

The Use of Items and Item Parcels in Nonlinear Structural Equation Models

Karina Rdz-Navarro ^{ab} , Rodrigo A. Asún ^a 

[a] Faculty of Social Sciences, University of Chile, Chile. [b] Faculty of Psychology, University of Talca, Chile.

Methodology, 2020, Vol. 16(1), 1–20, <https://doi.org/10.5964/meth.2305>

Received: 2017-12-12 • Accepted: 2019-09-17 • Published (VoR): 2020-04-06

Corresponding Author: Karina Rdz-Navarro, Faculty of Social Sciences, University of Chile. Av. Ignacio Carrera Pinto 1045, Nunoa, Santiago, Chile. E-mail: rdznavarro@uchile.cl

Abstract

Nonlinear structural equation models within the frequentist framework were developed to work with continuous items. Applied researchers who usually work with Likert-type items choose between two strategies to estimate such models: treat items as continuous variables or create item parcels. Two Monte Carlo studies were conducted to evaluate the effects of each strategy on estimates and Type I errors for models with interaction and quadratic effects estimated using LMS. The first study evaluated the effect of asymmetry type and item quantity. The second assessed the use of item parcels and parcel configuration under equivalent conditions. Results reveal that treating items as continuous variables is not problematic when item categories are symmetrical or have opposite-direction asymmetries; however, meaningful parameter bias and increased Type I errors are produced in the case of same-direction asymmetry. Use of parcels does not overcome these problems. The results are discussed to provide recommendations for applied researchers.

Keywords

item parcels, interaction effects, quadratic effects, nonlinear SEM, latent moderated structural equations

Researchers are often interested in using Structural Equation Models (SEM) to assess the nonlinear relationships between latent variables (e.g., Graves, Sarkis, & Zhu, 2013; Jackson, 2015; Hardy et al., 2013; Masland & Lease, 2016). In these applied research scenarios, latent variables are usually measured by items with discrete response categories whose level of measurement is ordinal at best (Michell, 2009). However, frequentist nonlinear SEM modeling techniques currently available (e.g., Kelava & Brandt, 2009; Klein & Moosbrugger, 2000; Marsh, Wen, & Hau, 2004) have been developed to work with factors measured by continuous observed variables rather than categorical variables.

Although some researchers (e.g., Jamieson, 2004; Norman, 2010) sustain that many statistical tools are robust to ordinal data, evidence indicates that overlooking the cat-



egorical nature of data produces invalid results and conclusions (Bernstein & Teng, 1989; DiStefano, 2002), unless item categories are symmetrical or exhibit low levels of asymmetry (Asún, Rdz-Navarro, & Alvarado, 2016). Measurement modeling procedures—such as item factor analysis—and linear SEM allow most of the problems produced by category asymmetry to be overcome. This is accomplished by replacing mean vectors with thresholds, and variance-covariance matrices with polychoric or tetrachoric correlation matrices used to estimate model parameters (Rhemtulla, Brosseau-Liard, & Savalei, 2012).

In nonlinear SEM, estimation of the model is not possible using correlation matrices. This is because correlations can only capture linear relationships between variables. Furthermore, nonlinearity produces non-normal dependent variables (Kelava et al., 2011) and non-normal underlying response variables. The multivariate normality assumption of poly-tetrachoric correlations will therefore not hold true.

Given that frequentist nonlinear SEM procedures assume measurement models comprised of continuous indicators, applied researchers willing to estimate nonlinear SEM models need to decide between two options (Little, Rhemtulla, Gibson, & Schoemann, 2013): (a) to treat ordinal items as if they were continuous variables in the hope that this will not seriously distort the results; or (b) to create item parcels as a means of avoiding the potential problems caused by the use of ordinal variables. An argument in favor of the second option may be that parcels tend to approximate normal distributions better than isolated items do, and the fact that they have more categories means that they are closer to being a continuous distribution (Bandalos, 2002).

According to evidence gathered by our team, more than 200 applied research articles using nonlinear SEM models were published between 2011 and 2016 across the social sciences. Most of them use the Latent Moderated Structural equation method (LMS; Klein & Moosbrugger, 2000) to estimate the model, and all of them either treat categorical items as if they were continuous indicators (e.g., Jackson, 2015; Masland & Lease, 2016), or create parcels (e.g., Graves et al., 2013; Hardy et al., 2013) that are used as indicators of factors. Despite the popularity of these decisions, their impact on nonlinear estimates is still unknown.

The present study addresses this gap by evaluating the impact of treating items as continuous indicators and creating item parcels to estimate nonlinear structural models using the LMS method. It focuses on the effects of type and degree of category asymmetry, how parcels are configured, parameter bias, standard errors (*SE*), and Type I error rates of nonlinear effects. The decision to focus on Type I error for nonlinear effects is based on the fact that this represents false positives, which are traditionally considered more serious than Type II errors (Jackson, 2014). Inflated Type I error rates produce unnecessarily overparametrized models, and this compromises parsimony and replication of results (Ioannidis, 2005; Simmons, Nelson, & Simonsohn, 2011). Moreover, if Type I errors are not guaranteed at a given significance level (e.g., $\alpha = .05$) for a statistical

procedure, its statistical power will also be compromised (Agresti & Finlay, 1997), and researchers will not be able to distinguish between true and false effects. The present research will therefore focus on situations where structural models are linear in the population, and where the analysis model estimates interactions and/or quadratic effects.

Nonlinear SEM Modeling

Substantive theory in the social sciences often suggests the presence of relationships between latent variables that are nonlinear, such as interactions and U-shaped (quadratic) relationships. In these circumstances, researchers may estimate models with an interaction (MI), such as that presented in Equation 1, or models with simultaneous interaction and quadratic (MIQ) effects, such as that presented in Equation 2, where η is an endogenous latent variable predicted by two exogenous latent variables (ξ_1 and ξ_2). Here, α is a latent intercept, γ_1 and γ_2 are the linear slopes of the exogenous factors, and the parameters ω_{ij} represent the slopes of the multiplicative (nonlinear) effects of predictors on η . The term $\xi_1\xi_2$ represents a two-way interaction between predictors, ξ_1^2 and ξ_2^2 represent quadratic effects, and ζ represents a latent prediction error.

$$\text{MI: } \eta = \alpha + \gamma_1\xi_1 + \gamma_2\xi_2 + \omega_{12}\xi_1\xi_2 + \zeta \quad (1)$$

$$\text{MIQ: } \eta = \alpha + \gamma_1\xi_1 + \gamma_2\xi_2 + \omega_{12}\xi_1\xi_2 + \omega_{11}\xi_1^2 + \omega_{22}\xi_2^2 + \zeta \quad (2)$$

Estimation of MI or MIQ models involves major methodological problems that are different to those found in linear SEM. On the one hand, dealing with nonlinearity implies inherent non-normality, because even when exogenous factors follow a standard normal distribution, their products ($\xi_1\xi_2$, ξ_1^2 and ξ_2^2) will not be normal, nor will their means be equal to zero (Aiken & West, 1991). Moreover, if at least one nonlinear effect in the structural model is not equal to zero, the endogenous latent variable (η) will depart from normality (Kelava et al., 2011). This violates the multivariate normality assumption of most estimation procedures. On the other hand, it is not possible to estimate nonlinear models simply by using sample means and variance-covariance matrices (or correlation matrices), because they only capture linear relationships between variables.

