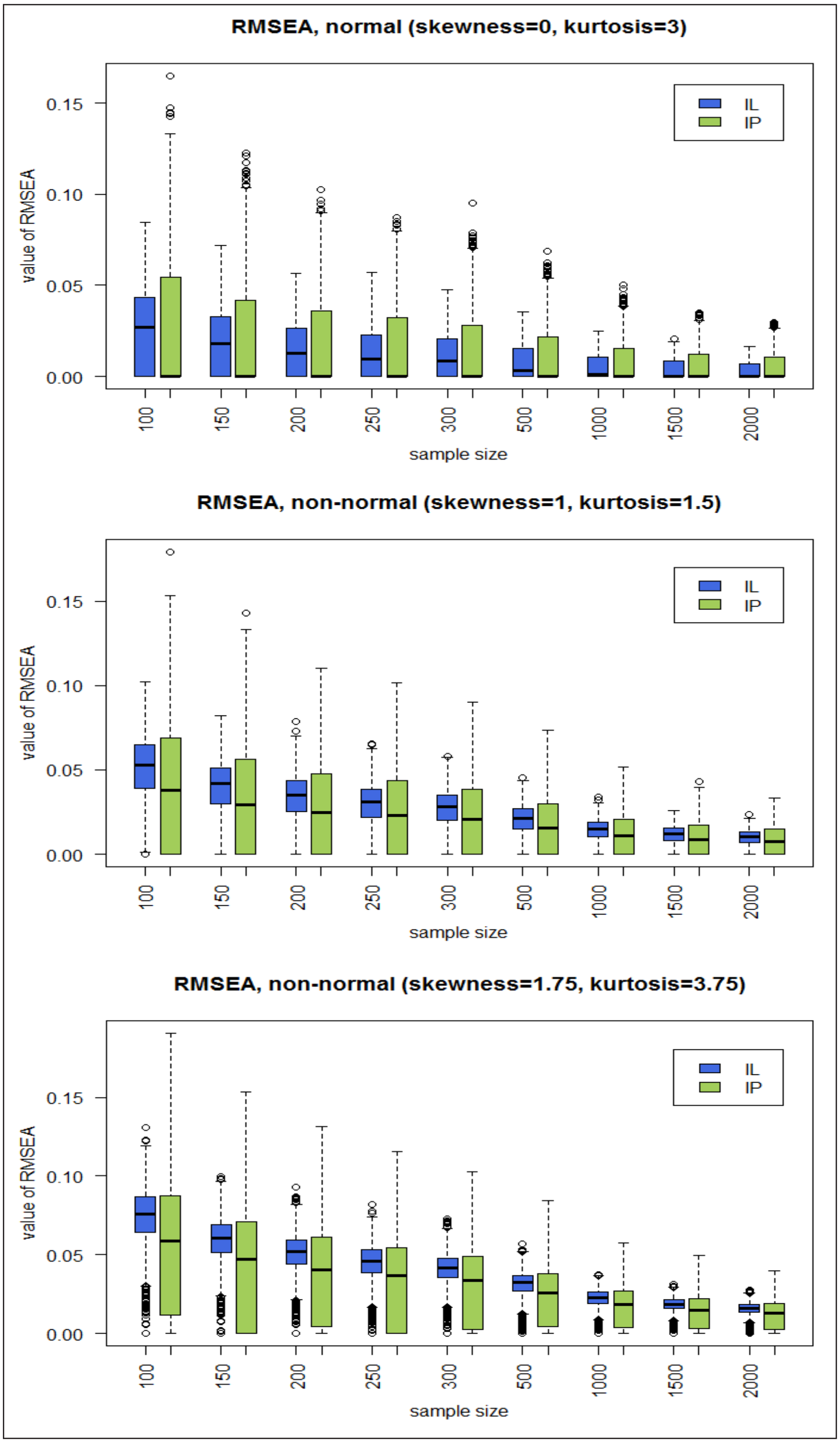

Figure 6. Box-plots for RMSEA for item-level (IL) model and item-parcel (IP) model

The box-plots in Figure 7 show that TLI is not influenced much by sample size and this finding is in line with the study of Fan et al. (1999). TLI for the IL model is more affected when data is non-normal and sample size is small (n = 100). The TLI for both models exceed the acceptable threshold level (>0.95) indicating that both models fit the data well. The box-plots show that the median of TLI is higher for the IP model and the medians of both models are close to each other for a large sample size. The dispersion is large for small sample sizes (n<500) for both models under normal and non-normal conditions.

