International Journal of Statistics and Probability; Vol. 7, No. 6; November 2018
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Approaches for Structural Investigations of Binary Data Using Confirmatory Factor Models

Karl Schweizer[1], Siegbert Reiß[1] & Stefan Troche[2]

[1] Institute of Psychology, Goethe University Frankfurt, Frankfurt a. M., Germany

[2] Department of Psychology and Psychotherapy, University of Witten/Herdecke, Germany

Correspondence: Karl Schweizer, Institute of Psychology, Goethe University Frankfurt, Theodor-W.- Adorno-Platz 6, 60323 Frankfurt a. M., Germany.

Received: June 27, 2018 Accepted: August 15, 2018 Online Published: September 12, 2018

doi:10.5539/ijsp.v7n6p68 URL: https://doi.org/10.5539/ijsp.v7n6p68

Abstract

An investigation of the suitability of threshold-based and threshold-free approaches for structural investigations of binary data is reported. Both approaches implicitly establish a relationship between binary data following the binomial distribution on one hand and continuous random variables assuming a normal distribution on the other hand. In two simulation studies we investigated: whether the fit results confirm the establishment of such a relationship, whether the differences between correct and incorrect models are retained and to what degree the sample size influences the results. Both approaches proved to establish the relationship. Using the threshold-free approach it was achieved by customary ML estimation whereas robust ML estimation was necessary in the threshold-based approach. Discrimination between correct and incorrect models was observed for both approaches. Larger CFI differences were found for the threshold-free approach than for the threshold-based approach. Dependency on sample size characterized the threshold-based approach but not the threshold-free approach. The threshold-based approach tended to perform better in large sample sizes, while the threshold-free approach performed better in smaller sample sizes.

Keywords: binary data, confirmatory factor analysis, constrained factor loadings, probability-based covariance, tetrachoric correlation

1. Introduction

The paper presents an evaluation of two approaches that are recommended for the investigation of the psychometric quality of psychological scales with items yielding binary data; i.e. items that only allow for correct or incorrect responses. The two approaches comprise methods for overcoming the following problem: on one hand there are binary data achieved by the items of psychometric scales and on the other hand there are statistical procedures that have been designed for investigating continuous and normally distributed data; investigating the binary data by means of these procedures means a violation of assumptions. An established way of overcoming this problem is the use of a link transformation as part of a generalized linear model (McCullagh & Nelder, 1985; Nelder & Wedderburn, 1972). However, this way requires the use of a link function from the so-called exponential family that is not suitable for the statistical procedures normally employed for the investigation of the psychometric quality of psychological scales: confirmatory factor analysis (CFA). The two approaches in the focus of this study propose ways of overcoming the problem in the framework of CFA when the data are binary.

The two approaches are the threshold-based and threshold-free approaches. The threshold is defined as the point that separates the two response options on the underlying ability or trait dimension. A major characteristic of the threshold-based approach is the estimation of thresholds. In this case either tetrachoric correlations are computed and used as input to CFA (Muthén, 1984) or item factor analysis (Bock, Gibbons, & Muraki, 1988) is conducted including the estimation of thresholds. In contrast, the threshold-free approach uses a link transformation that is conducted as part of CFA (Schweizer, Ren, & Wang, 2015) in order to overcome the difference between data and statistical model. The evaluation of these approaches is conducted by means of two simulation studies. These studies are designed to provide empirical evidence with respect to the following three research questions: (1) Do both approaches perform the transformation from binary that also means binomially distributed to continuous and normally distributed completely or is an additional adjustment necessary? (2) How does the sample size influence the outcome? (3) Are these approaches appropriate for distinguishing between correct and incorrect models?

A major characteristic of the investigations in the present paper is that the factor loadings are fixed to pre-specified values

source: https://doi.org/10.7892/boris.127842 | downloaded: 25.4.2024

in order to assure that the latent variable represents a specific source of responding. In the past many models of measurement with fixed parameters have been considered but only those allowing errors to vary found acceptance (Millsap, 2001; Millsap & Everson, 1991). Such models are the tau-equivalent model, the early growth curve model and the fixed-links model. Constrained factor loadings are not a disadvantage since models with free and fixed factor loadings lead to the same fit results if the model is correct (Schweizer, Ren, Wang, & Zeller, 2015; Schweizer, Troche, & Reiß, 2018). In contrast, free factor loadings perform better than constrained factor loadings in the case of an incorrect model because of their better adaptability to minor additional effects. Furthermore, factor loadings are considered as analogous to discriminability parameters of models in the framework of the item-response theory (IRT) (Lucke, 2005), where the fixation of parameters is quite common. Examples are the Rasch (1960) model, the corresponding one-parameter model (Birnbaum, 1968), and the Rasch model-based linear logistic test model (Scheiblechner, 1972).

1.1 The Characteristics of the Threshold-Based Approach

The concept of threshold values for overcoming the difference between distributions was developed for estimating the tetrachoric correlation (Pearson, 1900). While different methods are available for computing this correlation, the most popular method includes the estimation of latent thresholds by means of the maximum likelihood estimation method (Tallis 1962). To illustrate the transformation, assume a binary, random variable, X, with zero and one as possible values that represents the response to an item. Then the probability of X being equal to one [i.e., Pr(X=1)] is set equal to the probability that the continuous and normally distributed variable U is larger than the threshold τ that characterizes the item:

$$\Pr(X=1)=\Pr(U>\tau). \tag{1}$$

Furthermore, it is assumed that the observable variable X that is binary and shows a binomial distribution is the result of dichotomizing the non-observable variable X^* that is continuous and normally distributed (Muthén, 1984). This means that there is the function $f_{\text{Dichotomization_using_}\tau^*}$ such that

$$X = f_{\text{Dichotomization_using_}} \quad (X^*) . \tag{2}$$

After having achieved the thresholds for two random variables X_i and X_j (i,j=1, …, p), these thresholds are used for computing the correlation between X_i and X_j. Because of the switch from binomial to normal, the tetrachoric correlation is treated as correlation between two continuous, normally distributed random variables (i.e., latent correlation).

The tetrachoric correlations are integrated into the $p\times p$ matrix of correlations among all variables, $\mathbf{R}_{\text{tetrachoric}}$. The transformation preceding the computation of the correlations is expected to assure that this matrix represents relations among underlying variables at the latent level, thus satisfying the requirements of CFA (Muthén, 1984, 1993). Although this approach presents a convincing theoretical basis, the results obtained in empirical research are not always in line with the expectations. Thus, there have been many recent attempts to the improvement of the efficiency of computing tetrachoric correlations (e.g., Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009; Maydeu-Olivares, 2006; Savalei, Bonett, & Bentler, 2013; Savalei & Rhemtulla, 2013).

