Article

Educational and Psychological Measurement
2020, Vol. 80(1) 41–66

DOI: 10.1177/0013164419845039
journals.sagepub.com/home/epm
SAGE

Fitting Ordinal Factor Analysis Models With Missing Data: A Comparison Between Pairwise Deletion and Multiple Imputation

Dexin Shi[1], Taehun Lee[2], Amanda J. Fairchild[1] and Alberto Maydeu-Olivares[1,3]

Abstract

This study compares two missing data procedures in the context of ordinal factor analysis models: pairwise deletion (PD; the default setting in Mplus) and multiple imputation (MI). We examine which procedure demonstrates parameter estimates and model fit indices closer to those of complete data. The performance of PD and MI are compared under a wide range of conditions, including number of response categories, sample size, percent of missingness, and degree of model misfit. Results indicate that both PD and MI yield parameter estimates similar to those from analysis of complete data under conditions where the data are missing completely at random (MCAR). When the data are missing at random (MAR), PD parameter estimates are shown to be severely biased across parameter combinations in the study. When the percentage of missingness is less than 50%, MI yields parameter estimates that are similar to results from complete data. However, the fit indices (i.e., χ^2, RMSEA, and WRMR) yield estimates that suggested a worse fit than results observed in complete data. We recommend that applied researchers use MI when fitting ordinal factor models with missing data. We further recommend interpreting model fit based on the TLI and CFI incremental fit indices.

Keywords

ordinal factor analysis, missing data, pairwise deletion, multiple imputation

[1]University of South Carolina, Columbia, SC, USA
[2]Chung-Ang University, Seoul, Korea
[3]University of Barcelona, Barcelona, Spain

Corresponding Author:
Taehun Lee, Department of Psychology, Chung-Ang University, 84 Heukseok-Ro, Dongjak-Gu, Seoul 60974, Korea.
Email: lee.taehun@gmail.com

Factor analysis models have been widely applied in educational and psychological measurement (Jöreskog, 1969; McDonald, 1999; Raykov & Marcoulides, 2011). Traditional factor models assume a liner relationship between the latent factor and the continuously distributed observed variables. Most data in the field of psychology and education are ordinal in nature, however (e.g., correct/wrong in educational testing, Likert-type scales in psychological measurement, etc.). In principle, these measures should be treated as discrete. Strictly speaking, fitting a common factor analysis model to ordinal data introduces model misspecification, as the relationship between a discrete outcome and a continuous factor cannot be linear (Maydeu-Olivares, Cai, & Hernández, 2011). Under the structural equation modeling (SEM) framework, a better approach that takes into account the true nature of the data is to fit an ordinal factor analysis model. This is especially applicable when the data have fewer than five ordered categories (DiStefano & Morgan, 2014; Maydeu-Olivares, Fairchild, & Hall, 2017; Muthén & Kaplan, 1992; Rhemtulla, Brosseau-Liard, & Savalei, 2012).

Missing data are very likely to occur in many psychological or educational testing scenarios. For example, attitude surveys may be incomplete because respondents refuse to answer certain questions for fear that their anonymity will not be protected. Other respondents may not have a chance to answer all of the questions, due to time constraints. Moreover, in large-scale educational assessments, such as the National Assessment of Educational Progress, planned missing data designs such as matrix sampling are often used, such that some questions are never intended to be asked to reduce the burden on respondents (Graham, Taylor, Olchowski, & Cumsille, 2006; Kaplan, 1995). In Peng, Harwell, Liou, and Ehman's (2006) survey among 11 major education and psychological journals, 48% of the articles clearly involved missing data and approximately 16% of articles did not clearly report missing data information.

Methodological research has shed light on available tools to handle incomplete observations. Two major modern techniques for handling missing data—full information maximum likelihood (FIML; Allison, 1987; Arbuckle, 1996; Finkbeiner, 1979) and multiple imputation (MI; Rubin, 1976, 1987) —have been developed. Specifically, according to a typology for missing data mechanisms developed by Rubin (1976), FIML and MI have been shown to outperform traditional methods; they yield less biased and more efficient parameter estimates when the missing data mechanism is either missing completely at random (MCAR; the probability that a data value is missing does not depend on the observed or missing values) or missing at random (MAR; the probability that a data value is missing may depend on *other* observed variables, but not on the variable which is missing).[1] A number of studies have been conducted to evaluate the performance of FIML and MI and have compared them with other missing data techniques in estimating various statistical models in psychological studies (e.g., multiple regression, Enders, 2001; SEM, Enders & Bandalos, 2001; Allison, 2003; Olinsky, Chen, & Harlow, 2003).

When fitting ordinal factor analysis models, FIML is actually one of the most commonly used estimation methods (i.e., under the framework of item response models; Baker & Kim, 2004). By using FIML, missing data can be conveniently addressed in the same procedure used to estimate the model parameters. However, FIML is not always the ideal method for estimating the ordinal factor analysis model, as it must be performed by integrating over the latent traits (factors), making it computationally burdensome. The computational demands associated with FIML increase dramatically as the number of factors increase (Baker & Kim, 2004; Forero & Maydeu-Olivares, 2009; Kamata, & Bauer, 2008), making it unfeasible when the number of factors is large (e.g., ≥ 5; Forero & Maydeu-Olivares, 2009). Moreover, although goodness-of-fit indices for ordinal factor models under FIML have been developed (Maydeu-Olivares & Joe, 2005), they have not been routinely implemented in SEM software.[2]

To overcome the issues of applying FIML for the estimation of ordinal factor analysis models, the limited information approach based on polychoric correlations has been proposed and widely applied in practice (i.e., diagonally weighted least squares; Jöreskog & Sörbom, 1988; Muthén, du Toit, & Spisic, 1997). In these methods, parameters are estimated in several stages. First, thresholds and polychoric correlations are modeled using only univariate and bivariate information. Then, the factor analysis parameters are estimated from the first-stage estimates using one of several least squares methods. When ordinal factor analysis models are estimated with diagonally weighted least squares (i.e., the WLSMV estimator in Mplus), the default method for handling missing data in prevailing software packages is pairwise deletion (PD). The pairwise deletion approach attempts to mitigate the loss of data by using all available cases in the data analysis procedure, yielding a different subset of observations used to compute each element in the polychoric correlation matrix. As noted by Asparouhov and Muthén (2010a), the pairwise deletion approach only performs well when the data are MCAR; biased parameter estimates are observed under MAR.

To resolve the deficiencies of handling missing data under the more general MAR assumption, the MI method could be used. The MI approach involves three phases (Enders, 2010). First, in the imputation phase, missing observations are imputed a large number of times using stochastic regression to produce multiple ‘‘complete’’ data sets. Then the proposed models are estimated separately using each imputed data set in the analysis phase. Finally, in the pooling phase, the multiple sets of outcomes (i.e., parameter estimates) are combined into a single set of results. For fitting ordinal factor analysis models with missing data, the MI method followed by the limited information diagonally weighted least squares estimator is available in SEM software packages and has been recommended as the procedure could produce unbiased parameter estimates under either MCAR or MAR (Asparouhov & Muthén, 2010a, 2010b, 2010c).

Within the context of ordinal data, several strategies have been proposed for imputing missing values (see Jia, 2016, for a review). We consider the latent variable approach, as implemented in Mplus (Asparouhov & Muthén, 2010b, 2010c). Simulation studies have shown that this approach performs well when fitting regression models and ordinal factor analysis models, regardless of the number of response categories, sample size, proportion of missing data, or asymmetry of item distribution (Jia, 2016; Wu, Jia, & Enders, 2015). Findings from previous work in ordinal factor analysis are limited, however, in that they only consider *correctly* specified models. In practice, most models are to some degree *incorrect* or misspecified (Box, 1979; MacCallum, 2003; Shi, Maydeu-Olivares, & DiStefano, 2018). Guidelines based exclusively on correct model specification are fallible in the sense that those guidelines may not generalize to more realistic conditions where the model is wrong to some degree (i.e., in the presence of model misspecification). Thus, evaluating the performance of MI and PD in the presence of model misspecification remains an outstanding need. Related, previous work has only evaluated how different missing data procedures impart an influence on bias in *parameter estimation*. No existing research has examined how *goodness-of-fit* indices are impacted by the choice of missing data procedures across different missing data mechanisms.

In an effort to fill these gaps, we conduct a Monte Carlo simulation to compare the two most commonly used procedures in ordinal factor analysis: pairwise deletion, which is the *default* estimator Mplus, and multiple imputation. We compare goodness-of-fit indices across different percentages of missingness, as well as across different missing data mechanisms. We additionally consider bias in parameter estimation to validate findings from previous work. In so doing, we aim to provide holistic recommendations to applied SEM researchers on how to handle missing data when the outcomes are ordinal.

Monte Carlo Simulations

We conducted a simulation study to compare the performance of pairwise deletion and multiple imputation for fitting ordinal factor analysis models with missing data. We considered both correctly specified and misspecified model scenarios. The population model for the correctly specified scenario was a one-factor confirmatory factor analysis model with 12 observed variables (i.e., X1-X12), and the same one factor model was fitted to the data. For misspecified conditions, a confirmatory factor analysis model with two correlated factors was used as the population model, and a one-factor model was fitted to the data. Each factor was measured by six observed indicators (i.e., $f1$ by X1 to X6, and $f2$ by X7 to X12). The population factor variances were set to 1.0. We set the values for all factor loadings to 0.70 and all error variances to 0.51. For the two-factor models, the population correlation between the factors varied according to the different levels of model misspecification.

Based on the population model, complete data with a multivariate normal distribution were first generated. The continuous data were then discretized using a set of thresholds to create ordinal categories. Incomplete data were obtained from the complete data by introducing missing values for six items (i.e., X7-X12), according to different mechanisms and percentages of missingness, as described below.

The simulation conditions were obtained by manipulating five variables: number of response categories, sample size, percentage of missingness, mechanism of missingness, and level of model misspecification.

Number of Response Categories

Both binary ($c = 2$) or polytomous data ($c = 5$) were created by altering the values of the threshold. When the data were binary (0/1), the thresholds were set to 0.5, which implied that about 70% of the respondents were expected to endorse the binary items. For data with five categories, the thresholds were chosen so that the expected area under the curve was 7%, 24%, 38%, 24%, and 7% of the responses for ordered categories 0 through 4, respectively. The threshold values used were based on previous simulation studies in ordinal factor analysis (Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009; Rhemtulla et al., 2012; Shi, DiStefano, McDaniel, & Jiang, 2018).

Sample Size

Sample sizes included 200, 500, or 1,000. The levels were chosen to represent relatively small, medium, or large samples observed in social science research.

Percentage of Missingness

Three levels were considered for the percentage of missingness on each of the variables containing incomplete values (i.e., X7-X12): 15%, 25%, and 50%. The levels were manipulated to represent relatively small, medium, or large proportions of missing data, in line with previous simulation studies on missing data (Larsen, 2011; Newman, 2003; Prevosti & Chemisquy, 2010; Scheffer, 2002). The rate of missingness was set to be equal across the variables.