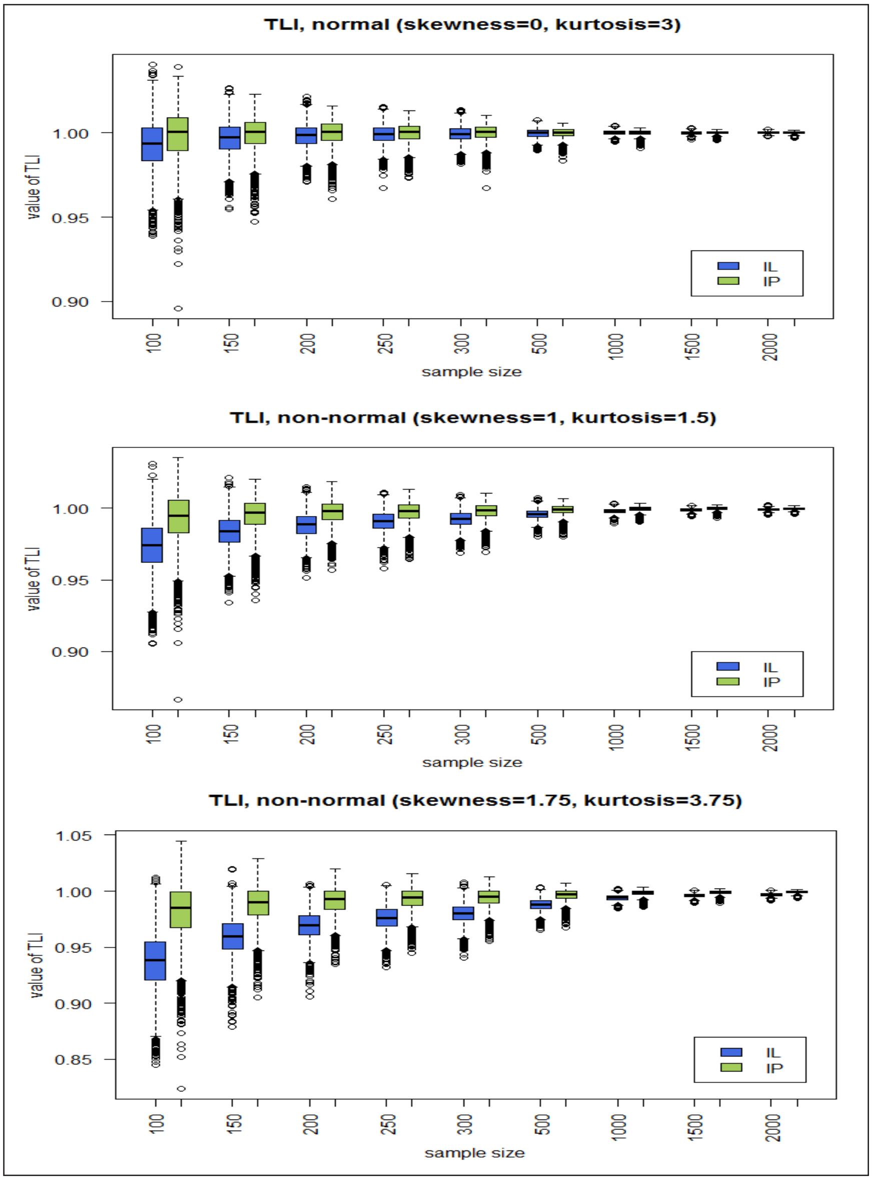

Figure 7. Box-plots for TLI for item-level (IL) model and item-parcel (IP) model

The box-plots in Figure 8 show that CFI is affected by model structure (IL model and IP model). The CFI also shows that the IL model and IP model have good model fit since the values exceed the acceptable threshold level (>0.95). The median of CFI is higher for the IP model compared to the IL model but for a large sample size, the medians for both models are close to each other. The dispersions are large for small sample sizes (n<500) for both models under normal and non-normal conditions.