1.2 The Characteristics of the Threshold-Free Approach

Within the threshold-free approach, the difference between the data and the model of measurement is overcome in three steps; however, only two of them are of importance for structural investigations concentrating on model fit (Schweizer, 2013; Schweizer, Ren, & Wang, 2015). These are the step from the binary to the continuous scale and the step from a binomial to a normal distribution. The third step (not of importance in model fit) is the disattenuation of factor loadings.

The step from the binary to the continuous scale is accomplished by computing probability-based covariances. The covariance of two binary random variables X_i and X_j (i,j=1, …, p) with zero and one as values representing incorrect and correct responses cov(X_i, X_j) is defined as

$$\operatorname{cov}(X_i, X_j)=\Pr(X_i=1\cap X_j=1)-\Pr(X_i=1)\Pr(X_j=1) \tag{3}$$

where Pr(X_i =1) and Pr(X_j =1) are the probabilities that X_i respectively X_j are equal to one (i.e., a correct response) and Pr(X_i =1 $\cap$ X_i =1) the probability that both X_i and X_j are equal to one. It does not require knowledge of the individual entries of a data matrix and can be computed on the basis of frequencies, that is, on the basis of binary information; however, outcomes are values following an interval scale. The probability-based covariance defined this way corresponds to a pre-stage of computing the Phi coefficient (see McDonald & Ahlawat, 1974; Schweizer, 2013).

The step from the binomial distribution to a normal distribution is conducted in concentrating on the variances and covariances, which ultimately serve as input to CFA. This means that the step of estimating and using thresholds when computing tetrachoric correlations is simply omitted. The variance of the continuous and normally distributed random variable X_i of the model var(X_i) [=cov(X_i, X_i)] is transformed by the link function g_{link} to correspond for the observed variance s_i:

$$s_i = g_{\text{link}}[\text{var}(X_i)] \ . \tag{4}$$

The link function of the threshold-free approach is realized as weight w_i (i =1, …, p). In CFA using fixed factor loadings w_i is defined as:

$$w_i = \sqrt{\Pr(X_i = 1)[1 - \Pr(X_i = 1)]/0.25} \tag{5}$$

where $\Pr(X_i = 1)$ is the probability of a correct response. This weight serves as a multiplier to the corresponding factor loadings when a covariance model is used (Jöreskog, 1970):

$$g_{\text{link}}[\text{var}(X_i)] = (w_i\lambda_i)\phi(w_i\lambda_i) + \vartheta_i \ . \tag{6}$$

Here, λ_i is the fixed factor loading, ϕ the variance of the latent variable and ϑ_i the error variance. As is shown in Equation 6, the link transformation extends only to the systematic variance but not to the error variance.

1.3 On the Evaluation of the Performance of Models With Constraints

Two strategies are employed for evaluating the outcomes of investigations of the structure of binary data using methods according to both two approaches. The first strategy focuses on model fit. The model of measurement plays a major role in this strategy. All models of measurement (see Graham, 2006) show the same basic structure that is adapted from the congeneric model (Jöreskog, 1971). The congeneric model is defined as

$$\mathbf{y} = \mathbf{\Lambda\eta} + \boldsymbol{\varepsilon} \tag{7}$$

where $\mathbf{y}$ is the $p\times1$ vector of centered observations, $\boldsymbol{\eta}$ the $q\times1$ vector of latent variables (=latent factors) and $\boldsymbol{\varepsilon}$ the $p\times1$ vector of error components.

Constraining factor loadings means the assignment of numbers to the elements of $\mathbf{\Lambda}$. Each column $\boldsymbol{\lambda}_j$ (j=1, …, q) of $\mathbf{\Lambda}$ refers to a specific latent variable. In a simulation study, the numbers included in $\boldsymbol{\lambda}_j$ can be expected to serve well in CFA if they correspond to those used in constructing the relational pattern for generating simulated data because CFA implicitly accomplishes the reproduction of the covariance or correlation matrix, as noted by:

$$\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Theta} \tag{8}$$

where $\mathbf{\Sigma}$ is the $p\times p$ model-implied matrix of variances and covariances, $\mathbf{\Lambda}$ is the $p\times q$ matrix of factor loadings, $\mathbf{\Phi}$ the $q\times q$ matrix of the variances and covariances of the latent variables and $\mathbf{\Theta}$ the $p\times p$ diagonal matrix of error variances (Jöreskog, 1970). If there is only one latent variable this model reduces to

$$\mathbf{\Sigma} = \boldsymbol{\lambda}\phi\boldsymbol{\lambda}' + \mathbf{\Theta} \ . \tag{9}$$

Equation 9 can also be perceived as a scheme that is implicitly used in the preparation of the relational pattern for data generation. Therefore, it is possible to create the situation for study where $\lambda_{\text{data_generation}}$ corresponds to $\lambda_{\text{data_analysis}}$:

$$\boldsymbol{\lambda}_{\text{data_generation}} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdot \\ \cdot \\ \lambda_p \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdot \\ \cdot \\ \lambda_p \end{bmatrix} = \boldsymbol{\lambda}_{\text{data_analysis}} \ . \tag{10}$$

If such a situation is given in a simulation study, CFA can be expected to lead to a good model fit although, of course, all the other important essentials, as for example a sufficient sample size, must be given.

The second strategy focuses on the comparison of the fit results obtained in investigating two different types of data: data requiring the steps from binary and binomial to continuous and normal, on one hand, and data only requiring the step from binary to continuous, on the other hand. Binary data obtained by simulating items showing a broad range of difficulties (i.e., means a broad range of different degrees of skewness) require a change of scale and a link transformation. These data are referred to as *biased data* in this paper. In contrast, in binary data requiring the step of the second type only, the data already show a symmetric distribution so that a transformation for eliminating skewness is not necessary. Such data are referred to as *unbiased data* in this paper. In unbiased data the probability of a correct response is .5.