Various methods have been proposed for estimation of nonlinear SEM models. They can be classified into four modeling frameworks: (a) the so-called product-indicator approaches (e.g., Kelava & Brandt, 2009; Kenny & Judd, 1984; Marsh et al., 2004); (b) distribution analytic methods (Klein & Moosbrugger, 2000; Klein & Muthén, 2007); (c) the method of moments approach (Mooijaart & Bentler, 2010; Wall & Amemiya, 2003); and (d) Bayesian methods (e.g., Lee, Song, & Tang, 2007).

All of these methods are designed to work with latent variables measured by continuous items. Each item is defined as a linear combination of the latent construct and measurement error (δ_i), as shown in Equations 3 and 4.

$$X_i = \mu_{0i} + \lambda_{ij}\xi_j + \delta_i \quad (3)$$

$$Y_i = \mu_{0i} + \lambda_i\eta + \varepsilon_i \quad (4)$$

In product-indicator approaches, estimation requires the creation of products of observed variables to represent nonlinear terms (Marsh et al., 2004). Model parameters are estimated using maximum likelihood estimation. Because creation of products means using indicators more than once, correlations between error terms and other constraints need to be specified (Kelava & Brandt, 2009). This makes its use highly error-prone, especially as the number of indicators, factors, and/or nonlinear effects increases. In addition, there is evidence that product-indicator approaches yield biased results when applied to congeneric indicators (Rdz-Navarro & Alvarado, 2015) typically found in applied research scenarios.

The method of moments and Bayesian methods approaches do not require the creation of products and constraints. However, they are complex, and discussion of their properties is highly technical (e.g., Mooijaart & Bentler, 2010; Wall & Amemiya, 2003). This has probably undermined their use in applied research. Indeed, in our review, we did not find a single research article that made use of these methods in applied research in the social and behavioral sciences. By contrast, we found that the distribution analytic procedure LMS (Klein & Moosbrugger, 2000) has become the most popular nonlinear SEM method among applied researchers in the social sciences.

Nonlinear SEM Using the LMS Method

What distinguishes LMS from other methods is the way in which model parameters are estimated. When latent predictors and model errors are normally distributed and the population model is linear, the distribution of η will also be normal. By contrast, when at least one nonlinear effect is not equal to zero, the non-normal distribution of latent products will be reflected by the distribution of η no longer being normal. This allows LMS to attempt to explain any departure from a normal distribution of η as the result of a nonlinear effect of exogenous predictors (Klein & Moosbrugger, 2000).

Under this assumption, LMS uses the Cholesky decomposition to split the distribution of η into its linear (normal) and nonlinear (non-normal) parts, and to represent both as a finite mixture of weighted normal distributions with different means and variances (for technical details, see Klein & Moosbrugger, 2000). Model parameters are obtained using robust Maximum Likelihood Estimation (i.e., MLR). LMS is readily implemented in Mplus (Muthén & Muthén, 1998-2012).

The Impact of Non-Normal and Categorical Items on LMS

It has been demonstrated that LMS yields unbiased, efficient and consistent parameter estimates when the normality assumption of predictors is true (Jackman, Leite, &

Cochrane, 2011; Kelava et al., 2011; Rdz-Navarro & Alvarado, 2015). However, because of the strong dependence of LMS on such an assumption, its properties may not remain true when predictors are not normal (Brandt, Kelava, & Klein, 2014). Evidence indicates that in the presence of non-normal latent predictors and non-normal continuous items, LMS produces biased nonlinear parameters, and inflated Type I errors for interaction and quadratic effects (Brandt et al., 2014; Cham, West, Ma, & Aiken, 2012; Wu, Wen, Marsh, & Hau, 2013).

Although this evidence points to limitations of LMS when running the analysis on non-normal exogenous factors and items, it is unclear whether such negative results are explained by having non-normal latent factors or non-normal items. Indeed, the key assumption of LMS is that exogenous factors and model errors are normally distributed (Klein & Moosbrugger, 2000). When this is the case, and items measuring each factor are continuous, items will also be normal. However, in real life applications, items may depart from normality for reasons other than non-normality of the latent factors. This will be the case when items are answered using discrete categories (k) coded with integer values (i.e., 1, 2, ..., k). In such situations, item distribution does not depend on factor distributions, but on the distribution of thresholds that define the limits between response categories, which in turn produce variables whose measurement levels are ordinal at best (Michell, 2009). Therefore, even when the factor normality assumption of LMS is not violated, categorical items may not reflect such a distribution and, to our knowledge, the consequences of this situation for nonlinear SEM estimates using LMS have not been studied.

In applied research, categorical items are often treated as continuous variables, although this practice is controversial. Some authors argue that ordinal variables can always be treated as continuous (e.g., Norman, 2010); others maintain that items can be treated as if they were continuous if specific conditions are met (e.g., Bollen & Barb, 1981); while others categorically deny this possibility (e.g., Jamieson, 2004). Nevertheless, evidence reveals that treating items as continuous variables produces spurious factors, attenuated variance-covariance matrices (Muthén & Kaplan, 1985), and parameter bias, especially when items have fewer than five response categories and/or item skewness is greater than $|1.0|$ (Asún et al., 2016; Bernstein & Teng, 1989; DiStefano, 2002; Rhemtulla, Brosseau-Liard, & Savalei, 2012).

Although treating categorical items as continuous variables is rather common in applied research that uses nonlinear SEM (e.g., Graves et al., 2013; Jackson, 2015; Hardy et al., 2013; Masland & Lease, 2016), the consequences of this practice are still unknown. Nevertheless, it is not unreasonable to hypothesize that such treatment of items would be problematic, especially when categories are asymmetrical.

