Why Ordinal Variables Can (Almost) Always Be Treated as Continuous Variables: Clarifying Assumptions of Robust Continuous and Ordinal Factor Analysis Estimation Methods

Alexander Robitzsch [1,2]*

[1] IPN–Leibniz Institute for Science and Mathematics Education, Kiel, Germany, [2] Center for International Student Assessment, Zentrum für Internationale Bildungsvergleichsstudien (ZIB), Kiel, Germany

The analysis of factor structures is one of the most critical psychometric applications. Frequently, variables (i.e., items or indicators) resulting from questionnaires using ordinal items with 2–7 categories are used. There are plenty of articles that recommend treating ordinal variables in a factor analysis by default as ordinal and not as continuous imposing a multivariate normal distribution assumption. In this article, we exhibit that the reasoning behind such suggestions is flawed. In our view, findings from simulation studies cannot tell about the right modeling strategy of ordinal variables in factor analysis. Moreover, it is argued that ordinal factor models impose a normality assumption for underlying continuous variables, which might also often be incorrect in empirical applications. However, researchers seldom opt for more flexible modeling strategies that involve correctly specified distributions. Finally, the consequences of modeling choices for validity, reliability, measurement invariance, handling of missing data, and the assessment of global model fit are discussed.

Keywords: factor analysis, ordinal variable, polychoric correlations, structural equation modeling, Gaussian copula model

1. INTRODUCTION

The analysis of factor structures is one of the most critical psychometric applications. Frequently, variables (i.e., items, or indicators) resulting from questionnaires are analyzed. These variables are often assessed with Likert scales that have a finite number of ordered categories. In many applications, the number of categories ranges between 2 and 7. An often posed question by applied researchers is about the most favorable approach for factor analysis in the presence of ordinal variables. First, ordinal variables could be treated as in the case of continuous variables, and the same estimation method would be used. Second, a factor model based on a distributional assumption for ordinal variables could be fitted (i.e., an ordinal factor model). There is a diversity of methodological literature addressing this issue (Dolan, 1994; DiStefano, 2002; Lei, 2009; Flora et al., 2012; Rhemtulla et al., 2012; Sass et al., 2014; Barendse et al., 2015; Asún et al., 2016; Li, 2016; Jia and Wu, 2019). The main message of most of the papers seems to be that ordinal variables should be treated as ordinal (i.e., not being treated as continuous variables) if there are only a few categories or the frequency distributions of variables are skewed (e.g., Rhemtulla et al., 2012). In this

article, it is argued that treating ordinal variables as continuous can (almost) always be defended, and the argument for doing so does neither depend on the number of categories nor the marginal distribution of ordinal variables.

The article is structured as follows. In section 2, the two major competing estimation approaches are contrasted, and their assumptions are clarified. In section 3, it is argued that there are many options for modeling ordinal variables, and the choice of the ordinal method that is recommended by default is as arbitrary as treating ordinal variables as continuous variables. Finally, in section 4, the consequences of modeling choices for validity, reliability, measurement invariance, and the assessment of global model fit are discussed.

2. NORMALITY ASSUMPTION OR LATENT NORMALITY ASSUMPTION?

2.1. Alternative Modeling Strategies

Assume that a vector of ordinal variables $Y = (Y_1, \ldots, Y_I)$ is given. For simplicity, we assume that each variable Y_i has values $0, 1, \ldots, K$. Let $p_{i,k} = P(Y_i \leq k)$ denotes the cumulative frequencies of variable i. If the ordinal variables would be treated as continuous, a linear factor model

$$Y = \mu + \Lambda F + E \qquad (1)$$

is assumed, where μ are intercepts, Λ is a loading matrix, F is a multidimensional factor variable, and E denotes a vector of residuals. Assume that $E(F) = 0$ and $E(E) = 0$. Let us define $\Phi = \mathrm{Var}(F)$ and $\Psi = \mathrm{Var}(E)$. Typically, Ψ is a diagonal matrix. The covariance matrix $\Sigma = \mathrm{Var}(Y)$ in the factor model given in Equation (1) is modeled as an implied covariance matrix $\Sigma_0(\theta)$

$$\Sigma \approx \Sigma_0(\theta) = \Lambda \Phi \Lambda^T + \Psi \qquad (2)$$

Hence, in (2), it is assumed that observed covariances of variables are represented by model parameters (i.e., loadings, factor covariances and variances, and residual variances). For ease of exposition, we interpret Σ as a Pearson product-moment correlation matrix, which is the covariance matrix if ordinal variables Y would have been standardized prior to analysis. Let θ denote the vector of parameters in Λ, Φ, and Ψ that are freely estimated.