Both types of data are assumed to originate from continuous data following the standard normal distribution [N(0,1)]. The $p\times p$ matrix $\mathbf{C}_{\text{continuous_and_normal}}$ (computed from these data) is assumed to include the variances and covariances obtained from the columns of the generated matrices. Unbiased binary data are achieved by rescaling continuous data into zeros and ones in such a way that the probability of one is .5 and the variance is .25. The variances and covariances of the unbiased binary data are included in the $p\times p$ matrix $\mathbf{C}_{\text{binary_and_symmetric}}$. Since $\mathbf{C}_{\text{continuous_and_normal}}$ and $\mathbf{C}_{\text{binary_and_symmetric}}$ originate from the same raw data [N(0,1)] and the transformation from continuous to binary affects the variances of all

columns of the data matrix in exactly the same way, there is a special relationship. This relationship refers to the expected matrices and can be described by the linear function g_{linear} (that is not a link function):

$$E(\mathbf{C}_{binary_and_symmetric}) = g_{linear}[\ E(\mathbf{C}_{continuous_and_normal})\ ] \tag{11}$$

while individual pairs of matrices may show a small degree of deviation because of random influences. As a consequence, the virtually same degree of model fit can be expected for $E(\mathbf{C}_{binary_and_symmetric})$ and for $E(\mathbf{C}_{continuous_and_normal})$ if the model is specified correctly.

In contrast, biased data originate from splitting up continuous and normally distributed data according to a variety of different proportions so that the variances $var(X_i)$ (i=1, …, p) vary between 0 and 0.25. Therefore, in this case it is not possible to describe the relationship by the linear function g_{linear}. Instead

$$E(\mathbf{C}_{binary_and_asymmetric}) \neq g_{linear}[\ E(\mathbf{C}_{continuous_and_normal})\ ] \tag{12}$$

where $\mathbf{C}_{binary_and_asymmetric}$ is a $p \times p$ covariance matrix of binary data showing a binomial (and mostly asymmetric) distribution. The inequity suggests that in investigations of model fit the investigation of $\mathbf{C}_{binary_and_asymmetric}$ should lead to worse results than the investigation of $\mathbf{C}_{continuous_and_normal}$ or $\mathbf{C}_{binary_and_symmetric}$ if there is no efficient link transformation or correct threshold estimation. Consequently, it depends on the link transformation or threshold estimation whether the investigation of biased data ($\mathbf{C}_{binary_and_asymmetric}$) leads to good model fit or model misfit.

1.4 The Objective of the Empirical Investigation

The main aim of the empirical part of the research was to investigate whether the threshold-based and the threshold-free approaches assured that the steps from binary and binomial to continuous and normal were accomplished correctly. Using confirmatory factor models that corresponded to the models used for data generation, it was expected that the successful accomplishment of the steps would find its expression in good model fit.

Another aim was to investigate whether the two approaches were appropriate for distinguishing between correct and incorrect models. This aim is important to determine whether the transformations retained the discriminability between correct and incorrect models.

A third aim was the investigation of the influence of the sample size on model fit. The sample size was an important topic in factor analytic investigations with a long tradition since a sufficiently large sample was found to be an important precondition for the achievement of valid results.

2. Method

2.1 First Study

Two 12×12 relational patterns of coefficients served the generation of simulated data. Since there were already studies using patterns obtained by means of equally sized factor loadings (e.g., Schweizer, Ren, Wang, & Zeller, 2015), more complex patterns were employed that still enabled a one-dimensional representation and could be expected to reveal differences between the threshold-based and the threshold-free approaches. In the first relational pattern addressed as Pattern A the coefficients showed a systematic increase from 0.15 to 0.35, and the coefficients of the diagonal were 1.00. The upper half of Table 1 shows the lower triangle of this pattern.

The second relational pattern was assumed to originate from two subsets of manifest variables. The coefficients of this pattern varied between .15 and .35. The lower half of Table 1 includes this Pattern B. Two hundred 1000 × 12 matrices were generated according to each pattern. To obtain relationships according to these patterns the unrelated columns of matrices including random data were combined using weights achieved by means of a procedure proposed by Jöreskog and Sörbom (2001). The matrices of simulated data obtained this way provided the outset for the transformation into binary data.

In the next step, the numbers of the columns of the matrices were dichotomized. Two types of splits were realized for obtaining biased data and unbiased data. First, the continuous data were split so that a broad range of probabilities of binary events was achieved. The following set of splits was realized: 0.100, 0.172, 0.225, 0.317, 0.390, 0.462, 0.535, 0.608, 0.681, 0.754, 0.827, and 0.900. For example, the 0.100 split in the first column required that the 10 percent of small random numbers in this column of each matrix were replaced by zeros and the remaining numbers by ones. In the next column, zeros substituted the 17.2 percent smaller numbers and the other numbers were substituted by ones. The remaining columns were treated in an analogous way. Second, dichotomization for achieving distributional equivalence was established by the 0.5 split of each column of each matrix. Afterwards, tetrachoric correlations and probability-based covariances were computed. The tetrachoric correlations were obtained by PRELIS (Jöreskog & Sörbom, 1994). PASCAL programs served the dichotomization of the continuous data and the estimation of the probability-based covariances.

Table 1. Patterns (A and B) Used for the Generation of Simulated Data

Pattern A	1.000											
	.150	1.000										
	.150	.170	1.000									
	.150	.170	.190	1.000								
	.150	.170	.190	.210	1.000							
	.150	.170	.190	.210	.230	1.000						
	.150	.170	.190	.210	.230	.250	1.000					
	.150	.170	.190	.210	.230	.250	.270	1.000				
	.150	.170	.190	.210	.230	.250	.270	.290	1.000			
	.150	.170	.190	.210	.230	.250	.270	.290	.310	1.000		
	.150	.170	.190	.210	.230	.250	.270	.290	.310	.330	1.000	
	.150	.170	.190	.210	.230	.250	.270	.290	.310	.330	.350	1.000
Pattern B	1.000											
	.150	1.000										
	.150	.150	1.000									
	.150	.150	.150	1.000								
	.150	.150	.150	.150	1.000							
	.150	.150	.150	.150	.150	1.000						
	.250	.250	.250	.250	.250	.250	1.000					
	.250	.250	.250	.250	.250	.250	.350	1.000				
	.250	.250	.250	.250	.250	.250	.350	.350	1.000			
	.250	.250	.250	.250	.250	.250	.350	.350	.350	1.000		
	.250	.250	.250	.250	.250	.250	.350	.350	.350	.350	1.000	
	.250	.250	.250	.250	.250	.250	.350	.350	.350	.350	.350	1.000

Correct and incorrect confirmatory models were specified. Each model consisted of one latent variable and 12 manifest variables loading on the latent variable. In the incorrect model, factor loadings were constrained to equal sizes in combination with setting the variance of the latent variable free for estimation. Correct specification with respect to data simulated according to Pattern A was achieved by using the following set of numbers as factor loadings: 0.387, 0.41, 0.434, 0.454, 0.473, 0.492, 0.510, 0.527, 0.544, 0.560, 0.576, and 0.592. In the case of a model according to the threshold-free approach weights according to Equation 5 additionally served as multipliers to the factor loadings. Using these numbers as constraints and Pattern A as input to CFA yielded good model fit, $\chi^2(65) = 42.4$, RMSEA = .000, SRMR = .023, GFI = .99, CFI = 1.00, NNFI = 1.01.