Mechanism of Missingness

Both MCAR and MAR were evaluated. In terms of MCAR, the occurrences of missingness have no correlation with any variables within the data set. For each complete data set, s% ($s = 15$, 25, or 50) of the X7 to X12 observations were randomly chosen and were set to be missing. For cases of MAR, the missingness of the six incomplete variables was determined by the percentile of the sum of the complete variables. If the sum scores were larger than its sth percentile, missing data were created by deleting the corresponding observations on X7 to X12. This approach used for generating missing data is consistent with that of previous studies (e.g., Zhang & Wang, 2013).

Level of Model Misspecification

The levels of model misspecification were manipulated by altering the sizes of the correlations between the latent factors in the population model. Five different correlations (1.0, 0.9, 0.8, 0.7, and 0.6) were included such that smaller correlations imply larger amounts of model misspecification. A correlation of 1.0 represents situations where the one factor model is correctly specified in the population. It is noted that we considered various levels of model misspecification that might occur in real data analysis. For example, when the population model is a two-factor model with an interfactor correlation of .90, most researchers would consider a one-factor model to be a close-fitting model. On the other hand, when the true model is a two-dimensional model with an interfactor correlation of .60, a one-factor model will demonstrate a poor fit to the data and thereby should be rejected.

In total, the number of conditions examined was 180 = 2 (number of response categories) × 3 (sample size levels) × 3 (percentages of missingness) × 2 (mechanisms of missingness) × 5 (levels of model misspecification). For each simulated condition, 500 replications of complete data sets were generated. Data generation and analyses were conducted with Mplus 7.4 (Muthén & Muthén, 1998-2010). SAS 9.4 was used to create missing values (according to the aforementioned missing data mechanisms) and to summarize the simulation results.

A unidimensional ordinal factor model was fit to each simulated data set using the robust diagonally weighted least squares estimator, with missing values addressed by PD and MI methods. For the MI conditions, the missing categorical observations were imputed using the latent variable approach implemented in Mplus (Asparouhov & Muthén, 2010a, 2010b). To achieve high efficiency and stabilize inference at the analysis phrase, we set the number of imputations to 100 (Graham, Olchowski, & Gilreath, 2007; Lu, 2017; Schafer & Olsen, 1998). In addition, the same ordinal factor analysis models were fitted using the same estimation method (i.e., WLSMV) to complete data, so direct comparisons could be made between results from applying the PD and MI methods with those that would be obtained from the complete data.

Specifically, we first focused on the average parameter estimates across replications for each parameter (i.e., the factor loadings and thresholds). In addition, we examined the performance of PD and MI in terms of estimating commonly used goodness-of-fit indices in ordinal factor analysis models, including the chi-square test statistic (χ^2), the root mean square error of approximation (RMSEA; Browne & Cudeck, 1993; Steiger, 1989, 1990), the comparative fit index (CFI; Bentler, 1990), the Tucker–Lewis index (TLI; Tucker & Lewis, 1973), and the weighted root mean square residual (WRMR; Yu, 2002; Yu & Muthén, 2002). The above goodness-of-fit indices have been routinely reported by applied researchers when evaluating model fit for ordinal factor analysis models (Garrido, Abad, & Ponsoda, 2016; Zhao, 2014).

It is noted that under MI, the procedures for combining model fit indices under ordinal factor analysis have yet to be developed. When fitting ordinal factor analysis models estimated with diagonally weighted least squares robust corrections (mean, or mean and variance adjustments) are applied to the chi-square test statistics

(Asparouhov & Muthén, 2010d; Muthén, 1993; Satorra & Bentler, 1994).[3] The existing pooling procedures (e.g., Meng and Rubin, 1992) are not applicable to chi-square test statistics with robust corrections from which most fit indices are derived. Currently, Mplus simply summarizes the results across imputations, and the average values of the fit indices are available to the users, which may be used as the naïve (average) estimate of the MI-based fit indices (Asparouhov & Muthén, 2010e).

To better measure the differences between the results from incomplete data analysis and complete data analysis, we computed the absolute relative difference (RD) for each outcome variables across all simulated conditions. The absolute RD is defined as

$$\text{Absolute RD} = \left| \frac{\bar{\hat{\theta}}_{\text{incomp}} - \bar{\hat{\theta}}_{\text{comp}}}{\bar{\hat{\theta}}_{\text{comp}}} \right|,$$

where $\bar{\hat{\theta}}_{\text{incomp}}$ represents the average values for the parameter estimates or the fit indices estimates across replications from incomplete data analysis using either PD or MI; $\bar{\hat{\theta}}_{\text{comp}}$ represents the corresponding average value from complete data analysis. Therefore, the RD evaluates the estimation bias due to missing data on a percentage scale in reference to the results from complete data analysis. The absolute RDs with smaller values suggest a smaller discrepancy between missing data and complete data analysis scenarios.

Results

The percentage of completed replications were more than 99% across all simulation conditions, except when ordinal data with five response categories were MAR, especially when the sample size was small (e.g., $N = 200$) and the percentage of missingness was large (i.e., 50%).[4] When summarizing the results, we excluded conditions with less than 50% of completed replications in this case.

Parameter Estimates

When fitting ordinal factor analysis models with binary data, one factor loading and one threshold were estimated for each item. The total number of estimated parameters was 24 (i.e., 2 model parameters × 12 items). For data with five response categories, one factor loading and four thresholds were estimated for each item. Therefore, in total, 60 parameters (i.e., 5 model parameters × 12 items) were estimated. For each estimated parameter, we computed the average point estimate obtained from PD and MI across replications and calculated the RD with reference to the result from the complete data analysis. To better compare and demonstrate patterns for each simulated condition, we summarized the results (RDs) for factor loading or threshold(s) across items selected based on whether missing data were present. That is, among the missing and complete blocks of items, we obtained the largest absolute RD for both factor loading and threshold. Tables 1 and 2 summarize the results for the simulation conditions under MCAR and MAR mechanisms, respectively. Consistent with previous studies, we considered RDs less than 0.10 (10%) acceptable (Muthén, Kaplan, & Hollis, 1987; L. K. Muthén & Muthén, 2002). Cases where absolute RD $>$ 0.10 are highlighted in bold.

As shown in Table 1, when the missing data mechanism was MCAR, the RDs were acceptable for both PD and MI across all simulated conditions. The results suggested that both PD and MI yielded average point estimates close to results from the complete data analyses. Under MAR, however, Table 2 demonstrates that PD produced absolute RDs larger than 10% across all simulated conditions if missing observations were present. This held true even when the percentage of missing data was small (i.e., 15%). For example, for binary data, the largest RD for factor loadings was 21.9% (correctly specified model, $N = 200$). The RDs for factor loadings were smaller with polytomous data; under the same conditions, the largest RD (for data with five categories) was 13.6%. In terms of the thresholds, similar patterns were observed for both binary data and data with five response categories. For example, when $N = 200$ and the model was correctly specified, the largest RDs were 43.4% and 45.6% for binary data and five-point ordinal data, respectively. The absolute RDs uniformly increased, as the percentage of missingness increased. Taking $\rho = .90$ and $N = 1{,}000$ as an example, as the percentage of missingness increased from 15% to 50%, the largest RDs for factor loadings of items with missing binary values increased from 21.5% to 72.2%.

On the other hand, when MI was applied, the RDs for parameter estimates with missing data were noticeably smaller than those obtained using PD. Generally speaking, the RDs for MI were acceptable if the percentage of missingness was less than 50%, regardless of sample size and the number of response categories. It is also noted that given a mixture of items with and without missing values, the parameter estimates for items with complete data were not affected by the missing values on the other items in the model, regardless of the mechanism and percentage of missing values. That is, as shown in both Tables 1 and 2, the RDs for parameters from the complete block of items were small across all simulated conditions.

Model Fit Indices

Across the simulated conditions, the average point estimates of the chi-square statistic, RMSEA, WRMR, CFI, and TLI were calculated for both PD and MI and compared with the results obtained from complete data analyses. The average estimates of each fit index for PD, MI, and the complete data analyses are summarized in Tables 3 to 7. In the tables, we have highlighted the results of PD and MI for cases where the absolute RD $>$ .10. It is also noted that as discussed earlier, when MI was applied, the fit indices were pooled by averaging the estimates across imputations.

As shown in Table 3, under MCAR, PD generally produced chi-square test statistics close to the complete data results if the fitted model was correctly specified.

Table 1. Relative Differences in Parameter Estimates: Missing Completely at Random (MCAR).

