

# Does the Constraint of Factor Loadings Impair Model Fit and Accuracy in Parameter Estimation?

Karl Schweizer<sup>1</sup>, Xuezhun Ren<sup>2</sup>, Tengfei Wang<sup>1</sup> & Florian Zeller<sup>1</sup>

<sup>1</sup> Department of Psychology, Goethe University Frankfurt, Frankfurt, Germany

<sup>2</sup> School of Education, Huazhong University of Science & Technology, Wuhan, China

Correspondence: Karl Schweizer, Department of Psychology, Goethe University Frankfurt, Frankfurt, Theodor-W.-Adorno-Platz 6, 60323 Frankfurt am Main, Germany. Tel: 0049/69/798 35355. E-mail: K.Schweizer@psych.uni-frankfurt.de

Received: August 16, 2015 Accepted: September 2, 2015 Online Published: September 23, 2015

doi:10.5539/ijsp.v4n4p40

URL: <http://dx.doi.org/10.5539/ijsp.v4n4p40>

## Abstract

The paper reports on the comparison of models of measurement with constrained and free factor loadings as part of confirmatory factor analysis in a simulation study. The comparison was conducted in order to find out whether constrained factor loadings that cause a reduced degree of adaptability to specificities of data mean a disadvantage in comparison to factor loadings that are freely estimated. Furthermore, the way of conducting the link transformation, the sample size and the number of variables were varied. The simulated data were dichotomous and constructed to conform to one underlying source of responding. The investigation of model fit and accuracy in estimating factor loadings yielded similar results for constrained and free factor loadings in confirmatory factor analysis. Furthermore, there were effects due to the type of link transformation and sample size.

**Keywords:** binary data, confirmatory factor analysis, link function, probability-based covariances, simulation, tetrachoric correlations

## 1. Introduction

The item discriminability is a characteristic of the model of measurement and reflects the relationship between the item and the corresponding latent attribute (Lucke, 2005). The model of measurement can show free or constrained discriminability. The shift from free discriminability to constrained discriminability reduces the adaptability to the specificities of data. Constrained discriminability characterizes the Rasch (1960) model and the corresponding one-parameter model (Birnbaum, 1968), the Rasch model-based linear logistic test model (Scheiblechner, 1972) and the tau-equivalent model (Lord & Novick, 1968). Despite the lack of discriminability, the Rasch model and the corresponding one-parameter model have so far played a major role in research and application guided by item-response theory (IRT), while the consideration of the linear logistic test model has been more or less restricted to investigations of specific effects, as for example the effects of the item position, learning and fatigue (Kubinger, 2008). These models are contrasted by the tau-equivalent model of measurement that can be employed for investigations in the framework of factor analysis. This model appears to originate from the relaxation of some constraints of the stricter parallel model of measurement, and the further relaxation of constraints leads to the less strict congeneric model of measurement (Graham, 2006).

It is the tau-equivalent model of measurement that is in the focus of the present investigation. The extended version of this model relates the  $p \times 1$  vector of observations  $\mathbf{y}$  to the  $p \times 1$  vector of intercepts  $\boldsymbol{\mu}$ , the product of the  $p \times q$  matrix of factor loadings  $\boldsymbol{\Lambda}$  and the  $q \times 1$  vector of latent variables (i.e., latent factors)  $\boldsymbol{\eta}$  and the  $p \times 1$  vector of error components  $\boldsymbol{\varepsilon}$ :

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon} . \quad (1)$$

Since the tau-equivalent model of measurement is usually considered with respect to data that give rise to the expectation of one underlying source of responding,  $\boldsymbol{\Lambda}$  is replaced by the  $p \times 1$  vector  $\boldsymbol{\lambda}$ :

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon} . \quad (2)$$

The special properties that distinguish this model of measurement from other models are equally sized factor loadings  $\lambda_1, \dots, \lambda_p$ :

$$\lambda_1 = \dots = \lambda_p \quad (3)$$

and error components  $\varepsilon_1, \dots, \varepsilon_p$  that may differ from each other.

The measurement model shows constrained discriminability because of the restriction represented by Equation 3. In an application it can be realized in two ways: on one hand, the variance of the latent variable is set equal to one while the factor loadings are estimated under the restriction that they show equal sizes; on the other hand, the factor loadings are set equal to one while the variance of the latent variable is estimated.

One reason for the wide disregard of the tau-equivalent model as compared to the other models that also apply to dichotomous data is the lack of a link transformation. This transformation establishes a relationship between observed and latent scores with respect to scale and distribution. Such a transformation is characteristic of the generalized linear model (McCullagh & Nelder, 1985; Nelder & Wedderburn, 1972; Skrondal & Rabe-Hesketh, 2004). The Rasch model includes such a transformation. It is accomplished by means of the logit as link function. The same function also applies to the linear logistic test model. Since the original tau-equivalent model does not comprise a link function, its application for investigating dichotomous data can be questioned (Schweizer, 2012).