The numbers for correct specification with respect to data simulated according to Pattern B were 0.387, 0.387, 0.387, 0.387, 0.387, 0.387, 0.591, 0.591, 0.591, 0.591, 0.591, and 0.591. The threshold-free approach additionally required weights as multipliers to the factor loadings. These numbers as constraints together with Pattern B as input to CFA also led to good model fit, $\chi^2(65) = 11.1$, RMSEA = .000, SRMR = .011, GFI = 1.00, CFI = 1.00, NNFI = 1.01.

For all investigations the software LISREL (Jöreskog & Sörbom, 2006) was used with ML (also addresses as customary ML) and MLR as estimation methods. ML was selected because it proved to discriminate well between models in binary data (Zeller, Krampen, Reiß, & Schweizer, 2017). The evaluation of the outcomes concentrated on model fit. The following fit indexes were considered in preparing the report of the results: chi-square, degree of freedom, normed chi-square, RMSEA, SRMR, CFI, TLI and GFI. Cut-offs provided by Hu and Bentler (1999) and Kline (2005) served the evaluation of the results (RMSEA $\leq$ .06, SRMR $\leq$ .08, CFI $\geq$ .95, TLI $\geq$ .95, GFI $\geq$ .95) (see also DiStefano, 2016). Furthermore, normed chi-squares below 2 (N=200), 3 (N=400) and 5 (N=1000 and larger) were regarded as indications of a good model-data fit. The CFI difference served the comparison of models. Following Cheung and Rensvold (2002) a CFI difference of .01 was considered as substantial.

2.2 Second Study

Two uniform, 9×9 pattern matrices were selected for the generation of simulated data. The off-diagonal coefficients of the first and second patterns were 0.25 and 0.30, respectively, and the diagonal coefficients 1.00 in both cases. Besides the two different uniform patterns four sample sizes were considered: 200, 400, 1000 and 2000. Sizes of 200 and 400 were selected because the sample sizes of empirical studies frequently range between 200 and 400. The other sample sizes were added because the use of tetrachoric correlations might necessitate larger sample sizes to obtain good model-data fit. Sets of 100 matrices were generated for each combination of a specific sample size and a specific size of off-diagonal coefficients of the relational pattern.

The structure according to the two uniform patterns was induced into the matrices of random data in the same way as in the first study. Furthermore, in dichotomizing the continuous data the following set of splits was used: 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80 and 0.90. The data were dichotomized analogously to the procedure described for Study 1.

The confirmatory factor model for investigating the matrices comprised one latent variable and nine manifest variables. The factor loadings were constrained to equal sizes in combination with setting the variance parameter free for estimation. Since the variance of the latent variable could be expected to compensate for lower or larger sizes of the number for fixing the factor loadings, these loadings were simply set equal to one. The threshold-free approach additionally required weights as multipliers to the factor loadings.

3. Results

3.1 First Study

The focus of this study was on model fit for the threshold-based and the threshold-free approaches, the difference in model fit between the applications to biased and unbiased data and the discrimination between correctly and incorrectly specified models.

3.1.1 Results for the Threshold-based Approach

The fit results observed when tetrachoric correlations based on Pattern A served as input to confirmatory factor analysis are reported in Table 2. Each row of this Table extends to means and corresponding standard deviations.

Table 2. Means and Standard Deviations (SD) of the Fit Results for Investigating Data Constructed According to Pattern A Using the Threshold-based Approach (200 Matrices of Simulated Data)

Model	Estimation	Probability range	χ^2	df	Normed χ^2	RMSEA	SRMR	CFI	TLI	GFI
Correct	ML	Broad	344.4	65	5.30	.065	.060	.92	.91	.95
SD			74.1		1.14	0.01	0.01	0.03	0.03	0.01
Incorrect	ML	Broad	417.4	65	6.42	.073	.077	.90	.90	.94
SD			78.6		1.21	0.01	0.01	0.03	0.03	0.01
Correct	ML	Zero	77.3	65	1.22	.013	.049	1.00	1.00	.96
SD			13.3		0.20	0.01	0.00	0.00	0.00	0.01
Incorrect	ML	Zero	108.4	65	1.67	.025	.069	.99	.99	.95
SD			16.9		0.26	0.01	0.01	0.01	0.01	0.01
Correct	MLR	Broad	74.2	65	1.14	.010	.060	1.00	1.00	.95
SD			16.3		0.25	0.01	0.01	0.00	0.00	0.01
Incorrect	MLR	Broad	90.8	65	1.40	.018	.077	.99	.99	.93
SD			17.6		0.27	0.01	0.01	0.01	0.01	0.01

The fit of the correct model with input originating from biased data using customary ML (first row) was good only with regard to SRMR and GFI. The comparison of the CFI results for input originating from biased and unbiased data (rows 1 and 3) signified a substantial CFI difference (.08). Robust estimation yielded fit results similar to the results for unbiased data (rows 3 and 5). Each comparison between the CFI results for the correct and incorrect models signified a substantial difference (rows 1 and 2: .02, rows 3 and 4: .01, and rows 5 and 6: .01). In sum, the threshold-based approach did not yield results to suggest that the step from binary and binomial to continuous and normal was performed adequately. Robust estimation was additionally necessary for achieving overall good model fit.

CFA with tetrachoric correlations based on Pattern B as input yielded the fit results of Table 3. Each row of this Table includes means and corresponding standard deviations.