			CAT = 2								CAT = 5							
			Factor loadings				Thresholds				Factor loadings				Thresholds			
			Missing		Complete		Missing		Complete		Missing		Complete		Missing		Complete	
Rho	*N*	PM	PD	MI	PD	MI	PD	MI	PD	MI	PD	MI	PD	MI	PD	MI	PD	MI
1.0	200	15	0.4	0.7	0.3	0.2	0.3	1.1	0.0	0.1	0.3	0.3	0.1	0.1	0.9	0.9	0.0	0.0
		25	0.5	1.7	0.2	0.3	0.8	1.1	0.0	0.1	0.3	0.7	0.1	0.1	0.9	2.3	0.0	0.0
		50	1.8	3.7	0.4	0.8	1.7	2.8	0.0	0.1	0.8	1.9	0.2	0.4	1.8	3.7	0.1	0.1
	500	15	0.2	0.6	0.1	0.1	0.3	0.6	0.0	0.3	0.2	0.3	0.0	0.0	0.5	1.0	0.0	0.0
		25	0.4	1.0	0.1	0.1	1.0	0.8	0.0	0.3	0.2	0.3	0.1	0.1	0.6	1.3	0.0	0.0
		50	0.5	1.5	0.3	0.3	0.9	1.4	0.0	0.3	0.3	0.7	0.2	0.3	1.0	2.7	0.0	0.0
	1,000	15	0.2	0.5	0.1	0.1	0.3	0.4	0.0	0.2	0.1	0.1	0.0	0.0	0.2	0.5	0.0	0.0
		25	0.3	0.6	0.1	0.2	0.3	0.5	0.0	0.2	0.1	0.4	0.1	0.1	0.4	0.6	0.0	0.0
		50	0.4	1.5	0.2	0.2	0.4	1.2	0.0	0.2	0.2	0.6	0.1	0.2	1.0	1.9	0.0	0.0
0.9	200	15	0.1	0.3	0.0	0.1	0.3	0.4	0.0	0.3	0.2	0.3	0.5	0.2	0.6	0.8	0.0	0.0
		25	0.2	0.5	0.0	0.1	0.2	0.5	0.0	0.3	0.3	0.7	0.6	0.3	1.1	1.9	0.0	0.0
		50	0.3	1.1	0.2	0.1	0.4	1.0	0.0	0.3	0.8	2.0	1.3	0.4	2.0	4.7	0.2	0.2
	500	15	0.3	1.2	0.4	0.2	0.3	1.2	0.0	0.1	0.2	0.3	0.4	0.1	0.5	1.0	0.0	0.0
		25	0.4	1.8	0.8	0.3	0.5	1.0	0.0	0.1	0.2	0.3	0.6	0.1	0.5	1.1	0.0	0.0
		50	0.9	3.9	1.5	0.8	1.5	3.3	0.0	0.1	0.6	0.7	1.3	0.3	0.8	2.4	0.0	0.0
	1,000	15	0.4	0.3	0.4	0.2	0.6	0.5	0.0	0.3	0.2	0.2	0.3	0.1	0.3	0.5	0.0	0.0
		25	0.4	0.9	0.7	0.3	0.7	0.5	0.0	0.3	0.2	0.4	0.6	0.1	0.5	0.8	0.0	0.0
		50	1.0	2.4	1.3	0.4	1.0	1.0	0.0	0.3	0.5	0.8	1.2	0.1	0.7	1.7	0.0	0.0
0.8	200	15	0.2	0.4	0.4	0.1	0.2	0.3	0.0	0.2	0.4	0.4	0.7	0.1	1.1	1.2	0.0	0.0
		25	0.4	0.5	0.7	0.2	0.3	0.4	0.0	0.2	0.7	0.7	1.2	0.2	1.1	1.4	0.0	0.0
		50	0.4	1.4	1.3	0.2	0.5	1.5	0.0	0.2	0.6	1.4	2.5	0.3	1.6	5.4	0.2	0.2
	500	15	0.2	0.2	0.4	0.1	0.1	0.3	0.0	0.3	0.2	0.2	0.7	0.1	0.5	1.2	0.0	0.0
		25	0.3	0.5	0.7	0.1	0.4	0.5	0.0	0.3	0.5	0.1	1.2	0.1	0.6	1.2	0.0	0.0
		50	0.6	0.8	1.4	0.1	0.5	0.8	0.0	0.3	1.0	0.9	2.5	0.2	1.0	2.5	0.0	0.0
	1,000	15	0.3	1.0	0.8	0.3	0.8	1.4	0.0	0.1	0.3	0.1	0.7	0.1	0.3	0.5	0.0	0.0
		25	0.7	1.8	1.5	0.7	1.0	2.3	0.0	0.1	0.6	0.3	1.2	0.1	0.4	1.1	0.0	0.0
		50	0.9	4.5	2.8	1.4	1.3	2.9	0.0	0.1	1.2	0.4	2.5	0.2	0.7	1.3	0.0	0.0
0.7	200	15	0.4	0.7	0.8	0.3	0.5	0.8	0.0	0.3	0.5	0.2	1.1	0.1	0.6	1.4	0.0	0.0
		25	0.7	1.1	1.2	0.2	0.7	0.7	0.0	0.3	0.8	1.0	1.9	0.3	0.8	2.3	0.0	0.0
		50	0.5	2.3	2.6	0.6	0.9	1.1	0.0	0.3	1.3	1.9	3.5	0.4	1.4	4.1	0.1	0.1
	500	15	0.4	0.4	0.8	0.1	0.2	0.3	0.0	0.2	0.5	0.3	1.1	0.1	0.5	0.8	0.0	0.0
		25	0.7	0.6	1.3	0.2	0.4	0.4	0.0	0.2	0.8	0.3	1.8	0.1	1.0	1.0	0.0	0.0
		50	1.2	1.5	2.6	0.3	0.7	1.1	0.0	0.2	1.4	1.0	3.8	0.3	0.9	2.8	0.0	0.0
	1,000	15	0.3	0.3	0.7	0.1	0.1	0.4	0.0	0.3	0.5	0.1	1.1	0.0	0.4	0.5	0.0	0.0
		25	0.4	0.5	1.2	0.1	0.3	0.4	0.0	0.3	0.8	0.4	1.8	0.1	0.3	0.9	0.0	0.0
		50	1.1	0.9	2.6	0.2	0.6	0.9	0.0	0.3	1.7	0.7	3.8	0.1	0.7	1.3	0.0	0.0
0.6	200	15	0.3	1.4	1.2	0.4	0.7	1.5	0.0	0.1	0.7	0.4	1.6	0.1	0.6	1.5	0.0	0.0
		25	0.9	2.3	1.9	0.6	0.7	1.5	0.0	0.1	1.2	1.4	2.6	0.3	0.9	2.6	0.0	0.0
		50	1.1	5.8	3.9	1.9	1.0	3.6	0.0	0.1	2.4	1.1	5.3	0.3	2.2	5.8	0.2	0.2

(continued)

Table 1. (continued)

			CAT = 2								CAT = 5							
			Factor loadings				Thresholds				Factor loadings				Thresholds			
			Missing		Complete		Missing		Complete		Missing		Complete		Missing		Complete	
Rho	*N*	PM	PD	MI	PD	MI	PD	MI	PD	MI	PD	MI	PD	MI	PD	MI	PD	MI
	500	15	0.6	0.5	1.2	0.3	0.5	0.7	0.0	0.3	0.8	0.2	1.5	0.0	0.4	0.9	0.0	0.0
		25	0.9	1.0	2.0	0.1	0.5	0.6	0.0	0.3	1.2	0.5	2.6	0.0	0.9	1.3	0.0	0.0
		50	1.4	2.0	4.0	0.7	0.8	1.3	0.0	0.3	2.5	0.8	5.4	0.1	0.8	2.6	0.0	0.0
	1,000	15	0.6	0.5	1.2	0.1	0.2	0.3	0.0	0.2	0.7	0.1	1.6	0.1	0.3	0.6	0.0	0.0
		25	0.9	0.8	1.9	0.2	0.3	0.6	0.0	0.2	1.2	0.4	2.6	0.1	0.4	0.8	0.0	0.0
		50	1.7	1.4	3.9	0.3	0.4	1.5	0.0	0.2	2.7	0.7	5.3	0.2	0.9	1.8	0.0	0.0

Note. Rho = interfactor correlation; *N* = sample size; PM = percentage of missingness; CAT = number of categories; PD = pairwise deletion; MI = multiple imputation.

However, for misspecifed models, PD could produce downwardly biased chi-square test statistics under MCAR, especially when the percentage of missingness was high and the magnitude of misspecification was large. For example, when ρ = .60, N = 1,000, and 50% of the binary observations were missing, the average PD-based chi-square statistics was 240.43, which was far smaller than the chi-square test statistic obtained from complete data for the same effect size condition (418.50; absolute RD = 42.5%). Interestingly, under MAR, across all simulated conditions, the PD-based chi-square statistics were noticeably larger than those from complete data analyses, even the percentage of missingness was small (i.e., 15%).

On the other hand, the MI-based chi-square statistics were generally upwardly biased, especially when a substantial proportion of data were missing under MAR, and the level of model misspecification was low. For example, when the model was correctly specified (i.e., ρ = 1.00), N = 1,000, number of categories = 2, and percentage of missingness was 50%, under MAR, the average chi-square statistics using MI was 341.01, which was far bigger than the average chi-square test statistics obtained from complete data (54.06, RD = 530%). The bias decreased as the number of response categories increased. Keeping the same conditions with the above example (i.e., correctly specified model, N = 1,000, and 50% of missingness), as the number of categories increased to five, the absolute RD decreased to 155%.

The behaviors of RMSEA and WRMR are summarized in Tables 4 and 5, respectively. Under MCAR, the PD-based RMSEA and WRMR were slightly downwardly biased with reference to the results from complete data analyses. In addition, the biases were considerable when the percentage of missingness was large (e.g., 50%) and the level of model misspecification was more severe. For example, when ρ = .60 and N = 1,000, for binary data, the average point estimates for RMSEA and WRMR with complete data were .08 and 2.02; if 50% of the observations were MCAR, the

Table 2. Relative Differences in Parameter Estimates: Missing Completely at Random (MAR).