The measurement model must be complemented by a way of conducting the link transformation in order to overcome the differences between the categorical and continuous scales and the associated distributions. The first way is characterized by the combination of the tau-equivalent model of measurement as core part of confirmatory factor analysis with tetrachoric correlations as input (Muthén, 1984, 1993). In this case the computation of tetrachoric correlations includes the shift from the categorical scale to the continuous scale and from the binomial distribution to the normal distribution (Pearson, 1900). It is accomplished by means of thresholds that refer to the normal distribution function. The second way is due to the opportunity to modify the model of measurement in such a way that a link transformation is conducted additionally (Schweizer, 2013; Schweizer & Reiss, 2014; Schweizer, Ren, & Wang, 2015). In this case probability-based covariances that are also known as a pre-stage reached in computing the Phi coefficient (McDonald & Ahlwardt, 1974) are recommended as input to confirmatory factor analysis in order to achieve interval scale. Their use presupposes a sample size of at least 200. Furthermore, weights reflecting the distributional differences between the dichotomous and continuous scores are included into the measurement model. The weights  $w_i$  that are assigned to the main diagonal of the  $p \times p$  diagonal matrix  $\mathbf{W}$  moderate the relationships between the observed and latent variables in such a way that there is an adjustment of the distributional differences between the variables regarding the variances and covariances:

$$\mathbf{y} = \boldsymbol{\mu} + (\mathbf{W}\boldsymbol{\lambda})\boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad (4)$$

The weights transform the tau-equivalent model into a weighted tau-equivalent model.

However, even in the presence of an appropriate link transformation researchers and practitioners may suspect a low degree of efficiency and avoid using the tau-equivalent model of measurement because of its reduced adaptability to the specificities of data. There are a number of properties of data that potentially exert an influence on model fit (Tanaka, 1993), and these properties may be demanding to the adaptability of the model. Some properties appear to be especially important. First there is the correctness of the model with respect to the data at hand. The tau-equivalent model can be expected to perform well in correct models and to show a higher degree of sensitivity for incorrect models than the more popular congeneric model (Jöreskog, 1971).

Second the appropriateness of the construct representation that, in the absence of specific construct facets, pertains to the number of variables is important (Hogarty, Hines, Kromrey, Ferron, & Mumford, 2005; MacCallum, Widaman, Zhang, & Hong, 1999). To some degree the number of necessary variables depends on the response format. If the response format includes two categories only, usually a larger number of variables is necessary than in the multi-category case.

Third there is the sample size that is known to exert influence on model fit and is commonly varied in simulation studies (Bandalos & Gagné, 2012; Finney & DiStefano, 2013). The sample size must be large enough to achieve stable parameter estimates. The larger the size of the sample drawn from the population, the better is the chance to attain stable estimates. However, there is also a downside of a large sample size. The downside is the effect on the sensitivity of the chi-square statistic (Bergh, 2015) that plays a major role in the evaluation of model fit and contributes to several other fit statistics. In large samples the sensitivity of the chi-square statistic is usually so high that even a minor deviation of the model from data leads to the indication of model misfit.

In the following sections of this manuscript a simulation study is reported. It investigates whether constrained discriminability means an impairment of model fit and accuracy in parameter estimation (i.e., factor loadings) as compared to free discriminability if the data are dichotomous and the model is correct. In order to identify effects

of potentially moderating factors, the type of the link transformation, the sample size and the number of variables are additionally considered.

## 2. The Simulation Study

The major objective of the simulation study was to compare models of measurement with either constrained or free discriminability regarding model fit and accuracy in parameter estimation under the condition that the data are dichotomous and the model is correct. It required confirmatory factor analysis of simulated data on the basis of models of measurement with constrained and free factor loadings. The second objective was to investigate whether the result regarding the first objective depended on the type of link transformation. The possibility of a moderating effect of sample size on the result regarding the first objective constituted the third objective. Finally there was the objective to find out whether the number of variables servings as indicators to the latent variable exerted an influence on the comparison of constrained and free discriminability.

The four objectives defined the characteristics of the data that had to be generated. The first objective demanded that the data showed a one-dimensional structure that could potentially be identified by means of a one-factor model. Such a structure could be generated by means of a uniform pattern representing the relationships among random variables (Jöreskog & Sörbom, 2001, p. 159). According to the first and second objectives the data had to be dichotomous on one hand and to be related to continuous latent scores on the other hand. This combination of properties suggested the following theoretical background: originally continuous data did undergo dichotomization in the process of the assessment of a latent construct. Since dichotomization was likely to create different degrees of easiness or difficulty and since a broad range of easiness or difficulty was found to be quite demanding to data analysis including a link transformation (Schweizer, 2013), such a range was considered as a good precondition for the projected comparisons and investigations. For detecting a possible effect of the sample size, as was necessary according to the third objective, three different sample sizes were generated:  $N = 300$ , 1000 and 2000. Furthermore, the fourth objective required different numbers of variables for representing the latent variable. In accordance to the practices of scale construction six variables were considered to be close to the lower limit whereas twelve variables were accepted as good.

### 2.1 Data Generation and Analysis

Two uniform patterns establishing relationships among six or twelve variables respectively served the construction of the simulated data. These patterns were correlation matrices including the number .32 as off-diagonal elements. This number was expected to give rise to completely standardized factor loadings of .57 (Note. The expected size was .5656 that was selected because it could be expected to create a small degree of variability when rounding the estimated factor loadings. Such variability should prevent excessively high values in using Hartley's  $F_{\max}$  test; for more details see below).

In the first step continuous random data following the normal distribution were generated:  $X \sim N(0,1)$ . The random data were arranged as matrices showing the following combinations of numbers of rows and columns:  $300 \times 12$ ,  $1000 \times 12$ ,  $2000 \times 12$ ,  $300 \times 6$ ,  $1000 \times 6$  and  $2000 \times 6$ . Since 200 replications were considered sufficient for the cells of a design (Bandalos & Gagné, 2012), this number of matrices was generated for each combination. In the second step the continuous random data were recombined by means of weights computed for the two uniform patterns. These weights were achieved by means of a procedure described by Jöreskog and Sörbom (2001, p. 159). In the third step the continuous data were transformed into dichotomous data:  $Y \sim \text{Bin}(2,p)$ . In order to arrive at a broad range of easiness or difficulty, different proportions were selected for the columns of the matrices.