Parcels as a Possible Solution

The recommendation to use item parcels as indicators of latent variables was made during the initial debate among experts as to the suitability of their application (for a summary of this controversy, cf., [Little, Cunningham, Shahar, & Widaman, 2002](#)).

The main arguments in favor of using parcels (a summary can be found in [Little et al., 2013](#), Table 3, p. 393) are based on the fact that they: (a) tend to approximate normal distributions better than isolated items do ([Bandalos, 2002](#)); (b) have more categories than isolated items do, such that they are close to being a continuous distribution ([Hall, Snell, & Foust, 1999](#)); (c) are more reliable than individual items ([Marsh, Hau, Balla, & Grayson, 1998](#)); (d) reduce the number of parameters to be estimated and model complexity, thereby producing more stable estimations ([Little et al., 2013](#)), especially in small samples ([Hau & Marsh, 2004](#)); and (e) reduce the global model error-variance ([Little et al., 2013](#)). By contrast, detractors of parcels argue that they: (a) may distort the dimensional structure of the data ([Bandalos, 2002](#)); (b) mask specification errors in the model ([Rogers & Schmitt, 2004](#)); (c) constitute a modification of the data which contaminates the results by the researcher's intervention; and (d) distort the metric of the scale that would be obtained if working directly with the items, possibly deforming some interpretations based on the total score distributions ([Little et al., 2002](#)).

Although research using simulated data (e.g., [Bandalos, 2002](#); [Hall et al., 1999](#); [Hau & Marsh, 2004](#); [Marsh et al., 1998](#)) shows that parcels have small or negligible effects on parameter recovery, it has been argued that their potentially positive effect depends on the manner in which they are constructed ([Little et al., 2002](#)) and on the context in which they are used. Discussions concerning the advantages and disadvantages of using parcels can also be found in some studies of nonlinear SEM models (e.g., [Jackman et al., 2011](#); [Wu et al., 2013](#)) that have identified neither positive nor negative effects of their use. These studies have focused on parcels of continuous items used to measure factors in SEM models that estimate interactions. It is not clear whether parcels of categorical items could produce positive or negative results for nonlinear SEM models, and whether parcels could work on models that estimate quadratic effects. This investigation will evaluate the performance of parcels in nonlinear SEM by implementing two simple alternatives for creating parcels ([Hau & Marsh, 2004](#)): counterbalancing and not counterbalancing category asymmetry within the parcel, when category asymmetries have opposite directions.

Simulation Studies

Two studies were carried out to assess the consequences of estimating nonlinear SEM models with the LMS method, firstly by treating categorical items as continuous indicators (Study 1), and secondly by using item parcels (Study 2). Parameter and *SE* biases,

as well as Type I error rates in the detection of nonlinear effects were assessed in both studies.

In order to assess Type I error, data were generated for each study using the model in Equation 2, setting linear parameters as $\gamma_1 = \gamma_2 = .3$, and nonlinear parameters equal to zero (i.e., $\omega_{12} = \omega_{11} = \omega_{22} = 0$). Latent predictors (ξ_1 and ξ_2) were created from an $N(0, 1)$ and covariance equal to $.3$. Prediction error (ζ) was simulated from an $N(0, 0.766)$ distribution such that η had a variance equal to one.

The endogenous factor η was measured by a single indicator with no measurement error (i.e., $\eta = Y$). The exogenous factors were measured with multiple items created in two steps. First, continuous items (X_i) were generated for each factor according to a simple structure (i.e., cross-loadings = 0) and the model in Equation 3. For simplicity, the factor loadings were set to $.5$, a value which has shown reasonable results in previous studies (Rdz-Navarro & Alvarado, 2015). The measurement errors (δ_i) were generated from an $N(0, 0.75)$ distribution, such that all X_i follow an $N(0, 1)$ distribution. Conditions were generated with four, eight, and 16 items per factor.

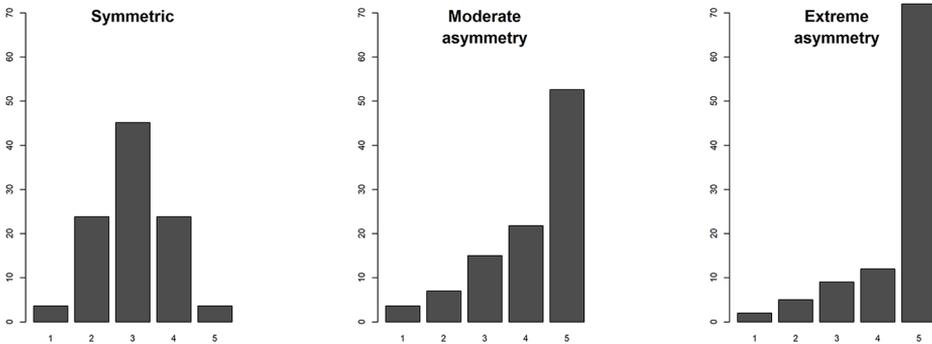
In the second step, continuous items (X_i) were transformed into categorical items (x_i). As with previous research (e.g., Rhemtulla et al., 2012), transformation was carried out by choosing four cutting points (i.e., thresholds) that yield five response categories to represent Likert-type items with different distributions, as shown in Figure 1. In symmetry conditions, thresholds were distributed symmetrically around the zero-mean of all X_i (i.e., thresholds were $-1.8, -0.6, 0.6,$ and 1.8). In asymmetry conditions, thresholds were selected such that the peak of the distribution was the highest response category. In moderate asymmetry conditions, threshold values had a mean equal to -0.942 (i.e., thresholds were $-1.799, -1.248, -0.656,$ and -0.065). In extreme asymmetry conditions, threshold values had a mean equal to -1.277 (i.e., thresholds were $-2.054, -1.476, -0.994,$ and -0.583).

Two additional conditions were created to represent moderate asymmetry-alternating and extreme asymmetry-alternating situations. Here, threshold values were the same as those used in asymmetry conditions, with the exception that the threshold sign was reversed for half of the items that measured a given factor¹. Thus, for example, in the moderate asymmetry-alternating condition with four categorical items, the first two categorical items of the factor were created using thresholds with negative values (i.e., $-1.799, -1.248, -0.656,$ and -0.065), and the other two were created using thresholds with positive values ($0.065, 0.656, 1.248$ and 1.799). This meant that half of the items peaked in the highest response category and the other half in the lowest response category.

1) Simulation code is available to researchers upon email request.

Figure 1

Simulated Items According to Their Distributions.