Two estimation methods for estimating θ can be distinguished (see Jöreskog, 2007, for an overview). First, a multivariate normal distribution can be assumed. Then, the estimated Pearson correlation matrix S for variables Y is a sufficient statistic for θ, and the fitting function (i.e., log-likelihood function in maximum likelihood estimation) is given as

$$F_{\mathrm{ML}}(\theta; S) = \log|\Sigma_0(\theta)| + \mathrm{tr}(S\Sigma_0(\theta)^{-1}) - \log|S| - I \quad (3)$$

The multivariate normal distribution will be misspecified if variables are ordinal. However, ML parameter estimates are consistent and converge to a parameter θ that maximizes the Kullback-Leibler information (White, 1982). The optimal parameter θ is obtained if S is replaced by the true population

covariance matrix Σ in (3) (Arminger and Schoenberg, 1989; Olsson et al., 2000; Yuan and Bentler, 2007). Note that the model assumption $\Sigma = \Sigma_0(\theta)$ can be correct when the data is not multivariate normally distributed. The choice of a misspecified ML function F_{ML} must not necessarily result in inconsistent (and, hence, biased) parameter estimates. Moreover, the so-called robust ML estimator (MLR; Savalei, 2014; Yuan and Bentler, 2007) provides valid statistical inference in a misspecified model (Satorra, 1992), and improvements have been proposed (Lai, 2019). Second, weighted least squares estimation based on the estimated correlation matrix S can be employed. Here, we consider diagonally weighted least squares (DWLS), and the fitting function is given as (Jöreskog, 2007)

$$F_{\mathrm{DWLS}}(\theta; S) = \mathrm{tr}[(S - \Sigma_0(\theta))W] \qquad (4)$$

where W is a diagonal weighting matrix. MacCallum et al. (2007) argue that using DWLS instead of ML possesses advantages in the case of misspecified models [i.e., $\Sigma \neq \Sigma_0(\theta)$].

Alternatively, ordinal variables can be modeled with ordinal marginal distributions (Forero et al., 2009; Yang-Wallentin et al., 2010). The idea is that there are underlying normally distributed variables $Y^* = (Y_1^*, \ldots, Y_I^*)$ and thresholds $-\infty = \tau_{i,0} < \tau_{i,1} < \ldots < \tau_{K+1,\infty} = \infty$ such that

$$Y_i = k \qquad \Leftrightarrow \qquad \tau_{i,k} \leq Y_i^* \leq \tau_{i,k+1} \qquad (k = 0, 1, \ldots, K) \qquad (5)$$

The thresholds are given as $\tau_{ik} = F^{-1}(p_{i,k+1})$, where F denotes the standard normal distribution function. In an ordinal treatment of ordinal variables, a linear factor model for the underlying latent normally distributed variables Y^* (referred to as latent normality in the sequel) is assumed:

$$Y^* = \mu + \Lambda F + E \qquad (6)$$

Then, the covariance structure of $\Sigma^* = \mathrm{Var}(Y^*)$ of latent variables is modeled as

$$\Sigma^* \approx \Sigma_0^*(\theta^*) = \Lambda^* \Phi^* \Lambda^{*T} + \Psi^* \qquad (7)$$

The covariance matrix Σ^* can be estimated by employing polychoric correlations (Muthén, 1984). Note that the estimated parameter θ in (2) when treating variables as continuous will typically be different from θ^* in (7) when treating variables as ordinal, even if original variables Y would have been standardized prior to analysis (Rhemtulla et al., 2012). Often, DWLS estimation based on the estimated polychoric correlation matrix S^* is conducted (Muthén, 1984; Yang-Wallentin et al., 2010):

$$F_{\mathrm{cat-DWLS}}(\theta; S^*) = \mathrm{tr}[(S^* - \Sigma_0^*(\theta^*))W^*] \qquad (8)$$

It should be emphasized that the ordinal factor models are also labeled as item response models (IRT). The assumption of latent normality corresponds to the graded response model with a probit link function (Takane and de Leeuw, 1987; Glockner-Rist and Hoijtink, 2003; Kamata and Bauer, 2008). In standard IRT software, ML estimation is typically utilized for

estimation by default (e.g., in the R package mirt; Chalmers, 2012) while in standard SEM software, limited information methods, such as DWLS are employed (e.g., in the R package lavaan; Rosseel, 2012).

Practitioners often seek advice from methodologists of how to analyze ordinal variables in factor analysis. This amounts to the question of whether Pearson correlations or polychoric correlations should be used in the factor analysis. In the following, we critically discuss the recommendations of some often cited methodological articles.

2.2. There Is No "Correct" Modeling Strategy

Many papers recommend that ordinal variables should be modeled as ordinal if there are only a few categories or if the frequency distributions are asymmetrical (Rhemtulla et al., 2012; Li, 2016). These articles often state that biased estimates (i.e., factor loadings) would have been obtained if the (misspecified) model assuming multivariate normality would be applied. Their reasoning is based on simulation studies. We now argue that simulation studies do not help for providing rationale of the appropriate modeling strategy because it would be relatively simple to design simulation studies with ordinal variables that fulfill the linear factor model for Pearson correlations (i.e., Equation 2) instead of the factor model for polychoric correlations (i.e., Equation 7).