The transformation of continuous into dichotomous data in the third step was conducted according to the following proportions: .900, .827, .755, .682, .609, .536, .464, .391, .318, .245, .173 and .100 (i.e., number of selected events / number of all events) for matrices including twelve columns. In the first column of a matrix the 10 percent smaller numbers were transformed into zeros and the 90 percent larger numbers into ones. In the second column zeros replaced the 17.3 percent smaller numbers whereas the other numbers were changed into ones. The third to twelfth columns were processed in the same way in considering the other proportions of the list in corresponding order. In the matrices including six columns the proportions for splits were: .900, .740, .580, .420, .260 and .100. Subsequently, the dichotomous data served the computation of tetrachoric correlations and probability-based covariances. Additionally customary covariances were computed from the continuous data.

The objectives of the study required the consideration of four methods of conducting confirmatory factor analysis. These methods differed in characteristics of the model of measurement and the type of input matrix. The first method combined the congeneric model of measurement (Jöreskog, 1971) and tetrachoric correlations

as input (TetCon). The combination of the congeneric model comprising weights and probability-based covariances as input gave rise to the second method. The second model was addressed as weighted congeneric model (PbCWCon). Free factor loadings meaning free discriminability characterized the models of measurement of the first and second methods. The other methods differed from the described methods by using constrained factor loadings instead of free ones. Because of the assumed uniformity of the relationships among the variables induced by the uniform patterns, equal-sized constraints were necessary. The modification gave rise to the third method characterized by the constrained model of measurement and tetrachoric correlations as input (TetTau) and the fourth method that also included weighted and constrained factor loadings and probability-based covariances as input (PbCWTau).

Since all the methods included provisions regarding deviations from the normal distributions, the parameter estimation was conducted by means of the maximum likelihood method. The following fit indexes were considered in the evaluation of the results: chi-square, degrees of freedom, normed chi-square, RMSEA, SRMR, CFI, TLI and GFI. Cut-offs provided by Kline (2005) and Hu and Bentler (1999) served the evaluation of the results ( $RMSEA \leq .06$ ,  $SRMR \leq .08$ ,  $CFI \geq .95$ ,  $TLI \geq .95$ ,  $GFI \geq .90$ ). Because of the known dependence of chi-square on sample size two different cut-offs regarding normed chi-square were taken into consideration (2 for  $N = 300$  and 5 for  $N = 1000$  and larger). Means and standard deviations were computed for each set of 200 matrices. In order to facilitate the reading of the tables and the aggregation of the results, the superscript “M” was added to a mean if this mean proved to be good in comparison to the corresponding cut-off. If additionally the confidence interval meaning 95 percent of the distribution of the observed results was beyond the cut-off in the range of good results, it was replaced by the superscript “CI”. For arriving at conclusions extending to different methods and data characteristics, the “M”s and “CI”s were counted in giving the weight 1.0 to each “M” and 2.0 to each “CI”. Finally the counts were used for comparisons.

The accuracy of the completely standardized factor loadings was not only important regarding freely estimated factor loadings but also constrained ones because the average size of factor loadings varied as a result of parameter estimation and because of the item-specific contributions of error in standardization. Two aspects marked the investigation of the completely standardized factor loadings. First, the absence of the dependence on the item marginal (Kubinger, 2003; Torgerson, 1958) was checked. Dependence on the item marginal was obvious from a systematic deviation of the observed factor loadings from the expected factor loadings. In the simulated data equal sizes of the completely standardized factor loadings signified the absence of this kind of dependence. This check was conducted by means of Hartley's  $F_{\max}$  test. Since the same size was expected for each completely standardized factor loading, the variances of these factor loadings obtained in investigating dichotomous data at the level of the means were compared with those achieved in investigating continuous data at the level of the means. In this check the  $F_{\max}$  statistic served more as a descriptive statistic than a significance test. Since the variability of the means observed in continuous data that served as denominator of  $F_{\max}$  was extremely small, it appeared to be overly sensitive to deviations. Second, the general size of the completely standardized factor loadings was checked.

## 2.2 The Results of Investigating Model-data Fit

In this section the results regarding model-data fit are presented. At first the results of investigating the covariance matrices computed from continuous data following the normal distribution by means of the congeneric model

Table 1. Means and Standard Deviations (in Parentheses) of the Fit Results for the Continuous Data in Different Sample Sizes ( $N_p = 300, 1000, 2000$ ) and Numbers of Variables ( $N_v = 6, 12$ ) Based on 200 Matrices

Input	$N_v$	$N_p$	$\chi^2$	df	Normed $\chi^2$	RMSEA	SRMR	CFI	TLI	GFI
Covariance	12	300	54.0 (10.3)	54	1.0 <sup>CI</sup> (0.2)	.012 <sup>CI</sup> (0.02)	.033 <sup>CI</sup> (0.00)	1.00 <sup>CI</sup> (0.00)	1.00 <sup>CI</sup> (0.01)	0.97 <sup>CI</sup> (0.01)
Covariance	12	1000	53.4 (10.9)	54	1.0 <sup>CI</sup> (0.2)	.012 <sup>CI</sup> (0.02)	.018 <sup>CI</sup> (0.00)	1.00 <sup>CI</sup> (0.00)	1.00 <sup>CI</sup> (0.00)	0.99 <sup>CI</sup> (0.00)
Covariance	12	2000	54.8 (10.9)	54	1.0 <sup>CI</sup> (0.2)	.004 <sup>CI</sup> (0.01)	.013 <sup>CI</sup> (0.00)	1.00 <sup>CI</sup> (0.00)	1.00 <sup>CI</sup> (0.00)	1.00 <sup>CI</sup> (0.00)
Covariance	6	300	9.1 (4.8)	9	1.0 <sup>CI</sup> (0.5)	.016 <sup>CI</sup> (0.01)	.025 <sup>CI</sup> (0.01)	1.00 <sup>CI</sup> (0.01)	1.00 <sup>CI</sup> (0.02)	0.99 <sup>CI</sup> (0.01)
Covariance	6	1000	9.2 (4.5)	9	1.0 <sup>CI</sup> (0.5)	.008 <sup>CI</sup> (0.01)	.014 <sup>CI</sup> (0.01)	1.00 <sup>CI</sup> (0.00)	1.00 <sup>CI</sup> (0.00)	1.00 <sup>CI</sup> (0.00)
Covariance	6	2000	8.9 (4.0)	9	1.0 <sup>CI</sup> (0.4)	.006 <sup>CI</sup> (0.01)	.010 <sup>CI</sup> (0.00)	1.00 <sup>CI</sup> (0.00)	1.00 <sup>CI</sup> (0.00)	1.00 <sup>CI</sup> (0.00)