The second study evaluated the performance of item parcels. Data were created under conditions equivalent to Study 1, although in this case, after categorical items were created, they were used to form parcels comprising two or four items. The specialized literature recommends a minimum of three or four parcels per factor (Marsh et al., 1998), because a single factor model is not identified (i.e., degrees of freedom are less than zero) when only two indicators are used. Thus, because four-item conditions only allow creation of two two-item parcels, this setting was discarded from the analysis. The same rationale was used for other configurations, eventually leaving only two-parcels per factor (i.e., two four-item parcels and two eight-item parcels). Thus, in eight-item conditions, four two-item parcels were created, and under sixteen-item conditions, four four-item parcels and eight two-item parcels were created. In the asymmetry-alternating conditions, the items of each parcel were grouped in two ways: counterbalanced within the parcel (i.e., items with opposite-direction asymmetries grouped within each parcel) or non-counterbalanced within the parcel (i.e., items with same-direction asymmetries grouped in each parcel). In the symmetry and asymmetry conditions, it was only possible to group parcels in a non-counterbalanced way. Given that λ_i was kept at .5, population factor loadings of two-item parcels were equal to .632, and .756 for four-item parcels. Mean skewness and excess kurtosis of item parcels for each simulated condition are displayed in Table 1.

Table 1*Skewness and Excess Kurtosis of Parcels*

Item category distribution	TP	Four Two-item		Eight Two-item		Four Four-item	
		SK	KU	SK	KU	SK	KU
Symmetry	NCB	-0.001	-0.151	0.001	-0.154	0.001	-0.116
Moderate asymmetry							
Same-direction	NCB	-0.918	0.360	-0.915	0.347	-0.820	0.347
Alternating	NCB	0.028	0.370	0.025	0.365	0.025	0.380
Alternating	CB	0.048	-0.093	0.048	-0.093	0.040	-0.148
Extreme asymmetry							
Same-direction	NCB	-1.542	2.045	-1.539	2.035	-1.379	1.871
Alternating	NCB	0.003	2.031	0.000	2.029	0.000	1.855
Alternating	CB	-0.001	0.920	0.000	0.915	0.001	0.575

Note. TP = type of parcel. SK = skewness. KU = excess kurtosis. NCB = non-counterbalanced parcel. CB = counterbalanced parcel.

Samples of 1,000 subjects were used, and 500 replicates were created for each condition in both studies. The data were analyzed with two types of nonlinear model: the MI model in Equation 1, and the MIQ model in Equation 2. Analyses were run in Mplus 7 (Muthén & Muthén, 1998-2012) using the LMS method. Results were considered acceptable upon meeting the following conditions: (a) 80% or more replicates produced convergent and admissible solutions (Forero & Maydeu-Olivares, 2009); (b) relative bias of linear parameters was equal to or less than $|0.05|$ (Hoogland & Boomsma, 1998); and (c) relative bias of *SE* was equal to or less than $|0.10|$. Given that relative bias of nonlinear parameters is not defined in this case because the population parameter is zero, the mean of nonlinear parameter estimates was assessed instead. No standard evaluation criterion is available to assess this mean, so an ad-hoc criterion of values less than or equal to $|0.025|$ was used². Following Bradley's liberal criterion (Serlin, 2000), Type I errors between 2.5% and 7.5% were considered adequate at a 95% confidence level.

2) This cut-off may sound too restrictive; however, given that the true parameter is exactly zero, almost any departure from that result could be considered unacceptable. The rationale behind this is the following. Relative bias is defined as $(\text{Parameter} - \text{Mean}(\text{Estimates})) / \text{Parameter}$. When the parameter is zero, the formula is not defined. However, if we replace the population value with any number as close to zero as possible, say 0.0001 or 0.001 to calculate the bias, obtaining a mean of estimates equal to $|0.025|$ will result in a relative bias of 250.0 (25,000%) and 25.0 (2,500%), respectively.

Study 1

Treating Categorical Items as Continuous

The first study assessed the impact of treating categorical items as if they were continuous. All 30 research conditions yielded convergent and admissible results. The following analysis will focus on parameter and *SE* recovery, and Type I error rates, as shown in Table 2.

Table 2

Parameter Estimates (Averages) and Type I Error Rates Using Items Treated as Continuous Indicators

	MI					MIQ								
	Parameters				%Sig	Parameters					%Sig			
TI	λ_i	γ_1	γ_2	ω_{12}	ω_{12}	λ_i	γ_1	γ_2	ω_{12}	ω_{11}	ω_{22}	ω_{12}	ω_{11}	ω_{22}
4 items														
SI	.472	.298	.301	.000	4.4	.472	.298	.302	-.001	.001	.000	6.2	4.2	6.4
M	.452	.301	.301	.040	15.4	.452	.316	.315	-.026	.044	.045	7.6	15.2	18.2
E	.409	.300	.297	.058	20.0	.410	.327	.324	-.020	.051	.052	6.8	21.2	19.8
M _A	.436	.298	.301	.004	5.0	.436	.299	.302	.002	.000	.003	4.8	5.6	6.8
E _A	.379	.305	.304	.000	7.0	.379	.308	.308	.002	-.001	-.001	6.2	5.8	7.4
8 items														
SI	.471	.302	.298	.001	5.0	.471	.302	.298	.000	-.001	.001	6.4	4.8	3.8
M	.453	.302	.302	.040	20.8	.453	.319	.319	-.012	.041	.041	7.2	21.6	19.0
E	.409	.298	.304	.059	32.0	.409	.334	.345	-.015	.055	.059	7.6	34.6	36.0
M _A	.439	.303	.301	.000	6.6	.439	.303	.301	.002	-.002	-.001	5.0	4.0	4.4
E _A	.384	.305	.300	.000	5.6	.384	.307	.301	.002	-.001	-.001	6.0	6.2	5.2
16 items														
SI	.472	.299	.302	.001	4.8	.472	.299	.302	-.002	.003	.001	5.0	4.2	5.4
M	.453	.302	.300	.042	28.4	.453	.332	.319	-.002	.039	.039	5.0	27.4	26.2
E	.410	.304	.302	.061	45.4	.410	.345	.344	.003	.050	.050	4.0	45.0	43.4
M _A	.440	.303	.300	-.001	4.0	.440	.303	.300	-.002	.001	.001	3.6	3.4	5.2
E _A	.386	.300	.302	.000	5.0	.386	.301	.302	.001	-.001	-.001	5.6	6.0	4.6