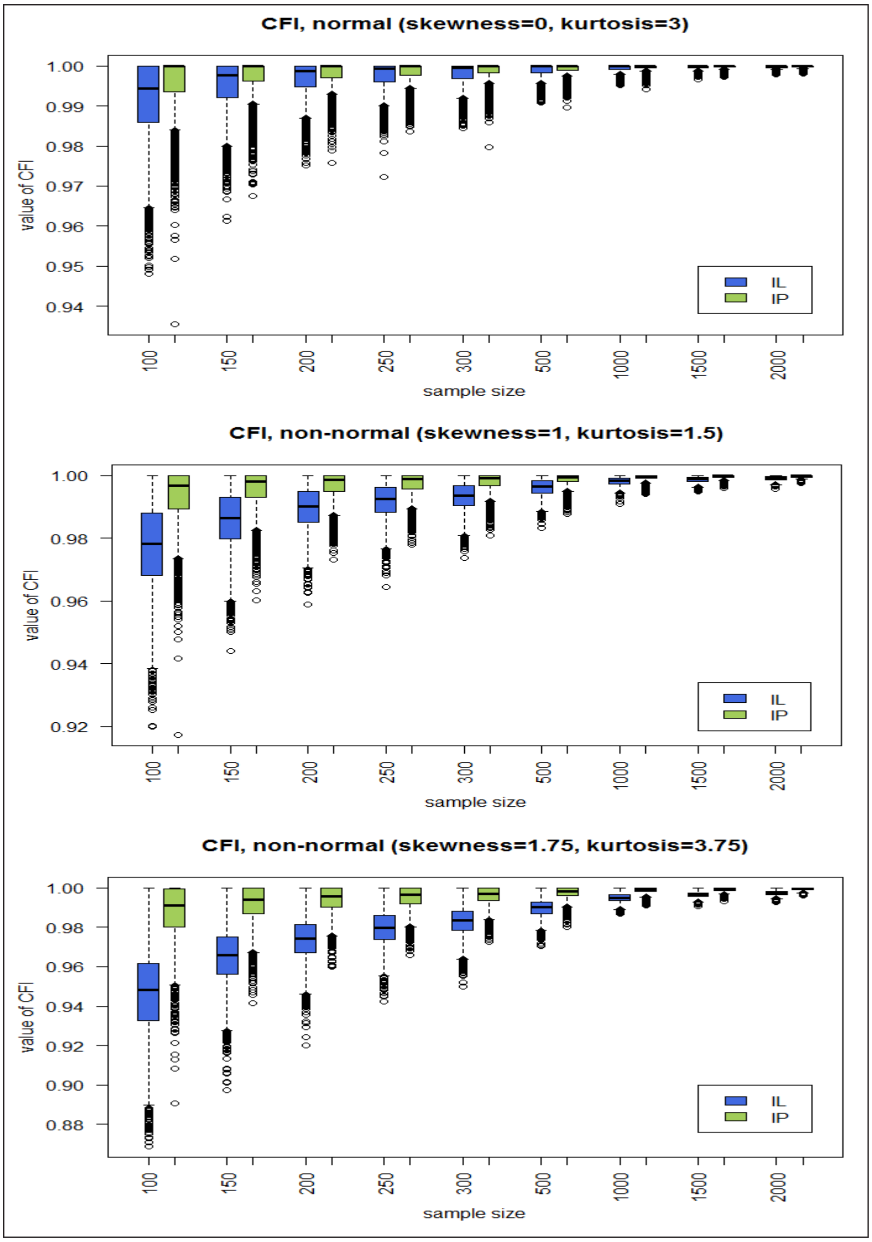

Figure 8. Box-plots for CFI for item-level (IL) model and item-parcel (IP) model

The simulation results and box-plots demonstrate that GoF measures improve with an increase in sample size and GoF measures are affected by model structure. In summary, based on parameter estimates and MSE, the IL model produces less biased parameter estimates than the IP model. However, the parameter estimates of the IL model are affected when data is severely non-normal. The summary of GoF measures presented in Table 4 shows that model fit for the IP model is better than the ILvel model under normal and non-normal conditions except for RMSEA under normal condition.

Table 4
Summary of GoF measures

	GFI		AGFI		NFI		TLI		CFI		RMSEA	
	N	NN	N	NN	N	NN	N	NN	N	NN	N	NN
Model 1: IL											/	
Model 2: IP	/	/	/	/	/	/	/	/	/	/		/

Note. N is normal. NN is non-normal. IL is item-level model. IP is item-parcel model. [/] represent the best model fit. GFI is the goodness-of-fit index. AGFI is the Adjusted goodness-of-fit index. RMSEA is the root mean square error of approximation. NFI is the Normed Fit Index. TLI is the Tucker-Lewis index. CFI is the comparative fit index.