Note. <sup>CI</sup> The 95 percent confidence interval indicates good fit. <sup>M</sup> The mean indicates good fit.

of measurement are presented. They serve as a comparison level for the results obtained from dichotomous data, especially for gaining variances that are necessary for computing the  $F_{\max}$  statistic.

The results presented in the upper half of Table 1 stem from matrices including 12 columns and the results of the lower half from matrices including 6 columns. The numbers that are not in parentheses are means and the numbers written in parentheses are standard deviations. All fit statistics obtained in investigating the continuous data and presented in this Table revealed that there was an overall good model-data fit. In all cases the confidence interval was completely below or above the corresponding cut-off. Furthermore, the sample size showed virtually no influence on model-data fit. Even the chi-square statistic did not display an influence of the sample size. Only the chi-squares observed in matrices including either 6 or 12 columns differed from each other. Matrices including twelve columns led to the larger chi-squares.

Table 2 provides the results observed in using the four methods described in the method section for investigating the matrices of dichotomous data comprising the larger number of columns ( $N_V = 12$ ).

Table 2. Means and Standard Deviations (in Parentheses) of the Fit Results Obtained for Binary Data in Considering Different Parameters Types (PType), Link Types (LType) and Sample Sizes ( $N_p = 300, 1000, 2000$ ) Based on 200 Matrices Including 12 Columns

PType	LType	$N_p$	$\chi^2$	df	Normed $\chi^2$	RMSEA	SRMR	CFI	TLI	GFI
Free <sup>1</sup>	TCorr <sup>3</sup>	300	343.3(198.2)	54	6.4 (3.7)	.124 (0.05)	.080 <sup>M</sup> (0.02)	0.77 (0.17)	0.72 (0.21)	0.85 (0.08)
		1000	418.4(293.5)	54	7.7 (5.4)	.078 (0.02)	.047 <sup>CI</sup> (0.01)	0.94 (0.05)	0.93 (0.06)	0.94 <sup>M</sup> (0.04)
		2000	346.3(163.6)	54	6.4 (3.0)	.051 <sup>M</sup> (0.01)	.031 <sup>CI</sup> (0.00)	0.98 <sup>CI</sup> (0.01)	0.97 <sup>M</sup> (0.02)	0.97 <sup>CI</sup> (0.01)
	WCov <sup>4</sup>	300	71.2 (12.5)	54	1.3 <sup>CI</sup> (0.2)	.016 <sup>CI</sup> (0.01)	.025 <sup>CI</sup> (0.00)	0.99 <sup>CI</sup> (0.02)	0.99 <sup>CI</sup> (0.02)	0.99 <sup>CI</sup> (0.01)
		1000	71.2 (13.8)	54	1.3 <sup>CI</sup> (0.3)	.016 <sup>CI</sup> (0.01)	.025 <sup>CI</sup> (0.00)	0.99 <sup>CI</sup> (0.01)	0.99 <sup>CI</sup> (0.01)	0.99 <sup>CI</sup> (0.01)
		2000	78.3 (14.6)	54	1.5 <sup>CI</sup> (0.3)	.014 <sup>CI</sup> (0.01)	.019 <sup>CI</sup> (0.00)	0.99 <sup>CI</sup> (0.00)	0.99 <sup>CI</sup> (0.00)	0.99 <sup>CI</sup> (0.00)
Fixed <sup>2</sup>	TCorr <sup>3</sup>	300	334.9(154.5)	65	5.2 (2.4)	.111 (0.04)	.097(0.02)	0.75 (0.18)	0.75 (0.18)	0.85 (0.06)
		1000	418.3(212.3)	65	6.4 (3.3)	.072 (0.02)	.058 <sup>CI</sup> (0.01)	0.94 (0.05)	0.93 (0.05)	0.94 (0.03)
		2000	369.1(144.6)	65	5.7 (2.2)	.048 <sup>M</sup> (0.01)	.039 <sup>CI</sup> (0.01)	0.98 <sup>CI</sup> (0.01)	0.98 <sup>CI</sup> (0.01)	0.97 <sup>CI</sup> (0.01)
	WCov <sup>4</sup>	300	72.5 (13.5)	65	1.1 <sup>CI</sup> (0.2)	.018 <sup>CI</sup> (0.01)	.057 <sup>CI</sup> (0.01)	0.98 <sup>M</sup> (0.02)	0.99 <sup>CI</sup> (0.02)	0.96 <sup>CI</sup> (0.01)
		1000	94.9 (15.1)	65	1.5 <sup>CI</sup> (0.2)	.021 <sup>CI</sup> (0.01)	.039 <sup>CI</sup> (0.00)	0.98 <sup>CI</sup> (0.01)	0.98 <sup>CI</sup> (0.01)	0.98 <sup>CI</sup> (0.00)
		2000	128.1 (17.6)	65	2.0 <sup>CI</sup> (0.3)	.022 <sup>CI</sup> (0.00)	.034 <sup>CI</sup> (0.00)	0.98 <sup>CI</sup> (0.00)	0.98 <sup>CI</sup> (0.00)	0.99 <sup>CI</sup> (0.00)

Note. <sup>1</sup> Free factor loadings. <sup>2</sup> Constrained factor loadings. <sup>3</sup> Tetrachoric correlations. <sup>4</sup> Weights and probability-based covariances. <sup>CI</sup> The 95 percent confidence interval indicates good fit. <sup>M</sup> The mean indicates good fit.