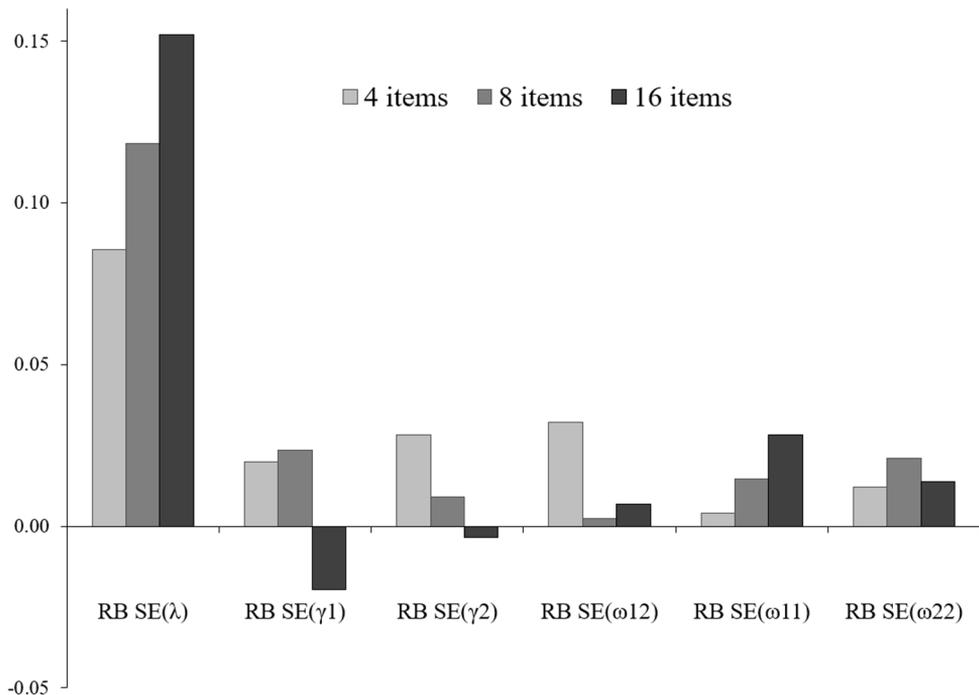
Note. MI = model with one interaction. MIQ = model with one interaction and two quadratic terms. %Sig = percentage of significant nonlinear effects (Type I error). TI = type of item distribution. SI = symmetrical items. M = moderate asymmetry. M_A = moderate asymmetry-alternating. E = extreme asymmetry. E_A = extreme asymmetry-alternating. Unacceptable results are in bold. Population parameters: $\lambda_i = .5$, $\gamma_1 = \gamma_2 = .3$, $\omega_{12} = \omega_{11} = \omega_{22} = 0$.

Factor loadings were underestimated in all conditions regardless of the analysis model. Bias was greater when item categories exhibited greater asymmetry, especially in

asymmetry-alternating conditions. Increasing the number of items slightly decreased the magnitude of bias for these parameters. The factor loading *SEs* were systematically overestimated, and such biases increased with the number of items, as shown in Figure 2. It should be noted that LMS uses MLR estimation, meaning that *SE* bias found here is not the result of an incorrect estimator, but a problem resulting from the treatment of categorical data as continuous.

Figure 2

Relative Bias of the Standard Errors Using Categorical Items Treated as Continuous



Note. RB = Relative bias; SE = Standard error.

Regarding structural model parameters, low biases and acceptable Type I error rates were observed for linear and nonlinear parameters in symmetry and asymmetry-alternating conditions. Under asymmetry conditions, linear effects were unbiased, but interaction effects and Type I errors increased for the MI analysis model. For the MIQ analysis model, linear effects were also affected, nonlinear parameters were overestimated, and Type I errors were severely inflated. Bias was greater for linear and quadratic estimates

than for interactions. This may indicate that the overestimation of interaction effects observed in the MI analysis was transferred to the quadratic effects when using the MIQ model. Under these conditions, use of more items per factor and higher threshold asymmetry levels, seemed to increase the magnitude of bias. The structural model *SEs* of all parameters were recovered with acceptable levels of bias.

Conclusions

Treatment of categorical items as continuous indicators in nonlinear SEM models estimated using LMS tends to generate estimation problems for the different parameters of the model. In measurement models, this treatment produces underestimation of the factor loadings and overestimation of *SEs*. In the structural model, it generates overestimation of nonlinear parameters and increases in the Type I error when items are asymmetrical. A set of exploratory simulations conducted to cross-validate these results enabled us to establish that using items with three, four or seven response categories generates results equivalent to those reported here when item asymmetry levels are also equivalent to those examined here. It may therefore not be the number of response categories that produces bias, but the threshold asymmetry that is not accounted for by the model. This is further supported by the fact that treatment of categorical items as continuous indicators seems unproblematic for the structural model when threshold distributions are symmetrical or alternating.

Study 2

The Impact of Working With Item Parcels

The second study evaluated the performance of item parcels and their ability to solve the problems detected. As in the first study, no convergence or admissibility problems were found. The research results presented in [Table 3](#) show that use of parcels does not solve any of the problems detected in Study 1.

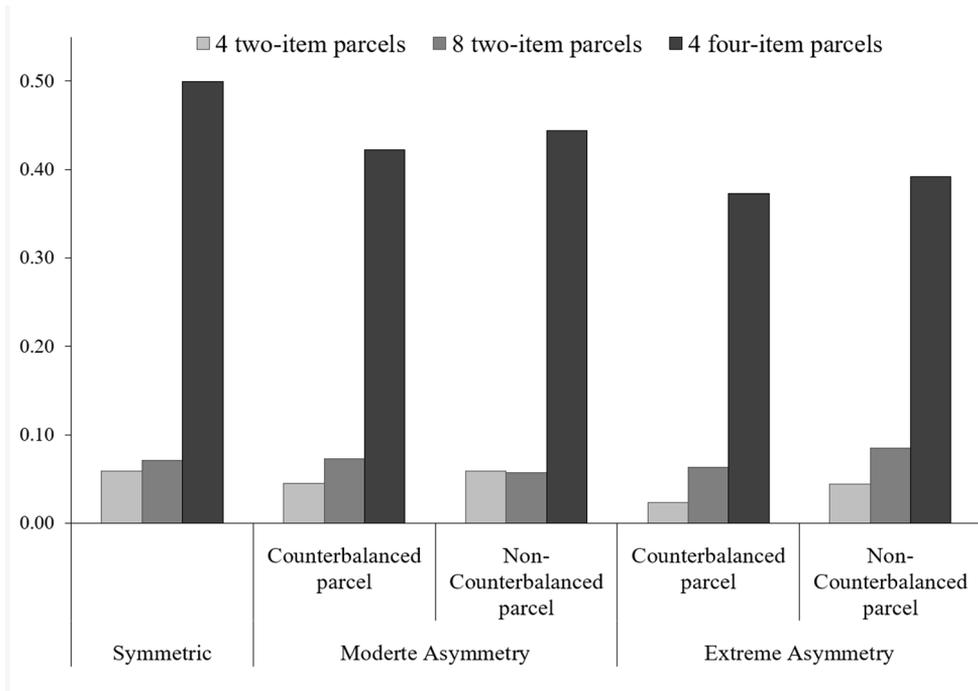
The parcel factor loadings were underestimated for all conditions. Small bias was found when the parcels comprised symmetrical items. Bias increased for parcels generated from more asymmetrical items. Counterbalancing item asymmetry within the parcels partially compensated for the underestimation of factor loadings. Unbiased factor loading *SEs* were found when four two-item parcels or eight two-item parcels were used (see [Figure 3](#)). However, when four four-item parcels were used, factor loading *SEs* displayed severe bias.