Researchers who claim the superiority of ordinal models for ordinal variables presuppose that Equation (7) holds, i.e., the matrix representation $\Sigma^* = \Lambda^* \Phi^* \Lambda^{*T} + \Psi$ holds for the polychoric correlation matrix (e.g., Li, 2016; Rhemtulla et al., 2012). Hence, they base inference on the latent normal variables Y^*. By employing the corresponding estimation function $F_{\text{cat}-\text{DWLS}}$, unbiased parameter estimates are obtained. However, if the variables would be analyzed as continuous (i.e., using F_{ML} or F_{DWLS}), biased parameter estimates would be obtained. This is almost trivial as the true model is ordinal, and the question is whether the incorrect model that treats variables as continuous can also recover true parameters (i.e., whether $\theta^* = \theta$ holds in the population). A typical argument found in several articles is as follows. Under the assumption that continuous variables underlie the observed ordinal variables, the matrix of Pearson correlations underestimates of the correlation matrix among the underlying continuous variables, that is the polychoric correlation matrix (Olsson, 1979; Rhemtulla et al., 2012). Hence, using the Pearson correlation constitutes the incorrect input of the factor analysis, while polychoric correlations would be the correct one, and only the ordinal factor analysis would provide unbiased parameter estimates.

We now sketch the design of a simulation that "shows" that treating ordinal variables as continuous (i.e., using Pearson correlations) result in unbiased estimation while treating them as ordinal in the model assuming latent normality will result in biased estimates. Assume predetermined thresholds $\tau_{i,k}$ ($k = 0, 1, \ldots, K$). The Pearson correlation σ_{ij} of variables i and j is a function of thresholds $\tau_{i,k}$ and $\tau_{i,j}$, and the polychoric correlation σ_{ij}^*. One can always find a polychoric correlation σ_{ij}^* such that the

covariance of i and j equals σ_{ij} (e.g., by applying some numerical procedure for finding a root). Moreover, assume that the linear factor model $\Sigma = \Lambda \Phi \Lambda^T + \Psi$ holds for observed covariances (i.e., Equation 2 holds true). Therefore, the ordinal factor model (Equation 7) will not be fulfilled, and hence, treating ordinal variables in an ordinal factor model assuming latent normality will result in biased estimates. Of course, a simulation exercise could be additionally carried to demonstrate this reasoning.

One could also argue that true scores (and consequently latent variables) are clearly defined in stochastic measurement theory (Steyer, 1989) as the expected value of an intraindividual distribution of an ordinal variable (see also Holland, 1990). Although observed variables are ordinal, item-specific true scores are bounded but non-integers. A factor model constitutes a model assumption for these true scores.

What can be learned from these observations? At best, findings from the literature that comes with recommendations to practitioners are only useful in identifying data constellations in which the continuous and the ordinal treatment of ordinal variables can result in similar parameter estimates. In our view, simulation studies or empirical data cannot be employed for deciding among the two competitive modeling strategies. Hence, a researcher must decide whether the factor structure should be posed on Pearson correlations or polychoric correlations.

As a cautionary note, one should add that categories of variables are stretched in the ordinal variable model according to their empirical frequencies when representing the latent variables F in the factor model. In contrast, by treating the variables as continuous, no such implicit transformation is carried out. The modeling choice relates to the question about the meaning of distances between categories of an ordinal variable. While it can be almost always be argued that assuming equal distances seem to be implausible in practice, distances that are derived on a sole empirical basis are equally implausible.

As a conclusion, we would not like to argue that dichotomous variables should be treated as continuous (see Maydeu-Olivares, 2005; Tran and Formann, 2010; for differences in parameter estimates for the two modeling strategies). However, we think that for items with 3–6 categories, using the linear factor model by treating variables as continuous is as defensible as treating them as ordinal. However, researchers should be aware that when estimating factor models with a misspecified distribution, the statistical inference should be obtained with the MLR estimator.

3. THE NORMALITY ASSUMPTION AND THE LATENT NORMALITY ASSUMPTION ARE EQUALLY RESTRICTIVE

In section 2, it was assumed that underlying continuous variables of the ordinal variables are normally distributed (i.e., latent normality holds). Typically, the latent normality assumption has often been taken for granted by applied researchers (Foldnes and Grønneberg, 2020). If there are violations of latent normality, parameter estimates based on the incorrect latent normality assumption can provide substantially biased estimates (Jin and Yang-Wallentin, 2017; Foldnes and Grønneberg, 2020). It has

shown that the latent normality assumption can be empirically tested (Maydeu-Olivares et al., 2009; Raykov and Marcoulides, 2015; Foldnes and Grønneberg, 2020). However, it seems that these tests are seldom conducted in practice.