The first to sixth rows include results due to free factor loadings (TetCon, PbCWCon) and the seventh to twelfth rows contain results achieved by means of constrained factor loadings (TetTau, PbCWTau). The results of the second set of rows virtually mirrored the results of the first set of rows. The counting of the weighted “M”s and “CI”s of the first and second sets led to scores of 48 and 46 for free and constrained factor loadings respectively. These numbers indicated that there was virtually no effect due to the type of factor loadings. In contrast, the results that were specific for the types of link transformation suggested dissimilarity. The majority of the results based on tetrachoric correlations (TetCon, TetTau) reported in the first to third and seventh to ninth rows indicated a bad model-data fit whereas the link transformation by weights and probability-based correlations (PbCWCon, PbCWTau) reported in the fourth to sixth and tenth to twelfth rows signified good degrees of model fit. The corresponding overall scores were 23 and 71. There was also an effect of the sample size. The scores for the sample sizes of 300, 1000 and 2000 were 24, 29 and 41 respectively. According to these scores the increase in sample size improved the model fit.

Furthermore, matrices comprising the smaller number of columns ( $N_V = 6$ ) were also investigated in order to find out whether the number of variables was important regarding model fit. The investigation of the matrices comprising 6 columns yielded the results of Table 3.

Table 3. Means and Standard Deviations (in Parentheses) of the Fit Results Obtained for Binary Data in Considering Different Parameters Types (PType), Link Types (LType) and Sample Sizes ( $N_p = 300, 1000, 2000$ ) Based on 200 Matrices Including 6 Columns

PType	LType	$N_p$	$\chi^2$	df	Normed $\chi^2$	RMSEA	SRMR	CFI	TLI	GFI
Free <sup>1</sup>	TCorr <sup>3</sup>	300	129.1(101.2)	9	14.3(11.2)	.192 (0.08)	.115 <sup>M</sup> (0.07)	0.80 (0.13)	0.66 (0.22)	0.88 (0.08)
		1000	83.6 (91.1)	9	9.3 (10.1)	.082 (0.04)	.040 <sup>CI</sup> (0.02)	0.96 (0.04)	0.93 (0.07)	0.97 <sup>M</sup> (0.03)
		2000	75.2 (48.1)	9	8.4 (5.3)	.057 <sup>M</sup> (0.02)	.027 <sup>CI</sup> (0.01)	0.98 <sup>CI</sup> (0.01)	0.96 <sup>M</sup> (0.02)	0.99 <sup>CI</sup> (0.01)
	WCov <sup>4</sup>	300	9.4 (4.6)	9	1.0 <sup>CI</sup> (0.5)	.017 <sup>CI</sup> (0.02)	.032 <sup>CI</sup> (0.01)	0.98 <sup>CI</sup> (0.02)	1.00 <sup>CI</sup> (0.03)	0.99 <sup>CI</sup> (0.01)
		1000	11.5 (5.5)	9	1.3 <sup>CI</sup> (0.6)	.014 <sup>CI</sup> (0.01)	.019 <sup>CI</sup> (0.00)	0.99 <sup>CI</sup> (0.01)	0.99 <sup>CI</sup> (0.01)	1.00 <sup>CI</sup> (0.00)
		2000	14.2 (6.9)	9	1.5 <sup>CI</sup> (0.8)	.015 <sup>CI</sup> (0.01)	.015 <sup>CI</sup> (0.00)	0.99 <sup>CI</sup> (0.00)	0.99 <sup>CI</sup> (0.01)	1.00 <sup>CI</sup> (0.00)
Fixed <sup>2</sup>	TCorr <sup>3</sup>	300	187.5(138.0)	14	13.4 (9.9)	.185 (0.09)	.132 (0.05)	0.59 (0.32)	0.53 (0.44)	0.84 (0.10)
		1000	135.6(218.1)	14	9.7 (15.6)	.079 (0.05)	.061 <sup>M</sup> (0.03)	0.93 (0.11)	0.93 (0.11)	0.96 <sup>M</sup> (0.05)
		2000	93.6 (50.6)	14	6.7 (3.6)	.051 <sup>M</sup> (0.02)	.040 <sup>CI</sup> (0.01)	0.97 <sup>CI</sup> (0.02)	0.97 <sup>CI</sup> (0.02)	0.98 <sup>CI</sup> (0.01)
	WCov <sup>4</sup>	300	17.4(6.2)	14	1.2 <sup>CI</sup> (0.4)	.025 <sup>CI</sup> (0.02)	.053 <sup>CI</sup> (0.01)	0.96 <sup>M</sup> (0.04)	0.97 <sup>M</sup> (0.06)	0.98 <sup>CI</sup> (0.01)
		1000	25.3(7.2)	14	1.8 <sup>CI</sup> (0.5)	.027 <sup>CI</sup> (0.01)	.037 <sup>CI</sup> (0.01)	0.97 <sup>CI</sup> (0.02)	0.97 <sup>CI</sup> (0.02)	0.99 <sup>CI</sup> (0.00)
		2000	39.7(10.6)	14	2.8 <sup>CI</sup> (0.8)	.029 <sup>CI</sup> (0.01)	.033 <sup>CI</sup> (0.00)	0.97 <sup>CI</sup> (0.01)	0.96 <sup>CI</sup> (0.01)	0.99 <sup>CI</sup> (0.00)

Note. <sup>1</sup> Free factor loadings. <sup>2</sup> Constrained factor loadings. <sup>3</sup> Tetrachoric correlations. <sup>4</sup> Weights and probability-based covariances. <sup>CI</sup> The 95 percent confidence interval indicates good fit. <sup>M</sup> The mean indicates good fit.