Table 3

Parameter Estimates (Averages) and Type I Error Rates Using Parcels

Np/Nip	MI					MIQ									
	Parameters				%Sig	Parameters					%Sig				
	λ_p	γ_1	γ_2	ω_{12}	ω_{12}	λ_p	γ_1	γ_2	ω_{12}	ω_{11}	ω_{22}	ω_{12}	ω_{11}	ω_{22}	
Type of item: SI															
4/2	.603	.302	.298	.000	5.0	.603	.302	.298	-.001	.000	.001	5.8	4.8	4.6	
4/4	.731	.299	.302	.001	5.6	.731	.299	.302	-.002	.003	.001	5.8	4.4	5.4	
8/2	.604	.299	.302	.001	5.2	.604	.299	.302	-.002	.003	.001	5.0	4.2	5.4	
Type of item: M / Parcel configuration: NCB															
4/2	.584	.302	.302	.040	19.6	.584	.318	.319	-.011	.040	.040	7.6	21.4	18.6	
4/4	.713	.303	.299	.041	28.0	.713	.321	.318	-.002	.039	.039	5.8	27.2	25.4	
8/2	.583	.302	.299	.041	28.0	.583	.321	.318	-.002	.039	.039	5.6	26.8	25.8	
Type of item: M _A / Parcel configuration: NCB															
4/2	.564	.303	.302	.000	6.6	.564	.304	.302	.002	-.002	-.002	4.8	5.4	4.0	
4/4	.690	.304	.301	-.001	4.0	.690	.305	.301	-.002	.002	.001	4.4	4.0	4.6	
8/2	.566	.304	.300	-.001	4.0	.566	.304	.300	-.002	.002	.001	4.2	3.6	5.2	
Type of item: M _A / Parcel configuration: CB															
4/2	.573	.302	.300	-.002	7.0	.573	.302	.301	.004	-.005	-.003	5.2	5.2	4.8	
4/4	.702	.303	.299	-.002	3.8	.702	.303	.299	-.002	.000	-.001	4.2	3.4	5.4	
8/2	.572	.303	.299	-.002	4.2	.572	.303	.299	-.001	.000	-.001	4.2	4.0	5.4	
Type of item: E / Parcel configuration: NCB															
4/2	.536	.298	.303	.059	31.6	.536	.333	.343	-.014	.054	.058	7.4	33.6	35.4	
4/4	.668	.304	.302	.060	45.0	.668	.344	.342	.003	.050	.050	4.0	44.8	43.6	
8/2	.536	.304	.302	.060	46.0	.536	.344	.343	.003	.050	.050	4.0	44.4	42.2	
Type of item: E _A / Parcel configuration: NCB															
4/2	.498	.307	.302	.000	5.8	.498	.310	.303	.003	-.001	-.002	5.4	6.4	6.2	
4/4	.619	.303	.304	-.001	4.6	.619	.304	.306	.001	-.002	-.001	4.8	6.0	4.8	
8/2	.503	.301	.303	.000	4.6	.503	.302	.303	.001	-.001	.000	5.0	5.8	4.8	
Type of item: E _A / Parcel configuration: CB															
4/2	.516	.304	.298	.000	6.6	.516	.304	.298	.002	.000	-.002	6.6	5.4	4.8	
4/4	.647	.299	.301	.000	3.8	.647	.299	.301	.001	-.001	.000	5.0	5.4	4.0	
8/2	.515	.299	.301	.000	4.2	.515	.299	.301	.001	-.001	.000	5.4	5.0	3.4	

Note. MI = model with one interaction. MIQ = model with one interaction and two quadratic terms. %Sig = percentage of significant nonlinear effects (Type I error). SI = symmetrical items. M = moderate asymmetry. M_A = moderate asymmetry-alternating. E = extreme asymmetry. E_A = extreme asymmetry-alternating. NCB = non-counterbalanced parcel. CB = counterbalanced parcel. Np/Nip = number of parcels created / number of items within each parcel. λ_p = parcel factor loading. Unacceptable results are in bold. Population parameters: $\gamma_1 = \gamma_2 = .3$, $\omega_{12} = \omega_{11} = \omega_{22} = 0$. Two-item $\lambda_p = .632$. Four-item $\lambda_p = .756$.

Figure 3*Relative Bias of Parcel Factor Loading Standard Errors*

Linear structural parameters were estimated with negligible bias when using an MI analysis model, but the tendency to obtain overestimated interaction parameters remained when parcels comprised same-direction asymmetry items. This produced inflation of Type I errors for the interaction. Upon analyzing the data with an MIQ model, severe overestimation of linear and nonlinear parameters, as well as a strong increase in Type I errors was observed for extreme asymmetry item parcels. No problems were observed in the recovery of the SEs of any structural model parameters. The manner of building the parcels (i.e., counterbalancing or non-counterbalancing for item asymmetries within the parcels) had no noticeable effect on the results.

Conclusions

Use of item parcels does not generate additional problems beyond those noted when dealing with items as continuous indicators; however, parcels does not offer a solution to the problems detected in Study 1. Indeed, contrary to our hypothesis, neither the

number of parcels nor the number of items forming each parcel seem to affect the results. The limited impact of using parcels may be due to the fact that only items with same-direction asymmetry were available for the conditions in which problems were observed. This obstacle is not eliminated by the use of parcels, as their scores retain an important part of this asymmetry. In alternating-asymmetry situations, counterbalancing asymmetry within the parcels did not offer an improvement in estimation compared to isolated items.