			CAT = 2								CAT = 5							
			Factor loadings				Thresholds				Factor loadings				Thresholds			
			Missing		Complete		Missing		Complete		Missing		Complete		Missing		Complete	
Rho	*N*	PM	PD	MI	PD	MI	PD	MI	PD	MI	PD	MI	PD	MI	PD	MI	PD	MI
1.0	200	15	**21.9**	2.5	3.7	0.5	43.4	3.4	0.0	0.1	**13.6**	0.3	2.0	0.2	**45.2**	1.8	0.4	0.4
		25	**33.8**	3.9	3.4	1.2	**69.1**	6.3	0.0	0.1	**24.7**	1.2	7.3	0.2	**70.3**	3.8	1.7	1.9
		50	**78.2**	**26.3**	1.0	1.1	**129.4**	**26.9**	0.0	0.1	—	—	—	—	—	—	—	—
	500	15	**21.7**	3.6	3.5	0.5	**42.4**	4.8	0.0	0.3	**13.6**	0.8	2.0	0.1	**44.8**	1.6	0.0	0.0
		25	**34.2**	5.5	3.0	1.1	**68.0**	8.5	0.0	0.3	**21.1**	0.7	2.7	0.1	**71.2**	2.7	0.1	0.1
		50	**75.6**	3.0	0.8	3.3	**128.1**	**18.1**	0.0	0.3	—	—	—	—	—	—	—	—
	1,000	15	**21.8**	3.3	3.4	0.4	**42.0**	4.0	0.0	0.2	**13.7**	0.6	1.9	0.0	**44.3**	1.4	0.0	0.0
		25	**34.5**	7.3	2.8	1.0	**67.0**	**10.2**	0.0	0.2	**21.1**	0.6	2.6	0.1	**70.1**	2.4	0.0	0.0
		50	**72.4**	**15.6**	0.6	3.2	**126.7**	**36.1**	0.0	0.2	**48.8**	1.4	4.6	0.2	**136.1**	7.5	0.5	0.5
0.9	200	15	**21.5**	3.8	3.3	0.4	**42.1**	4.5	0.0	0.3	**13.0**	0.4	1.9	0.3	**39.9**	1.8	0.2	0.2
		25	**34.3**	6.8	2.9	1.0	**66.9**	9.8	0.0	0.3	**20.5**	1.1	3.4	0.3	**61.1**	3.4	0.8	0.8
		50	**72.2**	**16.2**	0.6	3.2	**126.9**	**37.5**	0.0	0.3	—	—	—	—	—	—	—	—
	500	15	**19.4**	2.7	2.7	0.7	**37.9**	3.8	0.0	0.1	**12.6**	0.8	1.3	0.1	**39.1**	1.2	0.0	0.0
		25	**29.5**	3.7	2.2	1.4	**59.8**	6.4	0.0	0.1	**19.6**	0.8	1.7	0.2	**61.5**	2.3	0.0	0.0
		50	**74.6**	**24.8**	1.8	1.1	**111.0**	**19.1**	0.0	0.1	**49.8**	2.1	6.9	0.6	**116.8**	**11.2**	0.8	0.9
	1,000	15	**19.3**	3.8	2.6	0.6	**37.0**	5.1	0.0	0.3	**12.7**	0.6	1.3	0.1	**38.9**	1.2	0.0	0.0
		25	**30.3**	5.7	1.8	1.3	**58.8**	8.9	0.0	0.3	**19.7**	0.8	1.6	0.1	**60.5**	2.6	0.0	0.0
		50	**71.6**	3.0	2.1	4.3	**110.4**	**21.0**	0.0	0.3	**47.0**	1.7	0.4	0.3	**116.6**	7.9	0.1	0.1
0.8	200	15	**19.6**	3.3	2.4	0.5	**36.9**	4.0	0.0	0.2	**11.3**	0.5	0.8	0.3	**34.5**	1.6	0.1	0.1
		25	**30.7**	7.6	1.6	1.1	**58.0**	**10.7**	0.0	0.2	**18.0**	0.9	1.5	0.3	**52.8**	2.5	0.6	0.5
		50	**69.1**	**15.8**	2.0	4.0	**109.0**	**36.5**	0.0	0.2	—	—	—	—	—	—	—	—
	500	15	**19.2**	3.9	2.4	0.5	**36.8**	4.7	0.0	0.3	**11.3**	1.0	0.7	0.1	**33.9**	1.5	0.0	0.0
		25	**30.5**	6.9	1.6	1.1	**58.0**	**10.0**	0.0	0.3	**17.9**	1.1	0.7	0.2	**52.9**	1.9	0.0	0.0
		50	**69.1**	**16.4**	2.1	3.9	**109.0**	**37.4**	0.0	0.3	**46.4**	2.3	0.6	0.6	**100.4**	**10.2**	0.5	0.6
	1,000	15	**16.8**	2.8	2.0	0.9	**32.6**	4.1	0.0	0.1	**11.5**	0.6	0.6	0.1	**33.7**	1.0	0.0	0.0
		25	**25.5**	3.3	1.1	1.7	**51.4**	6.2	0.0	0.1	**18.0**	0.9	0.6	0.1	**52.1**	2.6	0.0	0.0
		50	**71.1**	**26.0**	3.8	0.4	**94.8**	**15.4**	0.0	0.1	**46.0**	1.7	2.1	0.4	**99.6**	7.6	0.0	0.0
0.7	200	15	**17.0**	3.7	1.7	0.7	**32.2**	5.0	0.0	0.3	**10.3**	0.5	0.6	0.3	**29.5**	1.5	0.1	0.1
		25	**26.7**	5.3	0.6	1.5	**50.8**	8.6	0.0	0.3	**16.3**	1.3	0.6	0.3	**45.1**	3.1	0.3	0.3
		50	**68.0**	5.5	4.2	5.5	**94.4**	**22.0**	0.0	0.3	**45.4**	3.8	6.5	1.0	**82.1**	**13.3**	1.9	2.0
	500	15	**17.3**	3.1	1.5	0.6	**32.0**	3.9	0.0	0.2	10.0	1.1	0.2	0.1	**29.0**	1.7	0.0	0.0
		25	**27.0**	7.3	0.3	1.2	**50.1**	**10.2**	0.0	0.2	**16.1**	1.2	0.4	0.2	**44.8**	1.7	0.0	0.0
		50	**65.5**	**15.3**	4.3	4.8	**93.3**	**36.4**	0.0	0.2	**44.8**	1.9	3.4	0.7	**84.7**	9.3	0.2	0.3
	1000	15	**16.9**	3.8	1.5	0.5	**32.0**	4.6	0.0	0.3	**10.1**	0.6	0.2	0.1	**28.7**	0.9	0.0	0.0
		25	**26.8**	6.7	0.3	1.2	**50.1**	9.7	0.0	0.3	**16.2**	1.0	0.5	0.1	**44.1**	2.4	0.0	0.0
		50	**65.5**	**16.0**	4.4	4.4	**93.4**	**36.6**	0.0	0.3	**45.4**	1.7	4.2	0.4	**83.9**	6.9	0.0	0.0
0.6	200	15	**14.4**	2.4	1.1	0.9	**27.9**	4.0	0.0	0.1	8.5	0.6	0.8	0.4	**25.0**	1.6	0.1	0.1
		25	**21.8**	3.0	0.1	2.0	**43.8**	6.0	0.0	0.1	**14.3**	1.8	0.7	0.2	**38.1**	3.1	0.2	0.2
		50	**67.1**	**24.6**	5.6	2.1	**80.4**	**11.0**	0.0	0.1	**41.0**	5.1	1.4	1.1	**67.7**	**10.2**	1.2	1.2
	500	15	**14.8**	3.5	0.8	0.7	**27.5**	5.0	0.0	0.3	8.6	1.1	0.9	0.1	**24.3**	1.5	0.0	0.0
		25	**23.1**	4.8	0.7	1.6	**43.1**	8.0	0.0	0.3	**14.3**	1.4	1.5	0.2	**37.4**	1.7	0.0	0.0
		50	**63.8**	8.1	6.3	6.1	**79.5**	**22.7**	0.0	0.3	**44.0**	1.6	5.9	0.7	**70.5**	8.9	0.2	0.2
	1,000	15	**21.9**	2.5	3.7	0.5	**43.4**	3.4	0.0	0.1	8.7	0.6	1.0	0.1	**24.1**	0.8	0.0	0.0
		25	**33.8**	3.9	3.4	1.2	**69.1**	6.3	0.0	0.1	**14.3**	0.9	1.6	0.2	**36.8**	2.1	0.0	0.0
		50	**78.2**	**26.3**	1.0	1.1	**129.4**	**26.9**	0.0	0.1	**44.8**	1.7	6.3	0.4	**69.8**	6.5	0.0	0.0

Note. Rho = interfactor correlation; *N* = sample size; PM = percentage of missingness; CAT = number of categories; PD = pairwise deletion; MI = multiple imputation. Conditions with less than 50% of completed replications were excluded from the table.

Table 3. The Behavior of the Chi-Square Test Statistics Under Missing Data.

			CAT = 2					**CAT = 5**				
				MCAR		MAR			MCAR		MAR	
Rho	*N*	PM	C	PD	MI	PD	MI	C	PD	MI	PD	MI
1.0	200	15	55.05	55.37	67.35	**68.88**	**68.28**	55.07	55.18	**68.39**	**68.91**	**66.61**
		25	55.05	55.38	**76.86**	**74.94**	**90.35**	55.07	55.53	**80.45**	**88.24**	**78.94**
		50	55.05	55.23	**113.40**	**77.99**	**213.54**	55.07	55.51	**127.45**	—	—
	500	15	54.41	54.38	**67.94**	**92.61**	**69.60**	54.84	55.04	**69.57**	**89.96**	**67.74**
		25	54.41	54.18	**79.56**	**107.57**	**96.95**	54.84	54.83	**81.66**	**135.59**	**80.87**
		50	54.41	54.61	**128.65**	**107.64**	**307.50**	54.84	55.41	**135.31**	—	—
	1,000	15	54.06	53.73	**67.78**	**133.66**	**70.55**	54.44	54.32	**69.30**	**127.30**	**68.00**
		25	54.06	54.25	**80.76**	**163.66**	**100.33**	54.44	54.85	**82.90**	**219.33**	**81.89**
		50	54.06	53.84	**132.10**	**145.20**	**341.01**	54.44	55.01	**136.53**	**404.51**	**139.43**
0.9	200	15	60.23	59.57	**72.32**	**84.82**	**73.86**	71.72	68.98	**84.01**	**105.82**	**82.86**
		25	60.23	58.81	**81.85**	**90.82**	**95.82**	71.72	67.78	**96.83**	**132.83**	**97.24**
		50	60.23	58.19	**120.62**	**92.91**	**219.36**	71.72	**63.54**	**140.68**	—	—
	500	15	68.38	66.55	**82.52**	**135.24**	**79.99**	99.28	92.71	**112.43**	**188.28**	**112.74**
		25	68.38	64.84	**92.91**	**150.63**	**103.20**	99.28	**87.45**	**123.03**	**252.41**	**124.86**
		50	68.38	**60.82**	**141.80**	**141.36**	**306.04**	99.28	**76.78**	**173.26**	**357.96**	**173.12**
	1,000	15	83.35	79.47	**97.59**	**222.88**	**92.06**	147.82	134.70	161.60	**331.14**	161.11
		25	83.35	76.69	**110.40**	**253.86**	**112.56**	147.82	**124.06**	**172.42**	**459.53**	**171.19**
		50	83.35	**69.24**	**160.83**	**218.06**	**327.78**	147.82	**102.36**	**224.56**	**651.37**	**218.56**
0.8	200	15	73.86	70.99	**85.63**	**103.65**	**87.23**	114.09	106.21	**125.93**	**156.33**	124.68
		25	73.86	69.37	**95.58**	**108.81**	**109.56**	114.09	**99.05**	**137.27**	**185.20**	**141.18**
		50	73.86	**63.96**	**132.98**	**108.71**	**231.42**	114.09	**83.33**	**177.27**	—	—
	500	15	105.66	98.27	**119.64**	**186.08**	112.91	213.38	188.65	223.26	**321.02**	227.13
		25	105.66	**94.81**	**131.39**	**198.01**	**131.54**	213.38	**175.40**	**235.98**	**389.10**	**238.50**
		50	105.66	**79.88**	**178.87**	**179.84**	**325.27**	213.38	**133.24**	**279.67**	**484.30**	**273.50**
	1,000	15	162.32	147.63	176.64	**328.88**	161.22	384.34	**339.00**	397.51	**603.57**	396.47
		25	162.32	**136.41**	**188.33**	**353.31**	171.14	384.34	**303.03**	404.66	**739.25**	400.67
		50	162.32	**110.07**	**238.11**	**299.82**	**351.32**	384.34	**220.82**	**450.06**	**901.62**	**433.80**
0.7	200	15	94.70	88.32	**106.48**	**124.86**	**107.90**	173.54	**155.82**	183.15	**213.85**	183.10
		25	94.70	85.58	**116.63**	**127.92**	**130.91**	173.54	**144.08**	**195.37**	**241.31**	**201.84**
		50	94.70	**74.58**	**155.69**	**124.98**	**243.57**	173.54	**114.89**	**230.98**	**291.69**	**245.11**
	500	15	161.98	146.91	176.48	**241.96**	164.78	367.81	322.12	376.55	**467.29**	381.71
		25	161.98	**136.69**	**185.39**	**249.02**	**180.26**	367.81	**292.55**	387.28	**530.76**	392.98
		50	161.98	**108.53**	**233.87**	**221.09**	**361.74**	367.81	**208.59**	**421.72**	**605.15**	**416.66**
	1,000	15	277.85	**246.25**	290.90	**442.82**	267.77	697.44	**604.56**	707.24	**897.64**	708.30
		25	277.85	**225.70**	304.56	**456.01**	269.12	697.44	**538.08**	715.56	**1023.89**	708.16
		50	277.85	**170.89**	**352.62**	**384.39**	**412.88**	697.44	**375.91**	755.13	**1145.17**	730.40
0.6	200	15	120.72	111.09	**132.87**	**147.12**	**133.69**	242.10	**214.15**	250.97	**272.86**	250.15
		25	120.72	**104.95**	**143.01**	**147.06**	**157.14**	242.10	**196.21**	263.09	**296.20**	**271.53**
		50	120.72	**87.85**	**180.72**	**140.74**	**256.83**	242.10	**148.89**	**293.68**	**335.85**	**313.91**
	500	15	230.94	**205.33**	243.80	**300.40**	231.07	540.42	**469.23**	549.17	**614.19**	553.47
		25	230.94	**189.00**	252.92	**299.36**	244.43	540.42	**419.29**	553.56	**667.91**	565.83
		50	230.94	**143.94**	**300.88**	**261.48**	**406.80**	540.42	**294.86**	588.39	**716.40**	581.47
	1,000	15	418.50	**368.70**	431.39	**562.21**	403.10	1042.96	**897.08**	1048.33	**1191.35**	1052.61
		25	418.50	**332.53**	443.40	**559.26**	397.99	1042.96	**802.83**	1062.52	**1298.68**	1049.10
		50	418.50	**240.43**	**484.38**	**468.18**	**508.70**	1042.96	**551.09**	1093.32	**1371.51**	1062.77

Note. Rho = interfactor correlation; *N* = sample size; PM = percentage of missingness; C = complete data results; PD = pairwise deletion results; MI = multiple imputation results; CAT = number of categories; MCAR = missing completely at random; MAR = missing at random. Conditions with less than 50% of completed replications were excluded from the table.