Table 3. Means and Standard Deviations (SD) of the Fit Results for Investigating Data Constructed According to Pattern B Using the Threshold-based Approach (200 Matrices of Simulated Data)

Model	Estimation	Probability range	χ^2	df	Normed χ^2	RMSEA	SRMR	CFI	TLI	GFI
Correct	ML	Broad	393.6	65	6.05	.070	.062	.91	.91	.94
SD			96.3		1.48	0.01	0.01	0.04	0.04	0.01
Incorrect	ML	Broad	495.8	65	7.63	.081	.093	.89	.89	.92
SD			96.7		1.49	0.01	0.01	0.04	0.04	0.01
Correct	ML	Zero	90.9	65	1.40	.019	.052	.99	.99	.96
SD			14.7		0.23	0.01	0.01	0.00	0.00	0.01
Incorrect	ML	Zero	131.4	65	2.02	.032	.086	.98	.98	.94
SD			17.7		0.27	0.00	0.01	0.00	0.00	0.01
Correct	MLR	Broad	79.7	65	1.23	.013	.062	1.00	1.00	.94
SD			19.3		0.30	0.01	0.01	0.00	0.00	0.01
Incorrect	MLR	Broad	101.9	65	1.57	.023	.093	.99	.99	.92
SD			19.6		0.30	0.01	0.01	0.00	0.00	0.01

The first row of Table 3 contains the fit statistics for the correct model and biased data using customary ML. Only the SRMR statistic indicated a good model fit. The comparison of the CFIs in investigating biased and unbiased data (rows 1 and 3) signified a substantial difference (.08). Conducting robust estimation led to an improvement even over customary ML applied to unbiased data (rows 3 and 5) according to the CFI difference (.01). All comparisons between the CFI results for the correct and incorrect models signified a substantial difference (rows 1 and 2: .02, rows 3 and 4: .01 and rows 5 and 6: .01). In sum, the results for the threshold-based approach without robust ML suggested that the step from binary and binomial to continuous and normal was not performed well. Robust estimation that accounts for distributional discrepancy was additionally necessary.

3.1.2 Results for the Threshold-Free Approach

Table 4 includes the results achieved by means of the threshold-free approach applied to data based on Pattern A. Also in this Table each row extends to means and corresponding standard deviations.

Table 4. Means and Standard Deviations (SD) of the Fit Results for Investigating Data Constructed According to Pattern A Using the Threshold-free Approach (200 Matrices of Simulated Data)

Model	Estimation	Probability range	χ^2	df	Normed χ^2	RMSEA	SRMR	CFI	TLI	GFI
Correct	ML	Broad	96.6	65	1.48	.021	.037	.97	.97	.98
SD			16.4		0.25	0.01	0.00	0.02	0.02	0.00
Incorrect	ML	Broad	125.2	65	1.92	.030	.045	.94	.94	.98
SD			21.0		0.32	0.01	0.00	0.02	0.02	0.00
Correct	ML	Zero	81.2	65	1.25	.014	.032	.99	.99	.99
SD			13.3		0.21	0.01	0.00	0.01	0.00	0.00
Incorrect	ML	Zero	118.4	65	1.82	.028	.044	.96	.96	.98
SD			19.1		0.29	0.01	0.00	0.01	0.01	0.00

The model fit for the correct model applied to biased data using customary ML (first row) led to a good fit according to all fit indices. The comparison of the CFI results for the biased and unbiased data as input (rows 1 and 3) signified a substantial CFI difference (.02). Each comparison between the CFI results for the correct and incorrect models signified a substantial difference (rows 1 and 2: .03, and rows 3 and 4: .03). Thus, with the threshold-free approach the step from binary and binomially distributed data to continuous and normally distributed data was performed to some degree, as is obvious in the good model fit observed when investigating biased data. However, the comparison of these results with those observed for unbiased data as input signified that the step was not performed perfectly in this type of data.

For Pattern B, CFA led to the fit results reported in Table 5. Again, each row of this Table includes both means and corresponding standard deviations.

Table 5. Means and Standard Deviations (SD) of the Fit Results for Investigating Data Constructed According to Pattern B Using the Threshold-free Approach (200 Matrices of Simulated Data)

Model	Estimation	Probability range	χ^2	df	Normed χ^2	RMSEA	SRMR	CFI	TLI	GFI
Correct	ML	Broad	87.9	65	1.35	.017	.035	.98	.98	.99
SD			18.2		0.28	0.01	0.00	0.01	0.02	0.00
Incorrect	ML	Broad	147.2	65	2.26	.035	.053	.93	.93	.98
SD			23.2		0.36	0.01	0.00	0.02	0.02	0.00
Correct	ML	Zero	92.0	65	1.41	.019	.034	.98	.98	.98
SD			16.3		0.25	0.01	0.00	0.01	0.01	0.00
Incorrect	ML	Zero	152.5	65	2.35	.036	.056	.95	.95	.98
SD			20.6		0.32	0.01	0.00	0.01	0.01	0.00

The results for the correct model with biased data as input and using customary ML (first row) indicated a good model fit according to all fit indices. The CFI results for the biased and unbiased data (rows 1 and 3) as input closely corresponded. The correspondence indicated that the steps between binary and binomial, on one hand, and continuous and normal, on the other hand, were nearly perfect. Each comparison between CFI results for the correct and incorrect models signified a substantial difference (rows 1 and 2: .05, and rows 3 and 4: .03). In sum, with the threshold-free approach the steps from binary and binomially distributed data to continuous and normally distributed data were performed well. Furthermore, there was no more a difference between using biased or unbiased data as input to CFA.

Overall, the results suggested that the threshold-free approach accomplished the steps from binary and binomially distributed data to continuous and normally distributed data better than the threshold-based approach. As a consequence, the threshold-based approach needed to be combined with robust ML estimation whereas the threshold-free approach could be conducted using customary ML estimation. Both approaches enabled the discrimination between correct and incorrect models. The mean CFI difference between correct and incorrect models was 0.013 for the threshold-based approach and 0.035 for the threshold-free approach indicating that is was more likely to discriminate between the two types of models within the threshold-free approach compared to the threshold-based approach.

3.2 Second Study

This study investigated the effect of the sample size on model fit. To facilitate the reading of the tables reporting fit results, the superscript “M” was added to a mean of fit statistics if the outcome of the comparison of the mean with the corresponding cut-off was positive. If even the 95 % confidence interval was above respectively below the corresponding cut-off, it was replaced by the superscript “CI”.

Means and standard deviations of the fit results of investigating the data matrices constructed according to the uniform pattern including off-diagonal coefficients of .25 are reported in Table 6. The results reported in the first to fourth rows were obtained by MLR estimation in combination with tetrachoric correlations as input to CFA. In the sample size of 200 only the mean normed χ^2 and RMSEA indicated good model fit. In the sample size of 400 the mean CFI and TLI were also good. All fit statistics indicated good model fit within the 95 % confidence interval when the sample size was 1000 or larger.