Given that problems observed when treating asymmetrical items as continuous indicators are not solved by the use of parcels, they do not appear to be an advisable alternative in these situations, and it may be presumed that other types of parcel configurations (e.g., a smaller or larger number of parcels, or using parcels comprised of a smaller or larger number of items) would produce results equivalent to those presented here.

General Discussion and Conclusions

Based on the results of this investigation, it can be asserted that treating categorical items as continuous indicators in nonlinear SEM using LMS does not seem to be problematic when items are symmetrical. However, even in this best-case scenario, this approach will produce underestimated factor loadings which might lead the researcher to believe that the items are of a lower quality than they actually are. Despite this, treatment of categorical items as continuous was not found to produce negative consequences for structural model parameter estimates when item thresholds are symmetrical or have alternating asymmetry, confirming previous studies (e.g., Rhemtulla et al., 2012). However, when item category distributions have same-direction asymmetry, treating them as continuous variables produces overestimated nonlinear effects. Such bias increases Type I errors, especially when larger tests or scales are used.

The bias problem detected in asymmetrical conditions remains unsolved when working with item parcels. The results confirm that while parceling does not generate further problems, as had been reported in previous literature (Hau & Marsh, 2004; Jackman et al., 2011; Wu et al., 2013), it does not produce additional benefits either. Use of parcels does not, therefore, appear to be an acceptable solution to the problems derived from threshold asymmetry. It is true that nonlinear SEM procedures able to handle non-normal data have been proposed within the frequentist framework (e.g., Brandt et al., 2014; Cham et al., 2012); however, they assume the presence of continuous items that are non-normal, because they belong to a factor that is not normal either. Because the distribution of categorical items depends on thresholds and not on the factors themselves, further research is needed to examine whether these procedures that are capable of handling non-normality in nonlinear SEM could also solve the problems encountered here.

The development of nonlinear SEM procedures able to handle categorical data is a challenging task, as estimation must consider two sources of nonlinearity at the same

time: nonlinearity in measurement models (due to categorical data), and nonlinearity in the structural model (due to the relationship between latent variables). Full development of such a methodology may take some time, although a number of proposals have emerged within Bayesian nonlinear SEM (Lee, Song, & Cai, 2010; Lee & Zhu, 2000). Evidence to date reveals that such methods yield unbiased linear and nonlinear parameter estimates when factors are measured with dichotomous or polytomous items (Lee et al., 2010). Although results look promising, the methodology is still under development. Parameter *SEs* show substantial bias, estimates are sensitive to prior misspecifications, and estimation requires sample sizes larger than those needed for continuous indicators.

Given all of the above, researchers should be aware that, in common applied research situations (e.g., items with moderate or large same-direction asymmetry), biased parameter estimates and inflated Type I error rates could be obtained as a consequence of item asymmetry in nonlinear models. This is particularly important given the fact that—to the best of our knowledge—current frequentist nonlinear SEM procedures assume that items are truly continuous. Therefore, researchers willing to fit nonlinear models using the LMS method (or any other method that assumes normality of latent predictors) should check the distribution of items before proceeding with the analysis to ensure the data set meets the conditions required for use of the method without jeopardizing the accuracy of results and statistical conclusions.

It should be noted that the findings presented here are restricted to situations where nonlinear effects are equal to zero in the population (i.e., Type I error conditions). Further studies are needed to evaluate whether these findings could be generalized to situations where true nonlinear effects exist in the population. This may be an important limitation of this study. However, a small simulation study (not reported here)—conducted as a validity evaluation under a subset of conditions equivalent to those in Study 1—revealed that the problem of overestimation bias of parameter estimates remains when interaction and/or quadratic effects exist in the population and the model is estimated using same-sign asymmetry items. Indeed, the trend of bias was comparable to that observed in Study 1 (i.e., bias increased with asymmetry), and the magnitude of bias was around 16% for true non-zero interaction and quadratic parameters. These results reinforce the idea that treating asymmetrical items as continuous variables in nonlinear models fitted using LMS is counterproductive, because even moderate asymmetries—which are usually not considered damaging (Rhemtulla et al., 2012)—may lead the researcher to believe that there is a nonlinear effect when in fact the effect is spurious, or that the nonlinear effects found are more important than they actually are due to overestimation bias.

Further studies are still required to evaluate the generalization of these results to other situations, such as those that involve non-normally distributed exogenous factors. As current research evidence (e.g., Brandt et al., 2014; Cham et al., 2012; Wu et al., 2013) reveals that, even with continuous items, problems may be worse for LMS when the

assumption of factor normality is not met, it is presumed that this could further affect work with categorical items.

Funding: This research was supported by the Chilean National Commission for Science and Technology (FONDECYT Project N°11160256) and the Chilean Ministry of Education (MECESUP Bicentenary Project, Faculty of Psychology, University of Talca).

Competing Interests: The authors have declared that no competing interests exist.

Acknowledgments: The authors have no support to report.

References

- Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences*. Upper Saddle River, NJ, USA: Prentice Hall.
- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. London, United Kingdom: Sage.
- Asún, R. A., Rdz-Navarro, K., & Alvarado, J. M. (2016). Developing multidimensional Likert scales using item factor analysis: The case of four-point items. *Sociological Methods & Research*, 45(1), 109-133. <https://doi.org/10.1177/0049124114566716>
- Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling*, 9(1), 78-102. https://doi.org/10.1207/S15328007SEM0901_5
- Bernstein, I. H., & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin*, 105(3), 467-477. <https://doi.org/10.1037/0033-2909.105.3.467>
- Bollen, K. A., & Barb, K. H. (1981). Pearson's r and coarsely categorized measures. *American Sociological Review*, 46(2), 232-239. <https://doi.org/10.2307/2094981>
- Brandt, H., Kelava, A., & Klein, A. (2014). A simulation study comparing recent approaches for the estimation of nonlinear effects in SEM under the condition of nonnormality. *Structural Equation Modeling*, 21(2), 181-195. <https://doi.org/10.1080/10705511.2014.882660>
- Cham, H., West, S. G., Ma, Y., & Aiken, L. S. (2012). Estimating latent variable interactions with nonnormal observed data: A comparison of four approaches. *Multivariate Behavioral Research*, 47(6), 840-876. <https://doi.org/10.1080/00273171.2012.732901>
- DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling*, 9(3), 327-346. https://doi.org/10.1207/S15328007SEM0903_2
- Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of IRT graded response models: Limited versus full information methods. *Psychological Methods*, 14(3), 275-299. <https://doi.org/10.1037/a0015825>