Table 4. The Behavior of RMSEA Under Missing Data.

			CAT = 2					CAT = 5				
				MCAR		MAR			MCAR		MAR	
Rho	*N*	PM	C	PD	MI	PD	MI	C	PD	MI	PD	MI
1.0	200	15	0.01	0.01	**0.03**	**0.03**	**0.03**	0.01	0.01	**0.03**	**0.03**	**0.03**
		25	0.01	0.01	**0.04**	**0.04**	**0.06**	0.01	0.01	**0.05**	**0.06**	**0.05**
		50	0.01	0.01	**0.07**	**0.04**	**0.12**	0.01	0.01	**0.08**	—	—
	500	15	0.01	0.01	**0.02**	**0.04**	**0.02**	0.01	0.01	**0.02**	**0.04**	**0.02**
		25	0.01	0.01	**0.03**	**0.04**	**0.04**	0.01	0.01	**0.03**	**0.05**	**0.03**
		50	0.01	0.01	**0.05**	**0.04**	**0.09**	0.01	0.01	**0.05**	—	—
	1,000	15	0.01	0.01	0.01	**0.04**	**0.02**	0.01	0.01	**0.02**	**0.04**	0.01
		25	0.01	0.01	**0.02**	**0.05**	**0.03**	0.01	0.01	**0.02**	**0.06**	**0.02**
		50	0.01	0.01	**0.04**	**0.04**	**0.07**	0.01	0.01	**0.04**	**0.08**	**0.04**
0.9	200	15	0.02	0.02	**0.04**	**0.05**	**0.04**	0.04	**0.03**	**0.05**	**0.07**	**0.05**
		25	0.02	0.02	**0.05**	**0.06**	**0.06**	0.04	**0.03**	**0.06**	**0.08**	**0.06**
		50	0.02	0.02	**0.08**	**0.06**	**0.12**	0.04	**0.03**	**0.09**	—	—
	500	15	0.02	0.02	**0.03**	**0.05**	**0.03**	0.04	0.04	**0.05**	**0.07**	**0.05**
		25	0.02	0.02	**0.04**	**0.06**	**0.04**	0.04	**0.03**	**0.05**	**0.09**	**0.05**
		50	0.02	0.01	**0.06**	**0.06**	**0.09**	0.04	**0.03**	**0.07**	**0.11**	**0.07**
	1,000	15	0.02	0.02	**0.03**	**0.06**	**0.03**	0.04	0.04	0.04	**0.07**	0.04
		25	0.02	0.02	**0.03**	**0.06**	**0.03**	0.04	0.04	**0.05**	**0.09**	**0.05**
		50	0.02	0.02	**0.04**	**0.06**	**0.07**	0.04	**0.03**	**0.06**	**0.11**	**0.05**
0.8	200	15	0.04	0.04	**0.05**	**0.07**	**0.05**	0.07	0.07	**0.08**	**0.10**	**0.08**
		25	0.04	**0.03**	**0.06**	**0.07**	**0.07**	0.07	**0.06**	**0.09**	**0.11**	**0.09**
		50	0.04	**0.03**	**0.08**	**0.07**	**0.12**	0.07	**0.05**	**0.11**	—	—
	500	15	0.04	0.04	**0.05**	**0.07**	**0.05**	0.08	**0.07**	0.08	**0.10**	0.08
		25	0.04	0.04	**0.05**	**0.07**	**0.05**	0.08	**0.07**	0.08	**0.11**	0.08
		50	0.04	**0.03**	**0.07**	**0.07**	**0.10**	0.08	**0.05**	**0.09**	**0.13**	**0.09**
	1,000	15	0.04	0.04	**0.05**	**0.07**	0.04	0.08	**0.07**	0.08	**0.10**	0.08
		25	0.04	0.04	**0.05**	**0.07**	**0.05**	0.08	**0.07**	0.08	**0.11**	0.08
		50	0.04	**0.03**	**0.06**	**0.07**	**0.07**	0.08	**0.06**	**0.09**	**0.13**	0.08
0.7	200	15	0.06	**0.05**	**0.07**	**0.08**	**0.07**	0.10	0.10	0.11	**0.12**	0.11
		25	0.06	**0.05**	**0.07**	**0.08**	**0.08**	0.10	0.09	0.11	**0.13**	0.11
		50	0.06	**0.04**	**0.10**	**0.08**	**0.13**	0.10	**0.07**	**0.13**	**0.15**	**0.13**
	500	15	0.06	0.06	**0.07**	**0.08**	0.06	0.11	0.10	0.11	0.12	0.11
		25	0.06	**0.05**	**0.07**	**0.09**	**0.07**	0.11	**0.09**	0.11	**0.13**	0.11
		50	0.06	**0.04**	**0.08**	**0.08**	**0.10**	0.11	**0.08**	0.12	**0.14**	0.11
	1,000	15	0.06	0.06	**0.07**	**0.09**	0.06	0.11	0.10	0.11	**0.13**	0.11
		25	0.06	0.06	**0.07**	**0.09**	0.06	0.11	0.09	0.11	**0.13**	0.11
		50	0.06	**0.05**	**0.07**	**0.08**	**0.08**	0.11	**0.08**	0.11	**0.14**	0.11
0.6	200	15	0.08	**0.07**	0.08	**0.09**	0.08	0.13	0.12	0.13	0.14	0.13
		25	0.08	**0.07**	**0.09**	**0.09**	**0.09**	0.13	0.11	0.14	**0.15**	0.14
		50	0.08	**0.05**	**0.11**	**0.09**	**0.13**	0.13	**0.09**	**0.15**	**0.16**	**0.15**
	500	15	0.08	**0.07**	0.08	**0.10**	0.08	0.13	0.12	0.13	0.14	0.14
		25	0.08	**0.07**	**0.09**	**0.10**	0.08	0.13	0.12	0.14	**0.15**	0.14
		50	0.08	**0.06**	**0.09**	**0.09**	**0.11**	0.13	**0.09**	0.14	**0.16**	0.14

(continued)

Table 4. (continued)

			CAT = 2					CAT = 5				
				MCAR		MAR			MCAR		MAR	
Rho	*N*	PM	C	PD	MI	PD	MI	C	PD	MI	PD	MI
	1,000	15	0.08	0.08	0.08	**0.10**	0.08	0.14	0.13	0.14	0.15	0.14
		25	0.08	0.07	0.08	**0.10**	0.08	0.14	**0.12**	0.14	0.15	0.14
		50	0.08	**0.06**	**0.09**	**0.09**	**0.09**	0.14	**0.10**	0.14	**0.16**	0.14

Note. Rho = interfactor correlation; *N* = sample size; PM = percentage of missingness; C = complete data results; PD = pairwise deletion results; MI = multiple imputation results; CAT = number of categories; RMSEA = root mean square error of approximation; MCAR = missing completely at random; MAR = missing at random. Conditions with less than 50% of completed replications were excluded from the table.

Table 5. The Behavior of WRMR Under Missing Data.

			CAT = 2					CAT = 5				
				MCAR		MAR			MCAR		MAR	
Rho	*N*	PM	C	PD	MI	PD	MI	C	PD	MI	PD	MI
1.0	200	15	0.66	0.66	**0.74**	**0.85**	**0.74**	0.50	0.51	**0.57**	**0.62**	**0.56**
		25	0.66	0.67	**0.8**	**0.94**	**0.87**	0.50	0.51	**0.63**	**0.74**	**0.62**
		50	0.66	0.68	**1.02**	**1.05**	**1.53**	0.50	0.53	**0.82**	—	—
	500	15	0.65	0.65	**0.73**	**0.98**	**0.74**	0.49	0.50	**0.56**	**0.70**	**0.55**
		25	0.65	0.66	**0.8**	**1.13**	**0.87**	0.49	0.50	**0.61**	**0.90**	**0.61**
		50	0.65	0.67	**1.04**	**1.24**	**1.66**	0.49	0.52	**0.80**	—	—
	1,000	15	0.65	0.65	**0.73**	**1.17**	**0.74**	0.49	0.49	0.55	**0.83**	**0.55**
		25	0.65	0.66	**0.8**	**1.38**	**0.87**	0.49	0.50	**0.61**	**1.14**	**0.60**
		50	0.65	0.66	**1.03**	**1.42**	**1.64**	0.49	0.51	**0.79**	**1.74**	**0.79**
0.9	200	15	0.70	0.7	**0.78**	**0.96**	**0.79**	0.59	0.59	0.65	**0.80**	0.64
		25	0.70	0.7	**0.84**	**1.06**	**0.91**	0.59	0.59	**0.71**	**0.94**	**0.70**
		50	0.70	0.71	**1.07**	**1.17**	**1.55**	0.59	0.58	**0.88**	—	—
	500	15	0.75	0.74	**0.83**	**1.2**	0.8	0.69	0.67	0.73	**1.05**	0.74
		25	0.75	0.73	**0.88**	**1.35**	**0.91**	0.69	0.66	**0.77**	**1.27**	**0.78**
		50	0.75	0.72	**1.1**	**1.44**	**1.66**	0.69	0.63	**0.93**	**1.69**	**0.92**
	1,000	15	0.82	0.81	0.89	**1.53**	0.86	0.83	0.81	0.87	**1.37**	0.87
		25	0.82	0.8	**0.95**	**1.74**	**0.94**	0.83	0.78	0.90	**1.69**	0.90
		50	0.82	0.77	**1.15**	**1.76**	**1.62**	0.83	**0.73**	**1.04**	**2.25**	**1.01**
0.8	200	15	0.80	0.79	0.88	**1.08**	0.88	0.79	0.77	0.83	**1.01**	0.83
		25	0.80	0.79	**0.94**	**1.17**	**1**	0.79	0.75	**0.88**	**1.15**	**0.89**
		50	0.80	0.77	**1.14**	**1.28**	**1.6**	0.79	**0.70**	**1.02**	—	—
	500	15	0.96	0.93	1.02	**1.43**	0.98	1.06	1.01	1.08	**1.42**	1.10
		25	0.96	0.92	**1.07**	**1.56**	1.05	1.06	0.98	1.12	**1.62**	1.12
		50	0.96	**0.85**	**1.27**	**1.63**	**1.73**	1.06	**0.88**	**1.23**	**2.01**	**1.20**