Study 2: Results of Empirical Example

This section presents the results of parameter estimates and model fit for the IL model and IP model for the empirical example in Study 2. The results for parameter estimates, p-value and standard error for the original (IL) model and IP (Model 1, Model 2, Model 3) models are presented in Table 5. The parameter estimates for Model 1 are closer to the original model followed by Model 2 and Model 3. Based on p-values, the main paths are significant for all models. Among the three IP models, Model 1 has the lowest standard error while Model 3 has the highest standard errors for all paths.

Table 5
Parameter estimates, p-value and standard error of empirical example

	IL	IP		
		Model 1	Model 2	Model 3
PSOC <- PSI	-0.489*** (0.101)	-0.505*** (0.090)	-0.538*** (0.095)	-0.551*** (0.106)
QoL<- PSOC	0.357** (4.015)	0.347** (4.049)	0.370** (5.264)	0.286** (6.009)
QoL<- PSI	-0.390*** (2.292)	-0.393*** (2.059)	-0.412*** (2.619)	-0.491*** (3.962)

Note. ***p-value < 0.01.**p-value < 0.05. *p-value < 0.10. PSOC is parenting sense of competence. IL is item-level model. IP is item-parcel model. PSI is parenting stress index. HRQoL is health related quality of life. The value in ( ) represent standard error.

Table 6 shows the GoF measures of the empirical example. The results indicate that all the GoF measures are higher for the IP model. However, the value of RMSEA for Model 1 is closer to the original (IL) model.

The empirical findings support the simulation results in that item parceling can improve model fit but it can also produce biased parameter estimates. Based on this empirical example, it is not advisable to average all items in a construct to form an observed variable (Model 3). We recommend forming parcels for a model which consists of many indicators in a construct (six and above). Researchers should also consider to use at least three indicators per construct as suggested by Ding et al. (1995), Hair et al. (2015) and Iacobucci (2010).

Table 6
GoF measures of empirical examples

Model	Chi-sq(df)	GFI	AGFI	RMSEA	NFI	TLI	CFI
IL	1331.512 (767)	0.740	0.708	0.064	0.711	0.841	0.851
IP :							
Model 1	645.831 (366)	0.808	0.771	0.065	0.804	0.892	0.903
Model 2	368.245 (180)	0.848	0.804	0.076	0.815	0.877	0.076
Model 3	48.607 (11)	0.934	0.833	0.138	0.894	0.836	0.914

Note. GFI is the goodness-of-fit index. AGFI is the Adjusted goodness-of-fit index. RMSEA is the root mean square error of approximation. NFI is the Normed Fit Index. TLI is the Tucker-Lewis index. CFI is the comparative fit index. IL is item-level model. IP is item-parcel model.

CONCLUSION

This simulation study investigated the performance of parameter estimates and model fit based on different sample sizes, model structures and distributions of data in SEM. The simulation results indicate that the IL model produced biased parameter estimates under non-normal distribution of data. The IP model also produced biased parameter estimates for both normal and non-normal data regardless of sample size. In terms of model fit, the IP model is better than the IL model under normal and non-normal conditions except for RMSEA under normal condition. The empirical example provided evidence that the IP model estimates were close to the ILevel model but the minimum number of indicators per construct must be at least three to reduce the biasness of parameter estimates. Averaging all items of a construct is not recommended as it will produce highly biased estimates.

Several limitations should be noted in this study because it only used a random item parceling technique and only a simple structural model in the simulation design. In future work, the simulation study could be extended to investigate how different parceling techniques can reduce the biasness of the parameter estimates of a structural equation model for a more complex model.

ACKNOWLEDGEMENT

We would like to express our gratitude to Professor Emeritus Dale Louis Goodhue from Department of Management Information Systems, Terry College of Business, University of Georgia for his permission to use his model as a basis to generate the data in this study.

REFERENCES

Bagozzi, R. R., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*(1), 74-94.

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*(1), 78-102.

Bearden, W. O., Sharma, S., & Teel, J. E. (1982). Sample size effects on chi square and other statistics used in evaluating causal models. *Journal of Marketing Research*, *19*, 425-430.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238-246.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd Ed.). London, New York: Routledge.

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*(l), 16-29.

Deng, L., Yang, M., & Marcoulides, K. M. (2018). Structural equation modeling with many variables: A systematic review of issues and developments. *Frontiers in Psychology*, *9*(580),1-14.

Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling*, *2*(2), 119-144.