If the ordinal nature of variables would be taken seriously, more sophisticated modeling strategies for factor analysis that try to estimate more flexible distributions are required (Jin and Yang-Wallentin, 2017; Foldnes and Grønneberg, 2019). A particularly attractive distribution class is the factor copula model (Krupskii and Joe, 2013; Nikoloulopoulos and Joe, 2015; Ackerer and Vatter, 2017; Krupskii and Genton, 2018). Copula models decompose a joint distribution for modeling into marginal distributions and modeling the dependency structure. Gaussian copula models pose a multivariate normal distribution for modeling the dependency structure while allowing a semiparametric estimation of the marginal distribution (Hoff, 2007; Gruhl et al., 2013; Murray et al., 2013). As a consequence, underlying latent variables can deviate from the latent normality assumption. More formally, it is assumed that there exists a vector of multivariate normally distributed variables Y^* with $\text{Var}(Y^*) = \Sigma^*$. For an ordinal variable Y_i, there exists an underlying latent variable $\tilde{Y}_i = G_i^{-1}(F(Y_i^*))$, where G_i denotes the distribution function of variable $\tilde{Y}_i$, and F is the normal distribution function. Like in Equation (5), the ordinal variable Y_i is obtained by discretizing the underlying continuous variable $\tilde{Y}_i$ with respect to thresholds $\tau_{i,k}$. If latent normality is fulfilled, it holds that $\tilde{Y}_i = Y_i^*$. For example, the distributions G_i could be, for example, the logistic, skew normal, skew t, or cloglog distribution. These distributions could be fixed or estimated using empirical data (Gruhl et al., 2013).

It should be noted that there is active research in factor analysis for continuous variables with non-normally distributed factors or residuals that can be skewed, bounded, or are mixtures of distributions (Song et al., 2010; Kelava and Brandt, 2014; Zhang et al., 2014; Asparouhov and Muthén, 2016; Lin et al., 2016; Revuelta et al., 2020). Using these more complex distributions would reduce the degree of distributional misspecification in the factor model. Moreover, in a few articles, the estimation of the link functions in item response models is considered (Peress, 2012; Liang and Browne, 2015; Feuerstahler, 2019). Notably, estimating the link function in IRT is equivalent to estimating the marginal distributions of the underlying latent variables Y_i^* in a Gaussian copula model.

To sum up, we believe that also estimating the marginal latent distributions of underlying continuous variables adds a further layer of complexity regarding estimation and interpretation in the analysis. However, if scholars suggest to always model ordinal variables appropriately by a well-fitting model, there is no reason for only fitting the ordinal model based on latent normality. In this regard, multivariate latent normality is a testable assumption as multivariate normality of the ordinal variables, and it can be supposed that both assumptions will typically be violated in practice. As a consequence, both treatments, the continuous (i.e., using Pearson correlations Σ) and the ordinal (i.e., using polychoric correlations Σ^*), correspond to misspecified models, and it is difficult to speculate about a plausible data-generating

model in a concrete application. It can be argued that by choosing a particular fitting function, the parameter of interest is defined, and one could consider the parameter estimate in a sample as an estimator that converges to some optimal parameter in the fitting procedure that would have been obtained in an infinite sample size. As it becomes clear in this case, again, simulation studies cannot decide about choosing an adequate modeling strategy (i.e., choosing a fitting function).

4. DISCUSSION

In this article, it is argued that the often found recommendations for not treating ordinal variables in factor models as continuous are not justified. The choice for a particular modeling strategy implies that it is assumed whether the linear factor model must hold for Pearson correlations (i.e., the normality assumption) or polychoric correlations (i.e., the latent normality assumption). This choice cannot be derived from an empirical dataset or simulation studies, although some articles argue otherwise.

It should be mentioned that alternative modeling choices are rarely discussed under the perspective of validity (but see Ferrando, 1999, for an exception). Using ordinal factor models implies the use of a non-linear scoring formula for factor scores, which is not the case when variables are treated as continuous. It has to be defended from a validity perspective (i.e., by studying relationships with external criterion variables) why the non-linear scoring rule from an ordinal factor model is preferable to a linear scoring rule from the factor model using Pearson correlations.

We would also like to emphasize that the computation of a reliability measure (e.g., ω) is defensible if ordinal variables would be treated as continuous by analyzing Pearson correlations in factor analysis (see also Chalmers, 2018). As it has been pointed out by Lucke (2005), the reliability of a sum score can be defined on a matrix decomposition into a part for the true score (i.e., $\Lambda \Phi \Lambda^T$) and a part for the error (i.e., Ψ). Unbiased estimation of reliability only needs the unbiased estimation of the model parameters that are involved in the computation. It is not required that the factor model is correctly specified with respect to distributional assumptions of variables. In our view, the assessment of reliability based on an ordinal factor model (Green and Yang, 2009) is not necessarily superior to using a factor model that treats ordinal variables as continuous. Moreover, we generally question the strong preference of a model-based assessment of reliability from a factor analysis (e.g., McNeish, 2018; Sijtsma, 2009; Yang and Green, 2011) compared to design-based reliability measures (i.e., internal consistency, see Meyer, 2010, for assumptions of Cronbach's alpha).

In factor analysis, the investigation of measurement invariance (Millsap, 2011) is heavily discussed. It has been recommended that invariance analysis for ordinal variables should also treat variables as ordinal (Chen et al., 2020; Svetina et al., 2020). By continuing our arguments, we think that the assessment of invariance can be equally defended by treating ordinal variables as continuous. Effect sizes of non-invariance

could be defined in the metric of raw scores (i.e., based on means and covariances; see Gunn et al., 2020).