The last four rows of Table 6 include the results of the investigations using the threshold-free approach. In the sample sizes of 200 and 400 normed χ^2, RMSEA, SRMR and GFI signified good model fit within the 95 % confidence interval while the mean CFI and TLI were also good. After the increase of the sample size to 1000 the CFI results were also good within the 95 % confidence interval and the further increase to 2000 led all fit statistics to comply with the 95 % confidence interval.

Table 6. Means and Standard Deviations (SD) of the Fit Results for Binary Data Constructed According to the 0.25-level Pattern

Input	Estimation	Sample size	χ^2	df	Normed χ^2	RMSEA	SRMR	CFI	TLI	GFI
Correlations	MLR	200	57.8	35	1.65^{M}	$.049^{M}$	.127	.91	.91	.83
SD			22.6		0.65	0.03	0.03	0.11	0.12	0.07
Correlations	MLR	400	50.1	35	1.43^{CI}	$.028^{M}$	.104	$.97^{M}$	$.97^{M}$	.88
SD			26.9		0.76	0.03	0.03	0.06	0.06	0.06
Correlations	MLR	1000	40.2	35	1.15^{CI}	008^{CI}	$.057^{CI}$	1.00^{CI}	1.00^{CI}	$.96^{CI}$
SD			23.4		0.70	0.01	0.03	0.01	0.01	0.02
Correlations	MLR	2000	36.9	35	1.05^{CI}	$.006^{CI}$	$.040^{CI}$	1.00^{CI}	1.00^{CI}	$.98^{CI}$
SD			11.4		0.32	0.01	0.01	0.01	0.01	0.01
Covariances	ML	200	36.2	35	1.03^{CI}	$.016^{CI}$	$.063^{CI}$	$.96^{M}$	$.99^{M}$	$.96^{CI}$
SD			7.8		0.22	0.02	0.01	0.05	0.08	0.01
Covariances	ML	400	37.5	35	1.07^{CI}	$.012^{CI}$	$.046^{CI}$	$.98^{M}$	$.99^{M}$	$.98^{CI}$
SD			8.6		0.25	0.01	0.01	0.03	0.04	0.00
Covariances	ML	1000	45.4	35	1.30^{CI}	$.015^{CI}$	$.034^{CI}$	$.98^{CI}$	$.98^{M}$	$.99^{CI}$
SD			10.1		0.29	0.01	0.00	0.01	0.02	0.00
Covariances	ML	2000	57.4	35	1.64^{CI}	$.017^{CI}$	$.028^{CI}$	$.98^{CI}$	$.98^{CI}$	$.99^{CI}$
SD			12.1		0.35	0.01	0.00	0.01	0.01	0.00

^{CI} The 95 % confidence interval is in the favourable area. ^{M} The mean is in the favourable area.

The data constructed according to the uniform pattern including off-diagonal coefficients of 0.30 were investigated in the same way as the data constructed according to the other pattern. The fit results are reported in Table 7.

Table 7. Means and Standard Deviations of the Fit Results for Binary Data Constructed According to the 0.30-level Pattern

Input	Estimation	Sample size	χ^2	df	Normed χ^2	RMSEA	SRMR	CFI	TLI	GFI
Correlations	MLR	200	61.2	35	1.75^{M}	$.054^{M}$	.118	.91	.91	.84
SD			22.8		0.65	0.03	0.03	0.10	0.10	0.07
Correlations	MLR	400	55.5	35	1.59^{CI}	$.029^{CI}$	.098	$.98^{M}$	$.98^{M}$	.87
SD			31.5		0.89	0.03	0.02	0.04	0.04	0.06
Correlations	MLR	1000	40.4	35	1.15^{CI}	$.008^{CI}$	$.058^{CI}$	1.00^{CI}	1.00^{CI}	$.95^{CI}$
SD			29.1		0.83	0.01	0.01	0.01	0.01	0.03
Correlations	MLR	2000	35.7	35	1.02^{CI}	$.005^{CI}$	$.039^{CI}$	1.00^{CI}	1.00^{CI}	$.98^{CI}$
SD			9.8		0.28	0.01	0.01	0.00	0.00	0.01
Covariances	ML	200	36.4	35	1.04^{CI}	$.016^{CI}$	$.064^{CI}$	$.96^{M}$	$.99^{M}$	$.96^{CI}$
SD			8.3		0.23	0.02	0.01	0.05	0.06	0.01
Covariances	ML	400	39.8	35	1.13^{CI}	$.017^{CI}$	$.048^{CI}$	$.97^{M}$	$.98^{M}$	$.98^{CI}$
SD			9.8		0.31	0.01	0.01	0.03	0.04	0.01
Covariances	ML	1000	51.3	35	1.47^{CI}	$.020^{CI}$	$.037^{CI}$	$.98^{M}$	$.98^{M}$	$.99^{CI}$
SD			11.5		0.33	0.01	0.01	0.02	0.02	0.00
Covariances	ML	2000	71.8	35	2.05^{CI}	$.022^{CI}$	$.032^{CI}$	$.98^{CI}$	$.98^{CI}$	$.99^{CI}$
SD			13.7		0.39	0.00	0.00	0.01	0.01	0.00

^{CI} The 95 % confidence interval is in the favourable area. ^{M} The mean is in the favourable area.

The outcomes of the evaluation of these results almost perfectly corresponded to what was reported as outcomes for Table 6. For sample sizes of 200 and 400, the investigations using the threshold-based approach led to model misfit according to the majority of fit indices whereas for sample sizes of 1000 and 2000 all fit statistics indicated a good model fit. In contrast, when the investigations were conducted according to the threshold-free approach, all fit statistics were at least good according to the mean with sample sizes of 200 and 400. The estimates improved when the sample size was increased. For the sample size of 2000, all fit statistics were included in the 95 % confidence interval.