(continued)

Table 5. (continued)

			CAT = 2					CAT = 5				
				MCAR		MAR			MCAR		MAR	
Rho	*N*	PM	C	PD	MI	PD	MI	C	PD	MI	PD	MI
	1,000	15	1.18	1.14	1.23	**1.89**	1.16	1.41	1.34	1.43	**1.92**	1.43
		25	1.18	1.1	1.27	**2.07**	1.18	1.41	1.28	1.45	**2.21**	1.44
		50	1.18	**1**	**1.44**	**2.08**	**1.7**	1.41	**1.13**	1.53	**2.70**	1.49
0.7	200	15	0.94	0.91	1.01	**1.21**	1.01	1.02	0.98	1.06	**1.23**	1.05
		25	0.94	0.91	**1.06**	**1.28**	**1.12**	1.02	0.95	1.10	**1.35**	1.12
		50	0.94	0.85	**1.27**	**1.39**	**1.65**	1.02	**0.87**	**1.22**	**1.62**	**1.24**
	500	15	1.23	1.18	1.28	**1.65**	1.22	1.47	1.39	1.48	**1.78**	1.50
		25	1.23	1.14	1.32	**1.76**	1.27	1.47	1.34	1.51	**1.96**	1.52
		50	1.23	**1.03**	**1.5**	**1.82**	**1.85**	1.47	**1.16**	1.58	**2.29**	1.56
	1,000	15	1.60	1.52	1.63	**2.22**	1.55	2.00	1.89	2.02	**2.44**	2.02
		25	1.60	1.46	1.67	**2.36**	1.53	2.00	1.81	2.03	**2.69**	2.02
		50	1.60	**1.3**	**1.81**	**2.37**	**1.87**	2.00	**1.56**	2.09	**3.11**	2.04
0.6	200	15	1.10	1.06	1.17	**1.33**	1.16	1.28	1.21	1.30	**1.44**	1.30
		25	1.10	1.04	**1.22**	**1.39**	**1.27**	1.28	1.17	1.34	**1.54**	1.36
		50	1.10	**0.96**	**1.4**	**1.47**	**1.7**	1.28	**1.05**	**1.44**	**1.77**	**1.47**
	500	15	1.51	1.44	1.55	**1.87**	1.49	1.88	1.78	1.90	**2.12**	1.91
		25	1.51	1.39	1.58	**1.95**	1.53	1.88	1.70	1.90	**2.27**	1.93
		50	1.51	**1.23**	**1.75**	**1.98**	**1.99**	1.88	**1.46**	1.97	**2.53**	1.94
	1,000	15	2.02	1.92	2.05	**2.53**	1.96	2.59	2.44	2.60	**2.93**	2.61
		25	2.02	1.83	2.08	**2.63**	1.92	2.59	**2.33**	2.62	**3.13**	2.60
		50	2.02	**1.59**	2.18	**2.62**	2.13	2.59	**1.99**	2.67	**3.47**	2.60

Note. Rho = interfactor correlation; *N* = sample size; PM = percentage of missingness; C = complete data results; PD = pairwise deletion results; MI = multiple imputation results; CAT = number of categories; WRMR = weighted root mean square residual; MCAR = missing completely at random; MAR = missing at random. Conditions with less than 50% of completed replications were excluded from the table.

average PD-based RMSEA and WRMR were .06 and 1.59, respectively. In contrast, when the missing data mechanism was MAR, regardless of the number of response categories, the PD-based RMSEA and WRMR were positively biased relative to the complete data results. Under the same condition described above (i.e., $\rho = .60$, N = 1,000 and binary data), if 50% of the observations were missing at random (MAR), the average PD-based RMSEA and WRMR were .09 and 2.62, respectively, which were bigger than the counterparts obtained using complete data.

The MI-based RMSEA and WRMR were upwardly biased, regardless of the missing data mechanism, suggesting that the model fit was worse than it should be if the researchers had access to complete data. Similar patterns were observed for both binary data and polytomous data with five categories. It was also noted that the RDs for MI were generally larger than those obtained using PD, especially when the level of model misspecification was minor, the sample size was small, and a large proportion of data were missing under MAR. For example, when $N = 200$ and 50% of the binary observations were missing under MAR, the average MI-based RMSEA and WRMR for correctly specified models were .12 and 1.53, respectively. According to the commonly used cutoffs (i.e., RMSEA $<$ .06, Hu & Bentler, 1998, 1999; WRMR $<$ 1.00, DiStefano, Liu, Jiang, & Shi, 2018), researchers would mistakenly conclude that the correctly specified model fits the data poorly.

Tables 6 and 7 demonstrate results for the CFI and TLI, respectively. As shown in the tables, similar patterns were observed for both binary data and polytomous data with five categories. Specifically, under MCAR, the PD-based CFI and TLI were close to the complete data results. Under MAR, the CFI and TLI using PD could be noticeably underestimated, however, especially when percentage of missingness was high, or the level of model misspecification became more severe. For example, when ρ = .60 and N = 200, the average point estimates for CFI and TLI using complete data were 0.90 and 0.88; given that 50% of the binary data were missing under MAR, the average CFI and TLI dropped to 0.75 and 0.70 when PD was employed. Alternatively, when MI was used, the average CFI and TLI were generally close to the complete data results, regardless of the missing data mechanism considered in the current study. That is, the RDs were acceptable for MI-based CFI and TLI across all simulated conditions, except for a few conditions where the sample size was small (i.e., N = 200) and the percentage of missingness reached 50%.

Discussion and Conclusions

This study compared pairwise deletion and multiple imputation in the context of ordinal factor analysis models with missing data. We investigated which procedure tends to show parameter estimates and model fit indices closer to those from analyses of complete data. Results show that when the data are MCAR, both PD and MI yield parameter estimates similar to those from analysis of complete data, regardless of the number of response categories, sample size, percentage of missingness, and level of model misspecification. However, when the data are MAR, the PD parameter estimates could be severely different from those obtained from the complete data analysis, especially as the percentage of missingness increases. The MI procedure yielded parameter estimates that were similar to the results using complete data, unless the percentage of missingness reached 50%. The findings regarding the parameter estimates using PD are consistent with the conclusions drawn by Asparouhov and Muthén (2010a, 2010c) for correctly specified models. As noted in the technical report, the diagonally weighted least squares (DWLS) estimation with listwise deletion is a special case of the DWLS estimation with PD. Therefore, it is not surprising to observe biased parameter estimates under MAR.

Our study additionally looked at the behaviors of fit indices under ordinal factor analysis models with missing data. When using PD, the fit indices were generally very similar to results from complete data analyses under conditions of MCAR, unless the percentage of missingness was large (i.e., 50%) and/or the model misspecification was severe. Under MAR, the goodness-of-fit indices were also biased, in

Table 6. The Behavior of CFI Under Missing Data.

			CAT = 2					CAT = 5				
				MCAR		MAR			MCAR		MAR	
Rho	*N*	PM	C	PD	MI	PD	MI	C	PD	MI	PD	MI
1.0	200	15	1.00	1.00	0.99	0.97	0.99	1.00	1.00	0.99	0.99	0.99
		25	1.00	1.00	0.98	0.95	0.97	1.00	1.00	0.99	0.97	0.99
		50	1.00	0.99	0.95	0.91	0.89	1.00	1.00	0.97	—	—
	500	15	1.00	1.00	0.99	0.97	0.99	1.00	1.00	1.00	0.99	1.00
		25	1.00	1.00	0.99	0.95	0.99	1.00	1.00	0.99	0.98	0.99
		50	1.00	1.00	0.97	0.93	0.94	1.00	1.00	0.98	—	—
	1,000	15	1.00	1.00	1.00	0.97	1.00	1.00	1.00	1.00	0.99	1.00
		25	1.00	1.00	0.99	0.95	0.99	1.00	1.00	1.00	0.97	1.00
		50	1.00	1.00	0.98	0.95	0.97	1.00	1.00	0.99	0.93	0.99
0.9	200	15	0.99	0.99	0.98	0.94	0.98	0.99	0.99	0.98	0.96	0.98
		25	0.99	0.99	0.97	0.92	0.96	0.99	0.99	0.98	0.94	0.98
		50	0.99	0.99	0.93	**0.87**	**0.88**	0.99	0.99	0.96	—	—
	500	15	0.99	0.99	0.99	0.94	0.99	0.99	0.99	0.99	0.96	0.99
		25	0.99	0.99	0.98	0.92	0.98	0.99	0.99	0.99	0.94	0.98
		50	0.99	0.99	0.96	**0.89**	0.93	0.99	0.99	0.97	**0.87**	0.98
	1,000	15	0.99	0.99	0.99	0.94	0.99	0.99	0.99	0.99	0.96	0.99
		25	0.99	0.99	0.99	0.92	0.99	0.99	0.99	0.99	0.93	0.99
		50	0.99	0.99	0.98	0.91	0.96	0.99	0.99	0.98	**0.88**	0.98
0.8	200	15	0.97	0.98	0.96	0.91	0.96	0.96	0.97	0.96	0.92	0.96
		25	0.97	0.98	0.95	0.88	0.93	0.96	0.97	0.95	0.89	0.95
		50	0.97	0.98	0.91	**0.82**	**0.86**	0.96	0.98	0.93	—	—
	500	15	0.97	0.98	0.97	0.90	0.97	0.96	0.96	0.96	0.92	0.96
		25	0.97	0.98	0.96	0.88	0.97	0.96	0.97	0.96	0.89	0.95
		50	0.97	0.98	0.94	**0.85**	0.91	0.96	0.97	0.95	**0.82**	0.95
	1,000	15	0.97	0.97	0.97	0.90	0.97	0.96	0.96	0.96	0.92	0.96
		25	0.97	0.98	0.97	**0.87**	0.97	0.96	0.96	0.96	0.88	0.96
		50	0.97	0.98	0.96	**0.86**	0.95	0.96	0.97	0.95	**0.82**	0.95
0.7	200	15	0.94	0.95	0.93	0.87	0.93	0.92	0.93	0.92	0.87	0.92
		25	0.94	0.95	0.92	0.84	0.90	0.92	0.93	0.91	0.84	0.90
		50	0.94	0.96	0.88	**0.79**	**0.84**	0.92	0.95	0.89	**0.76**	0.88
	500	15	0.94	0.94	0.93	0.86	0.94	0.92	0.92	0.91	0.87	0.91
		25	0.94	0.95	0.93	**0.83**	0.94	0.92	0.93	0.91	0.83	0.91
		50	0.94	0.96	0.91	**0.81**	0.89	0.92	0.94	0.90	**0.77**	0.91
	1,000	15	0.94	0.94	0.93	0.86	0.94	0.91	0.92	0.91	0.86	0.91
		25	0.94	0.94	0.93	**0.83**	0.95	0.91	0.92	0.91	0.83	0.91
		50	0.94	0.95	0.92	**0.82**	0.93	0.91	0.94	0.91	**0.77**	0.91
0.6	200	15	0.90	0.90	0.88	0.82	0.88	0.87	0.88	0.86	0.82	0.86
		25	0.90	0.91	0.87	**0.80**	0.85	0.87	0.89	0.86	0.79	0.85
		50	0.90	0.93	0.84	**0.75**	0.83	0.87	0.91	0.84	**0.71**	0.82
	500	15	0.89	0.90	0.89	0.82	0.90	0.86	0.87	0.86	0.82	0.85
		25	0.89	0.90	0.88	**0.79**	0.89	0.86	0.88	0.86	0.78	0.85
		50	0.89	0.92	0.86	**0.78**	0.86	0.86	0.90	0.85	**0.72**	0.85

(continued)

Table 6. (continued)

			CAT = 2					CAT = 5				
				MCAR		MAR			MCAR		MAR	
Rho	*N*	PM	C	PD	MI	PD	MI	C	PD	MI	PD	MI
	1,000	15	0.89	0.90	0.89	0.81	0.90	0.86	0.87	0.85	0.81	0.85
		25	0.89	0.90	0.88	**0.79**	0.91	0.86	0.87	0.85	0.78	0.86
		50	0.89	0.92	0.87	**0.79**	0.90	0.86	0.90	0.85	**0.73**	0.86

Note. Rho = interfactor correlation; *N* = sample size; PM = percentage of missingness; C = complete data results; PD = pairwise deletion results; MI = multiple imputation results; CAT = number of categories; CFI = comparative fit index; MCAR = missing completely at random; MAR = missing at random. Conditions with less than 50% of completed replications were excluded from the table.