The assessment of the global model fit in ordinal factor models utilizing fit statistics is more challenging than in models that presuppose multivariate normality. Recently, it has been proposed to define fit statistics for ordinal factor models that have the same rationale as in the normality case by replacing Pearson correlations (i.e., matrix Σ) by polychoric correlations (i.e., matrix Σ^*) in the definition of population effect sizes of global model fit (Savalei, 2020). Pursuing such a strategy implies that the latent normality assumption is taken for granted, and potential misspecification of latent normality is not quantified in the assessment of model fit. Hence, we think that the assessment of model fit should rely on bivariate distributions of observed variables (Maydeu-Olivares, 2013). In this respect, the assessment of model fit is independent of using a particular fitting function, a property which, however, might see some scholars as a disadvantage.

We argued that a researcher might be interested in interpreting model parameters of a factor model with a misspecified distribution. However, in the presence of missing data, fitting a misspecified factor model with ML or weighted least squares approaches only result in consistent parameter estimates if data are missing completely at random (MCAR; see Yuan, 2009). In the case of missing at random (MAR) data, in general, inconsistent and biased parameter estimates will be obtained except for linearly related variables (Yuan, 2009; Yuan and Bentler, 2010; Yuan et al., 2012). For ordinal variables, linear relations among variables are implausible. However, it could be argued that there are typically only mild deviations of the MCAR assumption in empirical datasets (Newman, 2014). If substantial deviations of MCAR are suspected, and a misspecified factor model should be estimated, imputation based approaches might be preferable (Gottschall et al., 2012; Jia and Wu, 2019). The use of sufficiently complex imputation models, such as the Gaussian copula model (Hollenbach et al., 2018), mixture models (Murray and Reiter, 2016), or latent class models (Vermunt et al., 2008; Si and Reiter, 2013) are advantageous to minimize possible distributional misspecifications for MAR data. Appropriate imputation models can also treat specific deviations from

MAR (missing not at random; MNAR; Harel and Schafer, 2009; Jung et al., 2011; Kano and Takai, 2011; Zhang and Reiser, 2015; Bartolucci et al., 2018; Kuha et al., 2018; Pohl and Becker, 2020).

As pointed out by a reviewer, the application of factor models based on the normal distribution requires fewer methodological skills than an ordinal factor model. Forcing the application of the more sophisticated ordinal factor model might have a biasing effect against research produced in developing countries because, in those countries, the use of factor models under the normal distribution still prevails. The reviewer remarked that our recommendation for also considering the more straightforward approach might enhance inclusion and promote diversity in academia. This aspect was indeed not the primary motivation for writing this article, but it could be a pleasant side effect.

To conclude, we tried to argue that simple suggestions derived from simulation studies cannot tell what the appropriate modeling strategy for handling ordinal variables in the factor analysis is. It is maybe improbable to reach a consensus about this issue. We tend to prefer the modeling approach that possesses higher validity. As a comprehensive assessment of validity cannot be reduced to a single quantitative measure, the choice of an appropriate factor modeling approach is not a purely statistical issue. However, it must be defended by a researcher with a concrete question at hand. Finally, we would like to emphasize that modeling ordinal variables with Pearson correlations as well as polychoric correlations can provide complementary information. Moreover, the results of alternative model specifications are worth to be reported (see also Steegen et al., 2016; Hoffmann et al., 2020).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

AR was responsible for writing the article.

REFERENCES

Ackerer, D., and Vatter, T. (2017). Dependent defaults and losses with factor copula models. *Depend. Model.* 5, 375–399. doi: 10.1515/demo-2017-0022

Arminger, G., and Schoenberg, R. J. (1989). Pseudo maximum likelihood estimation and a test for misspecification in mean and covariance structure models. *Psychometrika* 54, 409–425. doi: 10.1007/BF02294626

Asparouhov, T., and Muthén, B. (2016). Structural equation models and mixture models with continuous nonnormal skewed distributions. *Struct. Equat. Model.* 23, 1–19. doi: 10.1080/10705511.2014.947375

Asún, R. A., Rdz-Navarro, K., and Alvarado, J. M. (2016). Developing multidimensional Likert scales using item factor analysis: the case of four-point items. *Sociol. Methods Res.* 45, 109–133. doi: 10.1177/0049124114566716

Barendse, M. T., Oort, F. J., and Timmerman, M. E. (2015). Using exploratory factor analysis to determine the dimensionality of discrete

responses. *Struct. Equat. Model.* 22, 87–101. doi: 10.1080/10705511.2014.9 34850

Bartolucci, F., Montanari, G. E., and Pandolfi, S. (2018). Latent ignorability and item selection for nursing home case-mix evaluation. *J. Classif.* 35, 172–193. doi: 10.1007/s00357-017-9227-9

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *J. Stat. Softw.* 48, 1–29. doi: 10.18637/jss.v048.i06

Chalmers, R. P. (2018). On misconceptions and the limited usefulness of ordinal alpha. *Educ. Psychol. Meas.* 78, 1056–1071. doi: 10.1177/0013164417727036

Chen, P.-Y., Wu, W., Garnier-Villarreal, M., Kite, B. A., and Jia, F. (2020). Testing measurement invariance with ordinal missing data: a comparison of estimators and missing data techniques. *Multivar. Behav. Res.* 55, 87–101. doi: 10.1080/00273171.2019.1608799

DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Struct. Equat. Model.* 9, 327–346. doi: 10.1207/S15328007SEM0903_2

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: a comparison of categorical variable estimators using simulated data. *Brit. J. Math. Stat. Psychol.* 47, 309–326. doi: 10.1111/j.2044-8317.1994.tb01039.x

Ferrando, P. J. (1999). Likert scaling using continuous, censored, and graded response models: effects on criterion-related validity. *Appl. Psychol. Meas.* 23, 161–175. doi: 10.1177/01466219922031284

Feuerstahler, L. M. (2019). Metric transformations and the filtered monotonic polynomial item response model. *Psychometrika* 84, 105–123. doi: 10.1007/s11336-018-9642-9

Flora, D. B., LaBrish, C., and Chalmers, R. P. (2012). Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. *Front. Psychol.* 3:55. doi: 10.3389/fpsyg.2012.00055

Foldnes, N., and Grønneberg, S. (2019). On identification and non-normal simulation in ordinal covariance and item response models. *Psychometrika* 84, 1000–1017. doi: 10.1007/s11336-019-09688-z

Foldnes, N., and Grønneberg, S. (2020). Pernicious polychorics: the impact and detection of underlying non-normality. *Struct. Equat. Model.* 27, 525–543. doi: 10.1080/10705511.2019.1673168

Forero, C. G., Maydeu-Olivares, A., and Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: a Monte Carlo study comparing DWLS and ULS estimation. *Struct. Equat. Model.* 16, 625–641. doi: 10.1080/10705910903203573

Glockner-Rist, A., and Hoijtink, H. (2003). The best of both worlds: factor analysis of dichotomous data using item response theory and structural equation modeling. *Struct. Equat. Model.* 10, 544–565. doi: 10.1207/S15328007SEM1004_4

Gottschall, A. C., West, S. G., and Enders, C. K. (2012). A comparison of item-level and scale-level multiple imputation for questionnaire batteries. *Multivar. Behav. Res.* 47, 1–25. doi: 10.1080/00273171.2012.640589

Green, S. B., and Yang, Y. (2009). Reliability of summed item scores using structural equation modeling: an alternative to coefficient alpha. *Psychometrika* 74, 155–167. doi: 10.1007/s11336-008-9099-3

Gruhl, J., Erosheva, E. A., and Crane, P. K. (2013). A semiparametric approach to mixed outcome latent variable models: estimating the association between cognition and regional brain volumes. *Ann. Appl. Stat.* 7, 2361–2383. doi: 10.1214/13-AOAS675

Gunn, H. J., Grimm, K. J., and Edwards, M. C. (2020). Evaluation of six effect size measures of measurement non-invariance for continuous outcomes. *Struct. Equat. Model.* 27, 503–514. doi: 10.1080/10705511.2019.1689507

Harel, O., and Schafer, J. L. (2009). Partial and latent ignorability in missing-data problems. *Biometrika* 96, 37–50. doi: 10.1093/biomet/asn069

Hoff, P. D. (2007). Extending the rank likelihood for semiparametric copula estimation. *Ann. Appl. Stat.* 1, 265–283. doi: 10.1214/07-AOAS107

Hoffmann, S., Schönbrodt, F. D., Elsas, R., Wilson, R., Strasser, U., and Boulesteix, A. (2020). The multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines. *MetaArXiv.* doi: 10.31222/osf.io/afb9p

Holland, P. W. (1990). On the sampling theory roundations of item response theory models. *Psychometrika* 55, 577–601. doi: 10.1007/BF02294609

Hollenbach, F. M., Bojinov, I., Minhas, S., Metternich, N. W., Ward, M. D., and Volfovsky, A. (2018). Multiple imputation using Gaussian copulas. *Sociol. Methods Res.* doi: 10.1177/0049124118799381. [Epub ahead of print].

Jia, F., and Wu, W. (2019). Evaluating methods for handling missing ordinal data in structural equation modeling. *Behav. Res. Methods* 51, 2337–2355. doi: 10.3758/s13428-018-1187-4

Jin, S., and Yang-Wallentin, F. (2017). Asymptotic robustness study of the polychoric correlation estimation. *Psychometrika* 82, 67–85. doi: 10.1007/s11336-016-9512-2

Jöreskog, K. G. (2007). "Factor analysis and its extensions," in *Factor Analysis at 100*, eds R. Cudeck and R. C. MacCallum (Mahwah, NJ: Lawrence Erlbaum), 47–77.