All in all, for sample sizes of 200 and 400 the threshold-free approach led to better fit results than the threshold-based approach. In contrast, for the sample sizes of 1000 and 2000, there was a small numeric advantage for the threshold-based approach

4. Discussion

The investigation of binary data by CFA presupposes that a relationship is established between the properties characterizing the data on one hand and the properties characterizing the model integrated into the statistical method used in the investigation of the data on the other hand. The data are binary and binomially distributed whereas continuous and normally distributed variables characterize the model. The two considered approaches suggest different routes to the establishment of such a relationship. The results of the simulation studies indicate that both approaches accomplish the relationship successfully. However, the threshold-based approach requires robust maximum likelihood estimation that accomplishes an additional transformation regarding the distribution of data whereas the threshold-free approach does not make such an additional transformation necessary. Furthermore, the two approaches seem to differ according to their sensitivity for the sample size. The achievement of good model-data fit appears to depend on the sample size when using the threshold-based approach; a very large sample size seems to guarantee very good model-data fit whereas model misfit is likely in a small sample size. In contrast, the threshold-free approach appears to show almost no dependency on sample size.

Future progress in science using CFA requires efficient discrimination between correct and incorrect models. In the present study the model-data fit is in many cases good for the correctly and incorrectly specified models. Therefore, the discrimination between models becomes an important aspect of an investigation. Bentler's (1990) CFI is frequently used for this purpose since a CFI difference of .01 and larger means a substantial difference (Cheung & Rensvold, 2002). Using the threshold-based approach in four out of six comparisons the mean CFI difference was .01. Despite the just significant difference of means, many individual comparisons were not substantial. In contrast, using the threshold-free approach, all mean CFI differences were .03 or larger. Therefore, it seems to be more likely to discriminate between correct and incorrect models when using the threshold-free rather than the threshold-based approach.

The results demonstrate that using fixed factor loadings in CFA does not prevent the achievement of a good model-data fit. This observation runs counter to concerns based on the observation that models with free factor loadings, especially Jöreskog's (1971) congeneric model of measurement, dominate applied research. The presented results demonstrate that fixed factor loadings can substantially contribute to science and not just serve as provisions for amending estimation problems (Millsap, 2001). This important conclusion holds for both the threshold-based and the threshold-free approaches.

The use of models with fixed factor loadings may increase in future research on binary data if the trend to employ models considering more than one latent source of responding continues. As long as it was assumed that scales composed of binary items represent one systematic source of responding only, one-factor models with free factor loadings (Graham, 2006) could be expected to serve perfectly well for investigating the structure of binary data. However, this assumption is no more tenable. The multitrait-multimethod approach introduced by Campbell and Fiske (1959) has for more than 60 years produced evidence of method effects. Data usually show a considerable proportion of method variance besides trait variance. The consideration of sources of method variance as part of the model does not only control for impurity in measurement but also improves the replicability of results (Schweizer, Troche, & Rammsayer, 2011, 2018).

For example, there are a number of studies demonstrating that completing a homogeneous set of achievement items stimulates the item-position effect (e.g., Birney, Beckmann, Beckmann, & Double, 2017; Debeer & Janssen, 2013; Debeer, Buchholz, Hartig, & Janssen, 2014; Embretson, 1991; Hartig & Buchholz, 2012; Kubinger, 2008; Lozano, 2015; Ren et al., 2014; Schweizer, Schreiner, & Gold, 2009; Troche et al., 2016; Verguts & De Boeck, 2000; Zeller, Reiß, & Schweizer, 2017). A special characteristic of this effect is that it leads to a systematic increase of the proportion of systematic variance from the first to last items (Knowles, 1988). Another example of a method effect impairing the quality of measurement is the effect originating from the interaction of a time limit in testing and processing speed when the time limit prevents some or all participants of a sample from completing all items properly (Lu & Sireci, 2007; Oshima, 1994). The contribution of processing speed becomes apparent as lack of enough processing speed. This can mean either an omission or a random response. Since virtually all scales are applied with a time limit in testing, most performance data include (lack of) processing speed as another systematic source of responding to a larger or lesser degree. It is a source that usually influences the responses to the last few items of a scale and can be captured by a confirmatory factor model that takes the distribution of processing speed into consideration (Schweizer & Ren, 2013; Schweizer, Troche, & Reiß, 2018).

Overall the results demonstrate that confirmatory factor models with fixed factor loadings can serve well in investigating binary data independently of which approach is selected. The sample size seems to be a characteristic that distinguishes between the approaches and, therefore, needs to be considered in the selection of the approach for an investigation. Finally, it needs to be added that the reported results may be characteristic for the relational patterns used in data generation, and it is an open question to what degree they can be generalized to other relational patterns.

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246. https://doi.org/10.1037/0033-2909.107.2.238

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord, & M. R. Novick (Eds.), *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Birney, D. P., Beckmann, J. F., Beckmann, N., & Double, K. S. (2017). Beyond the intellect: complexity and learning trajectories in Raven's Progressive Matrices depend on self-regulatory processes and conative dispositions. *Intelligence.* Advance online publication. https://doi.org/10.1016/j.intell.2017.01.005

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement, 12,* 261-280. https://doi.org/10.1177/014662168801200305

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Debeer, D., & Janssen, R. (2013). Modeling item-position effects within an IRT framework. J*ournal of Educational Measurement, 50,* 164-185. https://doi.org/10.31021101699861455848 5

Debeer, D., Buchholz, J., Hartig, J., & Janssen, R. (2014). Student, school, and country differences in sustained test-taking effort in the 2009 PISA reading assessment. *Journal of Educational and Behavioral Statistics, 39*(6), 502–523. https://doi.org/10.3102/11016998614558485

DiStefano, C. (2016). Examining fit with structural equation models. In K.Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction* (pp. 26-51). Göttingen: Hogrefe.

Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. Psychometrika, *56*(3), 495–515. https://doi.org/10.1007/BF02294487

Forero, C. G., Maydeu-Olivares, A. & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling, 16,* 625-641. https://doi.org/10.1080/10705510903203573

Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability. *Educational and Psychological Measurement,* 66, 930-944. https://doi.org/10.1177/0013164406288165

Hartig, J., & Buchholz, J. (2012). A multilevel item response model for item position effects and individual persistence. *Psychological Test and Assessment Modelling, 54,* 418–431.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55. https://doi.org/10.1080/10705519909540118

Jöreskog, K. G. (1970). A general method for analysis of covariance structure. *Biometrika, 57,* 239-257. https://doi.org/10.2307/2334833

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36,* 109-133. https://doi.org/10.1007/BF02291393

Jöreskog, K. G., & Sörbom, D. (1994). *Prelis: A preprocessor for LISREL.* Chicago: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (2001). *Interactive LISREL: User's Guide.* Lincolnwood, IL: Scientific Software International Inc.

Jöreskog, K. G., & Sörbom, D. (2006). *LISREL 8.80.* Lincolnwood, IL: Scientific Software International Inc.