This Table shows the same structure as Table 2. The results reported in this Table proved to be very similar to the results reported in Table 2. For example, the overall score of Table 2 was 94 and of this Table 93. Only the chi-squares included in the two Tables differed from each other considerably. However, this kind of difference was not unexpected and not considered as a problem regarding model fit.

Because of the high degree of similarity of results of the two Tables the final evaluation extended to the results of both Tables. First the scores characterizing free and constrained factor loadings (96 and 91) were compared by means of the chi-square test. The difference was not substantial ( $\chi^2 = 0.13$ ,  $df = 1$ , n.s.). Second the two ways of conducting the link transformation that yielded scores of 141 and 46 were compared and found to differ from each other ( $\chi^2 = 48.26$ ,  $df = 1$ ,  $p < .05$ ). Third there were the three sample sizes giving rise to the scores of 47, 58 and 82. The chi-square test indicated a significant difference ( $\chi^2 = 10.27$ ,  $df = 2$ ,  $p < .05$ ). Fourth the scores characterizing the different numbers of variables were compared. They did not differ from each other ( $\chi^2 = 0.005$ ,  $df = 1$ , n.s.). Because of the substantial differences regarding the ways of conducting the link transformation and the sample sizes the two types of factor loadings were also compared separately for each way of conducting the link transformation and for each sample size. In none of these comparisons a significant difference was observed (regarding the ways of conducting the link transformation:  $p = .77$ ,  $p = .80$ ; different sample sizes:  $p = .47$ ,  $p = .79$ ,  $p = .82$ ).

In sum, constrained and free factor loadings as the two types of discriminability in confirmatory factor analysis do not differ from each other regarding model fit if the model is correct. Furthermore, the way of conducting the link transformation and the sample size influence model fit whereas the number of variables does not.

### 2.3 The Results of Investigating Accuracy in Parameter Estimation

In this section the results of investigating the accuracy of the completely standardized factor loadings are reported. The means and standard deviations of the completely standardized factor loadings achieved in investigating the covariance matrices computed from continuous data are presented in Table 4.

Table 4. Means and Standard Deviations (in Parentheses) of the Completely Standardized Factor Loadings Obtained for the Covariance Matrices (Mat) Computed from Continuous Data in Different Sample Sizes ( $N_p = 300, 1000, 2000$ ) and Matrices Including 6 or 12 Columns Based on 200 Matrices

Matrix type	$N_p$	Position of factor loading											
		1	2	3	4	5	6	7	8	9	10	11	12
Covariance	300	.56(.05)	.56(.04)	.57(.05)	.57(.04)	.56(.05)	.56(.04)	.56(.05)	.57(.04)	.57(.05)	.56(.04)	.57(.05)	.56(.04)
Covariance	1000	.57(.02)	.56(.02)	.57(.02)	.56(.03)	.57(.03)	.57(.02)	.56(.02)	.56(.03)	.57(.02)	.56(.03)	.57(.02)	.56(.02)
Covariance	2000	.56(.02)	.57(.02)	.56(.02)	.57(.02)	.57(.02)	.57(.02)	.57(.02)	.57(.02)	.57(.02)	.57(.02)	.57(.02)	.57(.02)
Covariance	300	.56(.05)	.55(.05)	.57(.05)	.57(.05)	.56(.05)	.56(.05)						
Covariance	1000	.57(.03)	.56(.03)	.57(.03)	.57(.03)	.57(.03)	.57(.03)						
Covariance	2000	.56(.02)	.56(.02)	.56(.02)	.57(.02)	.57(.02)	.57(.02)						

The means varied between .55 and .57. This kind of variation was expected because of the necessity of rounding parameter estimates. The sample size did not show any influence on the variability of the means nor did the number of variables. However, a decrease of the standard deviation of the individual results from 0.05 ( $N = 300$ ) to 0.02 ( $N = 2000$ ) was observed when the sample size increased.

The results obtained in investigating the completely standardized factor loadings computed for the matrices of binary data including 12 columns are presented in Table 5. The first to third rows of this Table provide the completely standardized factor loadings for the tetrachoric correlations in combination with the congeneric model of measurement (TetCon). In the sample size of 300 the mean factor loadings varied between .49 and .57, in 1000 between .56 and .59 and in 2000 between .56 and .57. Furthermore, the  $F_{\max}$  test results indicated deviations from the expected equality of the completely standardized factor loadings for the smaller sample sizes ( $N = 300$ :  $F_{\max}(2,11) = 167.37, p < .05$ ;  $N = 1000$ :  $F_{\max}(2,11) = 22.79, p < .05$ ;  $N = 2000$ :  $F_{\max}(2,11) = 1.15, n.s.$ ).