- Graves, L. M., Sarkis, J., & Zhu, Q. (2013). How transformational leadership and employee motivation combine to predict employee proenvironmental behaviors in China. *Journal of Environmental Psychology*, 35, 81-91. <https://doi.org/10.1016/j.jenvp.2013.05.002>
- Hall, R. J., Snell, A. F., & Foust, M. S. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods*, 2(3), 233-256. <https://doi.org/10.1177/109442819923002>
- Hardy, S. A., Francis, S. W., Zamboanga, B. L., Kim, S. Y., Anderson, S. G., & Forthun, L. F. (2013). The roles of identity formation and moral identity in college student mental health, health-risk behaviors, and psychological well-being. *Journal of Clinical Psychology*, 69(4), 364-382. <https://doi.org/10.1002/jclp.21913>
- Hau, K. T., & Marsh, H. W. (2004). The use of item parcels in structural equation modelling: Non-normal data and small sample sizes. *British Journal of Mathematical & Statistical Psychology*, 57(2), 327-351. <https://doi.org/10.1111/j.2044-8317.2004.tb00142.x>
- Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structural modeling: An overview and a meta-analysis. *Sociological Methods & Research*, 26(3), 329-367. <https://doi.org/10.1177/0049124198026003003>
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLOS Medicine*, 2(8), Article e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Jackman, M. G. A., Leite, W. L., & Cochrane, D. J. (2011). Estimating latent variable interactions with the unconstrained approach: A comparison of methods to form product indicators for large, unequal numbers of items. *Structural Equation Modeling*, 18(2), 274-288. <https://doi.org/10.1080/10705511.2011.557342>
- Jackson, J. (2015). Cognitive closure and risk sensitivity in the fear of crime. *Legal and Criminological Psychology*, 20(2), 222-240. <https://doi.org/10.1111/lcrp.12031>
- Jackson, S. L. (2014). *Research methods: A modular approach*. Boston, MA, USA: Cengage Learning.
- Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, 38(12), 1217-1218. <https://doi.org/10.1111/j.1365-2929.2004.02012.x>
- Kelava, A., & Brandt, H. (2009). Estimation of nonlinear latent structural equation models using the extended unconstrained approach. *Review of Psychology*, 16(2), 123-132.
- Kelava, A., Werner, C. S., Schermelleh-Engel, K., Moosbrugger, H., Zapf, D., Ma, Y., . . . West, S. G. (2011). Advanced nonlinear latent variable modeling: Distribution analytic LMS and QML estimators of interaction and quadratic effects. *Structural Equation Modeling*, 18(3), 465-491. <https://doi.org/10.1080/10705511.2011.582408>
- Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, 96(1), 201-210. <https://doi.org/10.1037/0033-2909.96.1.201>
- Klein, A. G., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, 65(4), 457-474. <https://doi.org/10.1007/BF02296338>
- Klein, A. G., & Muthén, B. O. (2007). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research*, 42(4), 647-673. <https://doi.org/10.1080/00273170701710205>

- Lee, S. Y., Song, X. Y., & Cai, J. H. (2010). A Bayesian approach for nonlinear structural equation models with dichotomous variables using logit and probit links. *Structural Equation Modeling*, 17(2), 280-302. <https://doi.org/10.1080/10705511003659425>
- Lee, S. Y., Song, X. Y., & Tang, N. S. (2007). Bayesian methods for analyzing structural equation models with covariates, interaction, and quadratic latent variables. *Structural Equation Modeling*, 14(3), 404-434. <https://doi.org/10.1080/10705510701301511>
- Lee, S. Y., & Zhu, H. T. (2000). Statistical analysis of nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical & Statistical Psychology*, 53(2), 209-232. <https://doi.org/10.1348/000711000159303>
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9(2), 151-173. https://doi.org/10.1207/S15328007SEM0902_1
- Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, 18(3), 285-300. <https://doi.org/10.1037/a0033266>
- Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, 33(2), 181-220. https://doi.org/10.1207/s15327906mbr3302_1
- Marsh, H. W., Wen, Z., & Hau, K. T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, 9(3), 275-300. <https://doi.org/10.1037/1082-989X.9.3.275>
- Masland, L. C., & Lease, A. M. (2016). Characteristics of academically-influential children: Achievement motivation and social status. *Social Psychology of Education*, 19(1), 195-215. <https://doi.org/10.1007/s11218-015-9314-x>
- Michell, J. (2009). The psychometricians' fallacy: Too clever by half? *British Journal of Mathematical & Statistical Psychology*, 62(1), 41-55. <https://doi.org/10.1348/000711007X243582>
- Mooijaart, A., & Bentler, P. M. (2010). An alternative approach for nonlinear latent variable models. *Structural Equation Modeling*, 17(3), 357-373. <https://doi.org/10.1080/10705511.2010.488997>
- Muthén, B. O., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical & Statistical Psychology*, 38(2), 171-189. <https://doi.org/10.1111/j.2044-8317.1985.tb00832.x>
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA, USA: Author.
- Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education : Theory and Practice*, 15(5), 625-632. <https://doi.org/10.1007/s10459-010-9222-y>
- Rdz-Navarro, K., & Alvarado, J. M. (2015). Reexamining nonlinear structural equation modeling procedures: The effect of parallel and congeneric measures. *Multivariate Behavioral Research*, 50(6), 645-661. <https://doi.org/10.1080/00273171.2015.1071236>

- Rhemtulla, M., Brosseau-Liard, P. É., & 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*(3), 354-373. <https://doi.org/10.1037/a0029315>
- Rogers, W. M., & Schmitt, N. (2004). Parameter recovery and model fit using multidimensional composites: A comparison of four empirical parceling algorithms. *Multivariate Behavioral Research, 39*(3), 379-412. https://doi.org/10.1207/S15327906MBR3903_1
- Serlin, R. C. (2000). Testing for robustness in Monte Carlo studies. *Psychological Methods, 5*(2), 230-240. <https://doi.org/10.1037/1082-989X.5.2.230>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359-1366. <https://doi.org/10.1177/0956797611417632>
- Wall, M. M., & Amemiya, Y. (2003). A method of moments technique for fitting interaction effects in structural equation models. *British Journal of Mathematical & Statistical Psychology, 56*(1), 47-63. <https://doi.org/10.1348/000711003321645331>
- Wu, Y., Wen, Z., Marsh, H. W., & Hau, K. T. (2013). A comparison of strategies for forming product indicators for unequal numbers of items in structural equation models of latent interactions. *Structural Equation Modeling, 20*(4), 551-567. <https://doi.org/10.1080/10705511.2013.824772>



Methodology is the official journal of the European Association of Methodology (EAM).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology Information (ZPID), Germany.