Table 7. The Behavior of TLI under Missing Data.

			CAT = 2					CAT = 5				
				MCAR		MAR			MCAR		MAR	
Rho	*N*	PM	C	PD	MI	PD	MI	C	PD	MI	PD	MI
1.0	200	15	1.00	1.00	0.98	0.97	0.98	1.00	1.00	0.99	0.99	0.99
		25	1.00	1.00	0.97	0.94	0.96	1.00	1.00	0.99	0.97	0.99
		50	1.00	1.00	0.94	**0.89**	0.86	1.00	1.00	0.96	—	—
	500	15	1.00	1.00	0.99	0.97	0.99	1.00	1.00	1.00	0.99	1.00
		25	1.00	1.00	0.99	0.94	0.98	1.00	1.00	0.99	0.97	0.99
		50	1.00	1.00	0.97	0.92	0.92	1.00	1.00	0.98	—	—
	1,000	15	1.00	1.00	1.00	0.97	1.00	1.00	1.00	1.00	0.99	1.00
		25	1.00	1.00	0.99	0.94	0.99	1.00	1.00	1.00	0.97	1.00
		50	1.00	1.00	0.98	0.93	0.96	1.00	1.00	0.99	0.91	0.99
0.9	200	15	0.99	0.99	0.98	0.93	0.97	0.99	0.99	0.98	0.96	0.98
		25	0.99	0.99	0.96	0.90	0.95	0.99	0.99	0.97	0.94	0.97
		50	0.99	0.99	0.92	**0.84**	**0.85**	0.99	0.99	0.95	—	—
	500	15	0.99	0.99	0.98	0.93	0.99	0.99	0.99	0.98	0.95	0.98
		25	0.99	0.99	0.98	0.90	0.98	0.99	0.99	0.98	0.92	0.98
		50	0.99	0.99	0.95	**0.87**	0.91	0.99	0.99	0.97	**0.87**	0.97
	1,000	15	0.99	0.99	0.99	0.93	0.99	0.99	0.99	0.99	0.95	0.99
		25	0.99	0.99	0.98	0.90	0.99	0.99	0.99	0.98	0.92	0.98
		50	0.99	0.99	0.97	**0.88**	0.96	0.99	0.99	0.98	**0.85**	0.98
0.8	200	15	0.97	0.97	0.95	0.89	0.95	0.96	0.96	0.95	0.91	0.95
		25	0.97	0.97	0.94	**0.85**	0.92	0.96	0.96	0.94	**0.89**	0.94
		50	0.97	0.98	0.89	**0.79**	**0.83**	0.96	0.97	0.92	—	—
	500	15	0.97	0.97	0.96	0.88	0.97	0.95	0.96	0.95	0.90	0.95
		25	0.97	0.97	0.95	**0.85**	0.96	0.95	0.96	0.95	0.86	0.94
		50	0.97	0.98	0.93	**0.82**	0.89	0.95	0.97	0.93	**0.78**	0.94

(continued)

Table 7. (continued)

			CAT = 2					CAT = 5				
				MCAR		MAR			MCAR		MAR	
Rho	*N*	PM	C	PD	MI	PD	MI	C	PD	MI	PD	MI
	1,000	15	0.97	0.97	0.96	0.88	0.97	0.95	0.95	0.95	0.90	0.95
		25	0.97	0.97	0.96	**0.85**	0.97	0.95	0.96	0.95	0.86	0.95
		50	0.97	0.97	0.95	**0.83**	0.94	0.95	0.96	0.94	**0.78**	0.94
0.7	200	15	0.93	0.94	0.91	0.84	0.91	0.91	0.91	0.90	0.85	0.90
		25	0.93	0.94	0.90	0.80	0.88	0.91	0.92	0.89	**0.80**	0.88
		50	0.93	0.95	0.85	**0.74**	**0.81**	0.91	0.93	0.87	**0.76**	0.85
	500	15	0.93	0.93	0.92	**0.83**	0.93	0.90	0.91	0.90	0.84	0.89
		25	0.93	0.93	0.91	**0.80**	0.92	0.90	0.91	0.89	**0.80**	0.89
		50	0.93	0.95	0.89	**0.77**	0.86	0.90	0.93	0.88	**0.72**	0.88
	1,000	15	0.92	0.93	0.92	0.83	0.93	0.89	0.90	0.89	0.83	0.89
		25	0.92	0.93	0.92	**0.80**	0.94	0.89	0.91	0.89	**0.79**	0.89
		50	0.92	0.94	0.90	**0.78**	0.92	0.89	0.93	0.89	**0.72**	0.89
0.6	200	15	0.88	0.88	0.86	**0.78**	0.86	0.84	0.85	0.83	0.78	0.83
		25	0.88	0.89	0.84	**0.76**	0.82	0.84	0.86	0.82	**0.74**	0.81
		50	0.88	0.91	0.80	**0.70**	**0.79**	0.84	0.89	0.80	**0.71**	0.79
	500	15	0.87	0.88	0.86	**0.78**	0.88	0.83	0.84	0.83	0.77	0.82
		25	0.87	0.88	0.86	**0.75**	0.87	0.83	0.85	0.83	**0.73**	0.82
		50	0.87	0.90	0.83	**0.73**	0.82	0.83	0.88	0.82	**0.66**	0.82
	1,000	15	0.87	0.87	0.86	**0.77**	0.88	0.82	0.84	0.82	0.77	0.82
		25	0.87	0.88	0.86	**0.75**	0.89	0.82	0.85	0.82	**0.73**	0.82
		50	0.87	0.90	0.85	**0.74**	0.88	0.82	0.88	0.82	**0.67**	0.83

Note. Rho = interfactor correlation; *N* = sample size; PM = percentage of missingness; C = complete data results; PD = pairwise deletion results; MI = multiple imputation results; CAT = number of categories; TLI = Tucker–Lewis index; MCAR = missing completely at random; MAR = missing at random. Conditions with less than 50% of completed replications were excluded from the table.

line with the bias observed in parameter estimates. More specifically, the PD-based chi-square test statistic, RMSEA and WRMR were larger as compared with complete data, whereas the PD-based CFI and TLI were smaller. All the above five goodness-of-fit indices suggest that the model fit worse than the goodness of fit that would have been obtained using compete data.

When using the MI procedure, the chi-square test statistic was upwardly inflated. It is worth noting that the MI-based fit indices are computed using the naïve average approach across imputations. When fitting factor analysis models with incomplete multivariate normal (continuous) data, Asparouhov and Muthén (2010e) also found that the MI-based chi-square test statistics pooled using the naïve average tended to be overestimated, such that worse fit was indicated. As a result, RMSEA and WRMR, which were computed based on the chi-square test statistic, were also expected to be overestimated, suggesting a worse fit than the counterparts that would have been obtained using complete data. In contrast, the MI-based CFI and TLI are close to the complete data results across all simulated conditions, except for a few conditions where the percentage of missingness reached 50%.

Why do the MI-based CFI and TLI behave similarly to the complete data results? It is noted that CFI and TLI are incremental fit indices, where two chi-square test statistics (one from the proposed model and another from the null model) are involved in the computation. As a result, the inflations on the naïve averaged chi-square test statistics due to missing data could be canceled out to some degree. For example, when the model was correctly specified, sample size was 100, and 50% of the binary data were missing at random, the MI-based chi-square test statistic (i.e., χ^2 = 341.01, degrees of freedom [*df*] = 54) was overestimated, as compared with the chi-square test statistic computed using complete data (i.e., χ^2 = 54.06, *df* = 54). For the same simulation condition, the MI-based chi-square test statistic for the baseline model (i.e., χ^2 = 9749.817, *df* = 66) was also larger than the baseline chi-square test statistic computed using complete data (i.e., χ^2 = 5120.457, *df* = 66). As a consequence, the MI-based CFI and TLI, which are computed using both chi-square test statistics (for the fitted model and the baseline model),[5] are not too different from the complete data results. Additional studies are needed to further investigate this issue. Future methodological studies should also explore and develop alternative strategies to combine the robust corrected chi-square test statistics (and fit indices) across imputations.

In summary, when fitting ordinal factor analysis models with missing data, the performance of PD and MI depends on the missing data mechanism and percentage of missingness. In light of study findings, we offer the following recommendations to researchers fitting and evaluating ordinal factor analysis models with incomplete data. First, if it is arguable that the data are missing completely at random (e.g., planned missing data design), pairwise deletion can be used for fitting ordinal factor analysis models. In addition, the PD-based robust chi-square test statistics, RMSEA and WRMR are similar to those obtained with complete data when the percentage of missing data is less than 50%. When the percentage of missing data is large (i.e., $\geq$ 50%), researchers should rely on the CFI and TLI to evaluate goodness of fit. Second, under more general assumptions where the data are MAR, researchers should not use pairwise deletion. To be clear, the default estimator used for ordinal factor models in Mplus will be problematic in this case. The MI procedure should be used as it produces parameter estimates closer to complete data results, given the percentage of missingness is less than 50%. In terms of evaluating model fit, using the naïve average approach, the MI-based robust chi-square test statistics, RMSEA and WRMR are not trustworthy. The incremental fit indices (CFI and TLI) are recommended for assessing goodness of fit.