Quite different results were observed for the tetrachoric correlations in combination with the model including constrained factor loadings (TetTau) (see the fourth to sixth rows of Table 5). The mean factor loadings varied between .53 and .54 when the sample size was 300. In the other sample sizes the mean factor loading was always .57. The  $F_{\max}$  statistic indicated a large deviation from the expected equality only for the sample size of 300 ( $N = 300$ :  $F_{\max}(2,11) = 133.37, p < .05$ ;  $N = 1000$ :  $F_{\max}(2,11) = 0.00, n.s.$ ;  $N = 2000$ :  $F_{\max}(2,11) = 0.00, n.s.$ ).

The completely standardized factor loadings obtained on the basis of probability-based covariances by means of the congeneric model (PbCWCon) are reported in the seventh to ninth rows of Table 5. In the sample sizes of 300 and 1000 the means varied between .56 and .58 and in the sample size of 2000 between .57 and .58. Investigating the dependence on the item marginal revealed significant deviations ( $N = 300$ :  $F_{\max}(2,11) = 9.06, p < .05$ ;  $N = 1000$ :  $F_{\max}(2,11) = 5.13, p < .05$ ;  $N = 2000$ :  $F_{\max}(2,11) = 3.71, p = .05$ ). In the case of the sample size of 2000 the ratio was only marginally significant, and in the other cases the  $F_{\max}$  values were considerably smaller than the  $F_{\max}$  values found for tetrachoric correlations.

Table 5. Means and Standard Deviations (in Parentheses) of the Completely Standardized Factor Loadings Computed from Binary Data in Different Sample Sizes ( $N_p = 300, 1000, 2000$ ) and Matrices Including 12 Columns Based on 200 Matrices in Considering Different Parameters Types (PType) and Link Types (LType)

PType/ LType	$N_p$	Position of factor loading											
		1	2	3	4	5	6	7	8	9	10	11	12
Free <sup>1</sup> /	300	.57(.16)	.53(.15)	.51(.12)	.50(.12)	.50(.11)	.49(.10)	.50(.11)	.49(.11)	.49(.11)	.50(.12)	.54(.14)	.56(.16)
TCorr <sup>3</sup>	1000	.59(.08)	.56(.05)	.57(.05)	.56(.04)	.56(.04)	.56(.04)	.56(.04)	.56(.04)	.56(.04)	.57(.05)	.56(.05)	.59(.08)
	2000	.57(.04)	.57(.04)	.56(.03)	.56(.03)	.56(.03)	.57(.03)	.56(.03)	.57(.03)	.57(.03)	.57(.03)	.57(.03)	.57(.04)
Fixed <sup>2</sup> /	300	.54(.08)	.53(.08)	.53(.08)	.53(.08)	.53(.08)	.53(.07)	.53(.07)	.53(.07)	.53(.07)	.53(.07)	.53(.07)	.53(.07)
TCorr <sup>3</sup>	1000	.57(.05)	.57(.04)	.57(.04)	.57(.04)	.57(.04)	.57(.04)	.57(.04)	.57(.04)	.57(.04)	.57(.04)	.57(.04)	.57(.05)
	2000	.57(.02)	.57(.01)	.57(.01)	.57(.01)	.57(.01)	.57(.01)	.57(.01)	.57(.01)	.57(.01)	.57(.01)	.57(.01)	.57(.02)
Free <sup>1</sup> /	300	.56(.09)	.57(.08)	.57(.07)	.58(.06)	.58(.07)	.58(.06)	.58(.07)	.58(.07)	.57(.07)	.57(.07)	.58(.08)	.57(.09)
WCov <sup>4</sup>	1000	.57(.05)	.56(.04)	.57(.04)	.58(.03)	.58(.03)	.58(.04)	.58(.04)	.58(.03)	.58(.04)	.58(.04)	.57(.04)	.57(.04)
	2000	.57(.03)	.57(.03)	.57(.03)	.57(.03)	.58(.02)	.58(.03)	.58(.02)	.58(.03)	.57(.03)	.57(.03)	.57(.03)	.57(.03)
Fixed <sup>2</sup> /	300	.57(.02)	.58(.02)	.58(.02)	.59(.02)	.59(.02)	.59(.02)	.59(.02)	.59(.02)	.59(.02)	.58(.02)	.58(.02)	.57(.02)
WCov <sup>4</sup>	1000	.57(.01)	.58(.01)	.58(.01)	.59(.01)	.59(.01)	.59(.01)	.59(.01)	.59(.01)	.59(.01)	.58(.01)	.58(.01)	.57(.01)
	2000	.57(.01)	.58(.01)	.58(.01)	.58(.01)	.59(.01)	.59(.01)	.59(.01)	.59(.01)	.58(.01)	.58(.01)	.58(.01)	.57(.01)

Note. <sup>1</sup> Free factor loadings. <sup>2</sup> Constrained factor loadings. <sup>3</sup> Tetrachoric correlations. <sup>4</sup> Weights and probability-based covariances.

Finally, there were the results achieved in investigating probability-based covariances in considering the model with constrained factor loadings (PbCWTau) (see the tenth to twelfth rows of Table 5). There was variation of the means between .57 and .59 in all sample sizes. Furthermore, the investigations of the equality of the factor loadings by means of the  $F_{\max}$  test revealed substantial differences for all sample sizes ( $N = 300$ :  $F_{\max}(2,11) = 7.69$ ,  $p < .05$ ;  $N = 1000$ :  $F_{\max}(2,11) = 7.29$ ,  $p < .05$ ;  $N = 2000$ :  $F_{\max}(2,11) = 7.19$ ,  $p < .05$ ). Given the significance of the  $F_{\max}$  values, again it needs to be highlighted that these  $F_{\max}$  values were small in comparison to the values observed for tetrachoric correlations computed from the smaller datasets.