Jung, H., Schafer, J. L., and Seo, B. (2011). A latent class selection model for nonignorably missing data. *Comp. Stat. Data An.* 55, 802–812. doi: 10.1016/j.csda.2010.07.002

Kamata, A., and Bauer, D. J. (2008). A note on the relation between factor analytic and item response theory models. *Struct. Equat. Model.* 15, 136–153. doi: 10.1080/10705510701758406

Kano, Y., and Takai, K. (2011). Analysis of NMAR missing data without specifying missing-data mechanisms in a linear latent variate model. *J. Multivar. Anal.* 102, 1241–1255. doi: 10.1016/j.jmva.2011.04.007

Kelava, A., and Brandt, H. (2014). A general non-linear multilevel structural equation mixture model. *Front. Psychol.* 5:748. doi: 10.3389/fpsyg.2014.00748

Krupskii, P., and Genton, M. G. (2018). Linear factor copula models and their properties. *Scand. J. Stat.* 45, 861–878. doi: 10.1111/sjos.12325

Krupskii, P., and Joe, H. (2013). Factor copula models for multivariate data. *J. Multivar. Anal.* 120, 85–101. doi: 10.1016/j.jmva.2013.05.001

Kuha, J., Katsikatsou, M., and Moustaki, I. (2018). Latent variable modelling with non-ignorable item nonresponse: multigroup response propensity models for cross-national analysis. *J. R. Stat. Soc. A Stat.* 181, 1169–1192. doi: 10.1111/rssa.12350

Lai, K. (2019). More robust standard error and confidence interval for SEM parameters given incorrect model and nonnormal data. *Struct. Equat. Model.* 26, 260–279. doi: 10.1080/10705511.2018.1505522

Lei, P.-W. (2009). Evaluating estimation methods for ordinal data in structural equation modeling. *Qual. Quant.* 43, 495–507. doi: 10.1007/s11135-007-9133-z

Li, C.-H. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychol. Methods* 21, 369–387. doi: 10.1037/met0000093

Liang, L., and Browne, M. W. (2015). A quasi-parametric method for fitting flexible item response functions. *J. Educ. Behav. Stat.* 40, 5–34. doi: 10.3102/1076998614556816

Lin, T.-I., McLachlan, G. J., and Lee, S. X. (2016). Extending mixtures of factor models using the restricted multivariate skew-normal distribution. *J. Multivar. Anal.* 143, 398–413. doi: 10.1016/j.jmva.2015.09.025

Lucke, J. F. (2005). The α and the ω of congeneric test theory: An extension of reliability and internal consistency to heterogeneous tests. *Appl. Psychol. Meas.* 29, 65–81. doi: 10.1177/0146621604270882

MacCallum, R. C., Browne, M. W., and Cai, L. (2007). "Factor analysis models as approximations," in *Factor Analysis at 100*, eds R. Cudeck and R. C. MacCallum (Mahwah, NJ: Lawrence Erlbaum), 153–175. doi: 10.4324/9780203936764

Maydeu-Olivares, A. (2005). Linear item response theory, nonlinear item response theory and factor analysis: a unified framework. in *Contemporary Psychometrics: A Festschrift for Roderick P. McDonald*, eds A. Maydeu-Olivares and J. J. McArdle (Mahwah, NJ: Lawrence Erlbaum Associates), 73–102. doi: 10.4324/9781410612977

Maydeu-Olivares, A. (2013). Goodness-of-fit assessment of item response theory models. *Meas. Interdiscipl. Res. Persp.* 11, 71–101. doi: 10.1080/15366367.2013.831680

Maydeu-Olivares, A., García-Forero, C., Gallardo-Pujol, D., and Renom, J. (2009). Testing categorized bivariate normality with two-stage polychoric correlation estimates. *Methodology* 5, 131–136. doi: 10.1027/1614-2241.5.4.131

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychol. Methods* 23, 412–433. doi: 10.1037/met0000144

Meyer, P. (2010). *Understanding Measurement: Reliability*. Cambridge: Oxford University Press.

Millsap, R. E. (2011). *Statistical Approaches to Measurement Invariance*. New York, NY: Routledge. doi: 10.4324/9780203821961

Murray, J. S., Dunson, D. B., Carin, L., and Lucas, J. E. (2013). Bayesian Gaussian copula factor models for mixed data. *J. Am. Stat. Assoc.* 108, 656–665. doi: 10.1080/01621459.2012.762328

Murray, J. S., and Reiter, J. P. (2016). Multiple imputation of missing categorical and continuous values via Bayesian mixture models with local dependence. *J. Am. Stat. Assoc.* 111, 1466–1479. doi: 10.1080/01621459.2016.1174132

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* 49, 115–132. doi: 10.1007/BF02294210

Newman, D. A. (2014). Missing data: five practical guidelines. *Organ. Res. Methods* 17, 372–411. doi: 10.1177/1094428114548590

Nikoloulopoulos, A. K., and Joe, H. (2015). Factor copula models for item response data. *Psychometrika* 80, 126–150. doi: 10.1007/s11336-013-9387-4

Olsson, U. (1979). On the robustness of factor analysis against crude classifications of the observations. *Multivar. Behav. Res.* 14, 485–500. doi: 10.1207/s15327906mbr1404_7

Olsson, U., Foss, T., Troye, S. V., and Howell, R. D. (2000). The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Struct. Equat. Model.* 7, 557–595. doi: 10.1207/S15328007SEM0704_3

Peress, M. (2012). Identification of a semiparametric item response model. *Psychometrika* 77, 223–243. doi: 10.1007/s11336-012-9253-9

Pohl, S., and Becker, B. (2020). Performance of missing data approaches under nonignorable missing data conditions. *Methodology* 16, 147–165. doi: 10.5964/meth.2805

Raykov, T., and Marcoulides, G. A. (2015). On examining the underlying normal variable assumption in latent variable models with categorical indicators. *Struct. Equat. Model.* 22, 581–587. doi: 10.1080/10705511.2014.937846

Revuelta, J., Hidalgo, B., and Alcazar-Córcoles, M. Á. (2020). Bayesian estimation and testing of a beta factor model for bounded continuous variables. *Multivar. Behav. Res.* doi: 10.1080/00273171.2020.1805582. [Epub ahead of print].