In the presence of missing data, the recommendations we made on the choice of goodness-of-fit indices were based on whether the estimates are similar to those that would be obtained from complete data. In practice, evaluating the goodness of model fit can be a challenging task. Previous methodological studies have shown that besides the level of model misspecification, the sample values of the fit indices also depend on other characteristics of the model and the data, such as the size of the model, sample size, and quantity of measurement (e.g., Hu & Bentler, 1998; Maydeu-Olivares, Shi, & Rosseel, 2018; McNeish, An, & Hancock, 2018; Saris, Satorra, & van der Veld, 2009; Shi, Lee, & Maydeu-Olivares, 2019; Shi, Lee, & Terry, 2018). Future studies on missing data should investigate the interaction effects of missing data and other model characteristics on goodness-of-fit indices. In addition, findings in the current study are based on binary and polytomous data with five response categories. The performance of missing data techniques under conditions with categorical data with various numbers of response categories and various items distributions (e.g., asymmetry) should be examined in future work. Finally, the Bayes estimator may be considered as another alternative for fitting ordinal factor analysis models with missing data (Asparouhov & Muthén, 2010c). It would be interesting to examine the performance of Bayesian parameter estimates and goodness-of-fit indices (e.g., the posterior predictive *p* values; Gelman, Meng, & Stern,1996; Meng, 1994) under various missing data conditions.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is supported by the Advanced Support for Innovative Research Excellence Grant No. 13580-17-44758 funded by the Office of the Vice President for Research at the University of South Carolina, the National Science Foundation under Grant No. SES-1659936, and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP; No. 2017R1C1B2012424).

ORCID iDs

Dexin Shi https://orcid.org/0000-0002-4120-6756
Taehun Lee https://orcid.org/0000-0001-8261-701X

Notes

1. See Little and Rubin (2002) for a detailed explanation of the missing data mechanisms.
2. In Mplus, when fitting ordinal factor analysis models with FIML, the Pearson's chi-square test statistics and the likelihood ratio test statistics are computed to assess how well the frequency of each possible response pattern is reproduced by the model. They can only be used when the number of possible patterns is small. In most applications of the ordinal factor analysis model, the number of possible patterns is large, and therefore, the goodness of fit of the model cannot be assessed. See Maydeu-Olivares and Joe (2005) for a detailed explanation.

3. As Muthén (1993) noted, estimation techniques for ordinal factor analysis can be written in a general form of a fit function $F = (\mathbf{r} - \rho_0)'\mathbf{W}(\mathbf{r} - \rho_0)'$, where $\mathbf{r}$ is a vector of the unduplicated elements of the polychoric correlation matrix, ρ_0 is the corresponding vector of correlations implied by the model, and $\mathbf{W}$ is a weight matrix. Under diagonally weighted least squares (DWLS) estimator, $\widehat{NF}$ does not follow a chi-square distribution. Therefore, robust corrections for mean, or mean and variance are typically applied to obtain a goodness-of-fit test.
4. Under MAR, as sample size decreased, number of response categories increased, and percentage of missingness increased, for the items with missing data, it is more likely to not observe a certain response category (from the data). In such cases, the proposed ordinal factor analysis model (with five response categories) cannot be estimated.
5. Let χ_0^2 and df_0 denote the chi-square statistic and degree of freedom for the baseline model, and χ_k^2 and df_k represent the chi-square statistic and degree of freedom for the target model, respectively. The sample CFI is computed as

$$\widehat{CFI} = \frac{\max(\chi_0^2 - df_0, 0) - \max(\chi_k^2 - df_k, 0)}{\max(\chi_0^2 - df_0, 0)}$$

and the sample TLI is given as

$$\widehat{TLI} = \frac{(\chi_0^2/df_0) - (\chi_k^2/df_k)}{(\chi_0^2/df_0) - 1}.$$

For the example provided, the sample CFI/TLI for complete data was 1.00/1.00 and the MI-based sample CFI/TLI (with missing data) was 0.97/0.96.

References

Allison, P. D. (1987). Estimation of linear models with incomplete data. *Sociological Methodology*, *17*, 71-103.

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

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumaker (Eds.), *Advanced structural equation modeling* (pp. 243-277). Mahwah, NJ: Lawrence Erlbaum.

Asparouhov, T., & Muthén, B. (2010a, August 14). *Weighted least squares estimation with missing data: Technical implementation*. Retrieved from https://www.statmodel.com/download/GstrucMissingRevision.pdf

Asparouhov, T., & Muthén, B. (2010b, September 29). *Multiple imputation with Mplus: Technical implementation*. Retrieved from http://statmodel2.com/download/Imputations7.pdf

Asparouhov, T., & Muthén, B. (2010c, September 29). *Bayesian analysis using Mplus: Technical implementation*. Retrieved from http://www.statmodel.com/download/Bayes3.pdf

Asparouhov, T., & Muthén, B. (2010d, May 3). *Simple second order chi-square correction: Technical implementation*. Retrieved from https://www.statmodel.com/download/WLSMV_new_chi21.pdf

Asparouhov, T., & Muthén, B. (2010e, July 27). *Chi-square statistics with multiple imputation: Technical implementation*. Retrieved from https://www.statmodel.com/download/MI7.pdf

Baker, F. B., & Kim, S. H. (2004). *Item response theory: Parameter estimation techniques*. Boca Raton, FL: CRC Press.

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

Box, G. E. (1979). Some problems of statistics and everyday life. *Journal of the American Statistical Association*, *74*(365), 1-4.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. s. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

DiStefano, C., Liu, J., Jiang, N., & Shi, D. (2018). Examination of the weighted root mean square residual: Evidence for trustworthiness? *Structural Equation Modeling*, *25*, 453-466.

DiStefano, C., & Morgan, G. B. (2014). A comparison of diagonal weighted least squares robust estimation techniques for ordinal data. *Structural Equation Modeling*, *21*, 425-438.

Enders, C. K. (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement*, *61*, 713-740.

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

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*, *8*, 430-457.

Finkbeiner, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika*, *44*, 409-420.

Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of IRT graded response models: Limited versus full information methods. *Psychological Methods*, *14*, 275-299.

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, *16*, 625-641.

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2016). Are fit indices really fit to estimate the number of factors with categorical variables? Some cautionary findings via Monte Carlo simulation. *Psychological Methods*, *21*, 93-111.

Gelman, A., Meng, X.-L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*, 733-760.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*, 206-213.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*, 323-343.

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424-453.

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

Jia, F. (2016). *Methods for handling missing non-normal data in structural equation modeling* (Unpublished doctoral dissertation). University of Kansas, Lawrence.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*, 183-202.

Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7: A guide to the program and applications* (2nd ed.). Chicago, IL: International Education Services.

Kamata, A., & Bauer, D. J. (2008). A note on the relation between factor analytic and item response theory models. *Structural Equation Modeling*, *15*, 136-153.

Kaplan, D. (1995). The impact of BIB spiraling-induced missing data patterns on goodness-of-fit tests in factor analysis. *Journal of Educational and Behavioral Statistics*, *20*, 69-82.

Larsen, R. (2011). Missing data imputation versus full information maximum likelihood with second-level dependencies. *Structural Equation Modeling*, *18*, 649-662.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: Wiley.

Lu, K. (2017). Number of imputations needed to stabilize estimated treatment difference in longitudinal data analysis. *Statistical Methods in Medical Research*, *26*, 674-690.

Maydeu-Olivares, A., Cai, L., & Hernández, A. (2011). Comparing the fit of item response theory and factor analysis models. *Structural Equation Modeling*, *18*, 333-356.

Maydeu-Olivares, A., Fairchild, A. J., & Hall, A. G. (2017). Goodness of fit in item factor analysis: Effect of the number of response alternatives. *Structural Equation Modeling*, *24*, 495-505.

Maydeu-Olivares, A., & Joe, H. (2005). Limited- and full-information estimation and goodness-of-fit testing in 2^n contingency tables: A unified framework. *Journal of the American Statistical Association*, *100*, 1009-1020.

Maydeu-Olivares, A., Shi, D., & Rosseel, Y. (2018). Assessing fit in structural equation models: A Monte-Carlo evaluation of RMSEA versus SRMR confidence intervals and tests of close fit. *Structural Equation Modeling*, *25*, 389-402.

MacCallum, R. C. (2003). 2001 Presidential address: Working with imperfect models. *Multivariate Behavioral Research*, *38*, 113-139.

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

McNeish, D., An, J., & Hancock, G. R. (2018). The thorny relation between measurement quality and fit index cutoffs in latent variable models. *Journal of Personality Assessment*, *100*, 43-52.

Meng, X.-L. (1994). Posterior predictive p-values. *Annals of Statistics*, *22*, 1142-1160.

Meng, X.-L., & Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, *79*, 103-111.

Muthén, B. O. (1993). Goodness of fit with categorical and other nonnormal variables. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 205-234). Newbury Park, CA: Sage.

Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997, November 18). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Retrieved from http://pages.gseis.ucla.edu/faculty/muthen/articles/Article_075.pdf

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

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*, 599-620.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods*, *6*, 328-362.

Olinsky, A., Chen, S., & Harlow, L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modeling. *European Journal of Operational Research*, *151*, 53-79.

Peng, C. Y. J., Harwell, M., Liou, S. M., & Ehman, L. H. (2006). Advances in missing data methods and implications for educational research. In S. Sawilowsky (Ed.), *Real data analysis* (pp. 31-78). Greenwich, CT: Information Age.

Prevosti, F. J., & Chemisquy, M. A. (2010). The impact of missing data on real morphological phylogenies: Influence of the number and distribution of missing entries. *Cladistics*, *26*, 326-339.

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory. An introduction to psychometric theory*. New York, NY: Routledge.

Rhemtulla, M., Brosseau-Liard, P. É. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*, 354-373.

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

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley.

Saris, W. E., Satorra, A., & Van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling*, *16*, 561-582.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. Von Eye & C. C. Clogg (Eds.), *Latent variable analysis. Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, *33*, 545-571.

Scheffer, J. (2002), Dealing with missing data. *Research Letters in the Information and Mathematical Sciences*, *3*, 153-160.

Shi, D., DiStefano, C., McDaniel, H. L., & Jiang, Z. (2018). Examining chi-square test statistics under conditions of large model size and ordinal data. *Structural Equation Modeling*, *25*, 924-945.

Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the model size effect on SEM fit indices. *Educational and Psychological Measurement*, *79*, 310-334.

Shi, D., Lee, T., & Terry, R. A. (2018). Revisiting the model size effect in structural equation modeling. *Structural Equation Modeling*, *25*, 21-40.

Shi, D., Maydeu-Olivares, A., & DiStefano, C. (2018). The relationship between the standardized root mean square residual and model misspecification in factor analysis models. *Multivariate Behavioral Research*, *53*, 676-694.

Steiger, J. H. (1989). *EzPATH: A supplementary module for SYSTAT and SYGRAPH*. Evanston, IL: Systat.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173-180.

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

Wu, W., Jia, F., & Enders, C. (2015). A comparison of imputation strategies for ordinal missing data on Likert scale variables. *Multivariate Behavioral Research*, *50*, 484-503.

Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (Unpublished doctoral dissertation). University of California, Los Angeles.

Yu, C.-Y., & Muthén, B. (2002, April 4). Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. *Paper presented at the annual conference of the American Educational Research Association*, New Orleans, LA.

Zhang, Z., & Wang, L. (2013). Methods for mediation analysis with missing data. *Psychometrika*, *78*, 154-184.

Zhao, Y. (2014). *The performance of model fit measures by robust weighted least squares estimators in confirmatory factor analysis* (Unpublished doctoral dissertation). Pennsylvania State University, State College.