DoĞan, İ., & Özdamar, K. (2017). The effect of different data structures, sample sizes on model fit measures. *Communications in Statistics-Simulation and Computation*, *46*(9), 7525-7533.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, *6*(1), 56-83.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*(4), 521-532.

Goodhue, D. L., Lewis, W., & Thompson, R. (2012). Does PLS have advantages for small sample size or non-normal data? *MIS Quarterly*, *36*(3), 981-1001.

Hair Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2015). *Multivariate data analysis* (7th Ed.). London, New York: Pearson Prentice Hall.

Hall, R. J., Snell, A. F., & Foust, M. S. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods*, *2*(3), 233-256.

Hamzah, A. H., Yap, B. W., Xie, X. J., & Ong, S. H. (2017). Investigating the power of goodness of fit tests for multinomial logistic regression. *Communication in Statistics: Simulation and Computation*, *47*(4), 1039-1055.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods*, *6*(1), 53-60.

Iacobucci, D. (2010). Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology*, *20*, 90-98.

Jöreskog, K. G., & Sörbom, D. (1982). Recent developments in structural equation modeling. *Journal of Marketing Research*, *19*, 404-416.

Kim, S., & Hagtvet, K. A. (2003). The impact of misspecified item parceling on representing latent variables in covariance structure modeling: A simulation study. *Structural Equation Modeling*, *10*(1), 101-127.

Landis, R. S., Beal, D. J., & Tesluk, P. E. (2000). A comparison of approaches to forming composite measures in structural equation models. *Organizational Research Methods*, *3*(2), 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*(2), 151-173.

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, *18*(3), 285-300.

Lyhagen, J., & Kraus, K. (2013). The small sample performance of estimators of the standard errors of structural equation models. *Journal of Statistical Computation and Simulation*, *83*(3), 458-471.

MacCallum, R. C., & Tucker, L. R. (1991). Representing sources of error in the common-factor model: Implications for theory and practice. *Psychological Bulletin*, *109*(3), 502-511.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*(3), 391-410.

Morgan, G. B., Hodge, K. J., & Baggett, A. R. (2016). Latent profile analysis with nonnormal mixtures: A Monte Carlo examination of model selection using fit indices. *Computational Statistics and Data Analysis*, *93*, 146-161.

Mulaik, S. A., James, L. R., Alstine, J. V., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, *105*(3), 430-445.

Nasir, N. N. M. (2015). *Investigating stress, self-efficacy and quality of life of mothers with autistic children: A structural equation modelling (SEM) approach* (Unpublished Master's Thesis). Universiti Teknologi Mara, Shah Alam, Malaysia.

Nasser, F., & Wisenbaker, J. (2006). A monte carlo study investigating the impact of item parceling strategies on parameter estimates and their standard errors in CFA. *Structural Equation Modeling*, *13*(2), 204-228.

Olsson, U. H., Foss, T., & Breivik, E. (2004). Two equivalent discrepancy functions for maximum likelihood estimation: Do their test statistics follow a non-central chi-square distribution under model misspecification? *Sociological Methods and Research*, *32*(4), 453-500.

Olsson, U. H., Foss, T., Troye, S. V., & Howell, R. D. (2000). The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling*, *7*(4), 557-595.

Orcan, F. (2013). *Use of item parceling in structural equation modeling with missing data* (PhD Dissertation). Florida State University, Tallahassee, Florida.

Ory, D. T., & Mokhtarian, P. L. (2010). The impact of non-normality , sample size and estimation technique on goodness-of-fit measures in structural equation modeling : evidence from ten empirical models of travel behavior. *Quality and Quantity*, *44*(3), 427-445.

Rocha, C. M., & Chelladurai, P. (2012). Item parcels in structural equation modeling: An applied study in sport management. *International Journal of Psychology and Behavioral Sciences*, *2*(1), 46-53.

Ryu, E. (2011). Effects of skewness and kurtosis on normal-theory based maximum likelihood test statistic in multilevel structural equation modeling. *Behavior Research Methods*, *43*, 1066-1074.

Shah, R., & Goldstein, S. M. (2006). Use of structural equation modeling in operations management research: Looking back and forward. *Journal of Operations management*, *24*(2), 148-169.

Sharma, S., Mukherjee, S., Kumar, A., & Dillon, W. R. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research, 58*, 935-943.

Sterba, S. K., & Maccallum, R. C. (2010). Variability in parameter estimates and model fit across repeated allocations of items to parcels. *Multivariate Behavioral Research, 4,* 322-358.