Rhemtulla, M., Brosseau-Liard, P. É., and Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol. Methods* 17, 354–373. doi: 10.1037/a0029315

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02

Sass, D. A., Schmitt, T. A., and Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: a comparison of estimators. *Struct. Equat. Model.* 21, 167–180. doi: 10.1080/10705511.2014.882658

Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. *Sociol. Methodol.* 22, 249–278. doi: 10.2307/270998

Savalei, V. (2014). Understanding robust corrections in structural equation modeling. *Struct. Equat. Model.* 21, 149–160. doi: 10.1080/10705511.2013.824793

Savalei, V. (2020). Improving fit indices in structural equation modeling with categorical data. *Multivar. Behav. Res.* doi: 10.1080/00273171.2020.1717922. [Epub ahead of print].

Si, Y., and Reiter, J. P. (2013). Nonparametric Bayesian multiple imputation for incomplete categorical variables in large-scale assessment surveys. *J. Educ. Behav. Stat.* 38, 499–521. doi: 10.3102/1076998613480394

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika* 74, 107–120. doi: 10.1007/S11336-008-9101-0

Song, X.-Y., Pan, J.-H., Kwok, T., Vandenput, L., Ohlsson, C., and Leung, P.-C. (2010). A semiparametric bayesian approach for structural equation models. *Biometrical J.* 52, 314–332. doi: 10.1002/bimj.200900135

Steegen, S., Tuerlinckx, F., Gelman, A., and Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspect. Psychol. Sci.* 11, 702–712. doi: 10.1177/1745691616658637

Steyer, R. (1989). Models of classical psychometric test theory as stochastic measurement models: representation, uniqueness, meaningfulness, identifiability, and testability. *Methodika* 3, 25–60.

Svetina, D., Rutkowski, L., and Rutkowski, D. (2020). Multiple-group invariance with categorical outcomes using updated guidelines: an illustration using

Mplus and the lavaan/semtools packages. *Struct. Equat. Model.* 27, 111–130. doi: 10.1080/10705511.2019.1602776

Takane, Y., and de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika* 52, 393–408. doi: 10.1007/BF02294363

Tran, U., and Formann, A. (2010). *IRT Modelling of Dichotomous Items With Linear Factor Analysis.* Technical report. doi: 10.2139/ssrn.2408956

Vermunt, J. K., van Ginkel, J. R., van der Ark, L. A., and Sijtsma, K. (2008). Multiple imputation of incomplete categorical data using latent class analysis. *Sociol. Methodol.* 38, 369–397. doi: 10.1111/j.1467-9531.2008.00202.x

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–25. doi: 10.2307/1912526

Yang, Y., and Green, S. B. (2011). Coefficient alpha: a reliability coefficient for the 21st century? *J. Psychoeduc. Assess.* 29, 377–392. doi: 10.1177/0734282911406668

Yang-Wallentin, F., Jöreskog, K. G., and Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Struct. Equat. Model.* 17, 392–423. doi: 10.1080/10705511.2010.489003

Yuan, K.-H. (2009). Normal distribution based pseudo ML for missing data: with applications to mean and covariance structure analysis. *J. Multivar. Anal.* 100, 1900–1918. doi: 10.1016/j.jmva.2009.05.001

Yuan, K.-H., and Bentler, P. M. (2007). "Robust procedures in structural equation modeling," in *Handbook of Latent Variable and Related Models*, ed S. Y. Lee (North-Holland: Elsevier), 367–397. doi: 10.1016/B978-044452044-9/50020-3

Yuan, K.-H., and Bentler, P. M. (2010). Consistency of normal-distribution-based pseudo maximum likelihood estimates when data are missing at random. *Am. Stat.* 64, 263–267. doi: 10.1198/tast.2010.09203

Yuan, K.-H., Yang-Wallentin, F., and Bentler, P. M. (2012). ML versus MI for missing data with violation of distribution conditions. *Sociol. Methods Res.* 41, 598–629. doi: 10.1177/0049124112460373

Zhang, J., Li, J., and Liu, C. (2014). Robust factor analysis using the multivariate *t*-distribution. *Stat. Sin.* 24, 291–312. doi: 10.5705/ss.2012.342

Zhang, J., and Reiser, M. (2015). "A continuous latent factor model for non-ignorable missing data," in *Innovative Statistical Methods for Public Health Data*, eds D. G. Chen and J. Wilson (New York, NY: Springer), 173–199. doi: 10.1007/978-3-319-18536-1_9