Kline, R. B. (2005). *Principles and practice of structural equation modeling.* 2nd ed. New York: Guliford Press.

Knowles, E. S. (1988). Item context effects on personality scales: Measuring changes the measure. J*ournal of Personality and Social Psychology, 55,* 312–320. https://doi.org/10.1037/0022-3514.55.2.312

Kubinger, K. D. (2003). On artificial results due to using factor analysis for dichotomous variables. *Psychology Science, 45,* 106-110.

Lozano, J. H. (2015). Are impulsivity and intelligence truly related constructs? Evidence based on the fixed-links model. *Personality and Individual Differences, 85,* 192–198. https://doi.org/10.1016/j.paid.2015.04.049

Lu, Y., & Sireci, S. G. (2007). Validity issues in test speededness. *Educational Measurement, 26,* 29-37. https://doi.org/10.1111/j.1745-3992.2007.00106.x

Maydeu-Olivares, A. (2006). Limited information estimation and testing of discretized multivariate normal structural models. *Psychometrika, 71*, 57-77. https://doi.org/10.1007/S11336-005-0773-4

McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics, 16*, 163-200. https://doi.org/10.1007/BF01065485

McCullagh, P., & Nelder, J. A. (1985). *Generalized linear models.* London: Chapman and Hall.

McDonald, R. P., & Ahlawat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology, 27*, 82–99. https://doi.org/10.1111/j.2044-8317.1974.tb00530.x

Millsap, R. E. (2001). When trivial constraints are not trivial: the choice of uniqueness constraints in confirmatory factor analysis. *Structural Equation Modeling, 8*, 1-17. https://doi.org/10.1207/S15328007SEM0801_1

Millsap, R. E., & Everson, H. (1991). Confirmatory measurement model comparisons using latent means. *Multivariate Behavioral Research, 26*, 479-497. https://doi.org/10.1207/s15327906mbr2603_6

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

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

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A, 135*, 370-384. https://doi.org/10.2307/2344614

Oshima, T. C. (1994). The effect of speededness on parameter estimation in item response theory. *Journal of Educational Measurement, 31*, 200-219. https://doi.org/10.1111/j.1745-3984.1994.tb00443.x

Pearson, K. (1900). Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Philosophical Transaction of the Royal Society, 195*, 1-47. https://doi.org/10.1098/rsta.1900.0022

Rasch, G. (1960). P*robabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research.

Ren, X., Wang, T., Altmeyer, M., & Schweizer, K. (2014). A learning-based account of fluid intelligence from the perspective of the position effect. *Learning and Individual Differences, 31*, 30–35. https://doi.org/10.1016/j.lindif.2014.01.002

Savalei, V., & Rhemtulla, M. (2013). The performance of robust test statistics with categorical data. *British Journal of Mathematical and Statistical Psychology, 66*, 201-223. https://doi.org/10.1111/j.2044-8317.2012.02049.x

Savalei, V., Bonett, D. G., & Bentler, P. M. (2013). CFA with binary variables in small samples: A comparison of two methods. Frontiers in Psychology, *5*, 1515. https://doi.org/10.3389/fpsyg.2014.01515

Scheiblechner, H. (1972). Das Lernen und Lösen komplexer aufgaben [Learning and solving complex cognitive problems]. *Zeitschrift für Experimentelle und Angewandte Psychologie, 19*, 476-505.

Schweizer, K. (2013). A threshold-free approach to the study of the structure of binary data. *International Journal of Statistics and Probability, 2*, 67-75. https://doi.org/10.5539/ijsp.v2n2p67

Schweizer, K., & Ren, X. (2013). The position effect in tests with a time limit: The consideration of interruption and working speed. *Psychological Test and Assessment Modelling, 55*, 62–78.

Schweizer, K., Ren, X., & Wang, T. (2015). A comparison of confirmatory factor analysis of binary data on the basis of tetrachoric correlations and of probability-based covariances: a simulation study. In R. E. Millsap, D. M. Bolt, L. A. van der Ark, & W.-C. Wang (Eds.), *Quantitative Psychology Research* (pp. 273-292). Heidelberg: Springer.

Schweizer, K., Ren, X., Wang, T., & Zeller, F. (2015). Does the constraint of factor loadings impair model fit and accuracy in parameter estimation? *International Journal of Statistics and Probability, 4*, 40-50. https://doi.org/10.5539/ijsp.v4n4p40

Schweizer, K., Schreiner, M., & Gold, A. (2009). The confirmatory investigation of APM items with loadings as a function of the position and easiness of items: a two-dimensional model of APM. *Psychology Science Quarterly, 51*, 47-64.

Schweizer, K., Troche, S., & Rammsayer, T. (2011). On the Special Relationship between Fluid and General Intelligence: New Evidence Obtained by Considering the Position Effect. *Personality and Individual Differences, 50*, 1249-1254. https://doi.org/10.1016/j.paid.2011.02.019

Schweizer, K., Troche, S., & Rammsayer, T. (2018). Does processing speed exert an influence on the special relationship of fluid and general intelligence? *Personality and Individual Differences, 131,* 57-60. Https://doi.org/10.1016/j.paid.2018.04.022

Schweizer, K., Troche, S., & Reiß, S. (2018). Does the effect of a time limit for testing impair structural investigations by means of confirmatory factor models? *Educational and Psychological Measurement.* https://doi.org/10.1177/0013164418770824

Tallis, G. M. (1962). The maximum likelihood estimation of correlation from contingency tables. *Biometrics, 18,* 342-353.

Troche, S., Wagner, F. L., Schweizer, K., & Rammsayer, T. H. (2016). The structural validity of the culture fair test under consideration of the item-position effect. E*uropean Journal of Psychological Assessment.* https://doi.org/10.1027/1015-5759/a000384

Verguts, T., & De Boeck, P. (2000). A Rasch model for detecting learning while solving an intelligence test. *Applied Psychological Measurement, 24,* 151–162. https://doi.org/10.1177/01466210022031589

Zeller, F., Krampen, D., Reiß, S., & Schweizer, K. (2017). Do adaptive representations of the item-position effect in APM improve model fit? a simulation study. *Educational and Psychological Measurement, 77,* 743-765. https://doi.org/10.1177/0013164416654946

Zeller, Reiß, S., & Schweizer, K. (2017). Is the item-position effect in achievement measures induced by increasing item Difficulty? *Structural Equation Modelling, 24,* 745-754. https://doi.org/10.1080/10705511.2017.1306706

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).