Next the degree of correspondence of the overall observed and expected sizes of the mean factor loadings was evaluated in considering the different sample sizes. The tetrachoric correlations in combination with the congeneric model of measurement (TetCon) led to the mean factor loadings of .52, .57 and .57 ( $SD = 0.12, 0.05, 0.03$ ) for the sample sizes of 300, 1000 and 2000. The replacement of the congeneric model by the model with constrained factor loadings (TetTau) yielded mean factor loadings of .53, .57 and .57 ( $SD = 0.07, 0.04, 0.01$ ) for the sample sizes of 300, 1000 and 2000. The investigation of probability-based covariances in considering the congeneric model (PbCWCon) yielded mean factor loadings of .57 ( $SD = 0.07, 0.04, 0.03$ ). These covariances in combination with the model including constrained factor loadings (PbCWTau) led to mean results of .58 ( $SD = 0.02, 0.01, 0.01$ ) for all the sample sizes.

The matrices with 6 columns were investigated in exactly the same way as the matrices with 12 columns. The results of these investigations are included in Table 6. The results of the first to third rows were obtained using tetrachoric correlations in combination with the congeneric model of measurement (TetCon). In the sample size of 300 the results are based on 72 percent of the 200 matrices only. The other matrices were eliminated since the completely standardized factor loadings were either larger than 1.0 or smaller than -1.0. In the remaining matrices the mean factor loadings varied between .52 and .61. In the sample size of 1000 the means were between .55 and .60 and in 2000 between .56 and .57. Furthermore, the  $F_{\max}$  test results indicated deviations from the expected equality of the completely standardized factor loadings for the smaller sample sizes ( $N = 300$ :  $F_{\max}(2,5) = 221.71$ ,  $p < .05$ ;  $N = 1000$ :  $F_{\max}(2,5) = 94.87$ ,  $p < .05$ ;  $N = 2000$ :  $F_{\max}(2,5) = 3.38$ , n.s.).

Table 6. Means and Standard Deviations (in Parentheses) of the Completely Standardized Factor Loadings Computed from Binary Data in Different Sample Sizes ( $N_p = 300, 1000, 2000$ ) and Matrices Including 6 Columns Based on 200 Matrices in Considering Different Parameters Types (PType) and Link Types (LType)

PType/ LType	$N_p$	Position of factor loading					
		1	2	3	4	5	6
Free <sup>1</sup> /	300 <sup>5</sup>	.61(.21)	.53(.13)	.52(.13)	.52(.14)	.52(.14)	.59(.20)
TCorr <sup>3</sup>	1000	.60(.12)	.55(.07)	.55(.07)	.55(.07)	.55(.08)	.60(.10)
	2000	.57(.05)	.56(.04)	.56(.04)	.57(.04)	.57(.04)	.57(.06)
Fixed <sup>2</sup> /	300	.61(.09)	.58(.06)	.58(.06)	.58(.06)	.58(.06)	.60(.08)
TCorr <sup>3</sup>	1000	.58(.04)	.57(.03)	.57(.03)	.57(.03)	.57(.03)	.58(.03)
	2000	.57(.02)	.57(.02)	.57(.02)	.57(.02)	.57(.02)	.57(.02)
Free <sup>1</sup> /	300	.55(.11)	.56(.09)	.58(.08)	.58(.09)	.57(.09)	.56(.11)
WCov <sup>4</sup>	1000	.57(.06)	.57(.05)	.58(.04)	.58(.05)	.57(.05)	.57(.05)
	2000	.56(.03)	.56(.03)	.58(.03)	.58(.03)	.57(.03)	.57(.04)
Fixed <sup>2</sup> /	300	.55(.03)	.56(.03)	.56(.03)	.56(.03)	.56(.03)	.55(.03)
WCov <sup>4</sup>	1000	.55(.02)	.56(.02)	.57(.02)	.57(.02)	.56(.02)	.55(.02)
	2000	.55(.01)	.56(.01)	.56(.01)	.56(.01)	.56(.01)	.55(.01)

Note. <sup>1</sup> Free factor loadings. <sup>2</sup> Constrained factor loadings. <sup>3</sup> Tetrachoric correlations. <sup>4</sup> Weights and probability-based covariances. <sup>5</sup> Datasets with factor loading  $>1.0$  or  $<-1.0$  were excluded leading to a reduction by 28 percent.

The replacement of the congeneric model by the model with constrained factor loadings (TetTau) led to quite different results (see the second part of Table 6). The mean factor loadings varied between .58 and .61 when the sample size was 300 and between .57 and .58 when the sample size was 1000. In the other sample size the mean was always .57. It indicated a deviation from the expected equality for the smallest sample size only ( $N = 300$ :



$F_{\max}(2,5) = 22.72, p < .05; N = 1000: F_{\max}(2,5) = 1.11, \text{ n.s.}; N = 2000: F_{\max}(2,5) = 0.00, \text{ n.s.}$ ). Furthermore, the lack of variance in the sample size of 2000 signified that there was no deviation from the equality in size.

The investigation of the completely standardized factor loadings achieved for probability-based covariances and the congeneric model (PbCWCon) led to the results reported in the third part of Table 6. In the sample sizes of 300 there was variation between .55 and .58, in the sample size of 1000 between .57 and .58 and in the sample size of 2000 between .56 and .58. The investigation of the dependence on the item marginal revealed two substantial deviations ( $N = 300: F_{\max}(2,5) = 63.13, p < .05; N = 1000: F_{\max}(2,5) = 4.23, \text{ n.s.}, N = 2000: F_{\max}(2,5) = 14.10, p < .05$ ).

The investigation of probability-based covariances in considering the model with constrained factor loading (PbCWTau) yielded the results reported in the last part of Table 6. There was variation of the mean factor loadings between .55 and .56 in sample sizes of 300 and 2000 and between .55 and .57 in the sample size of 1000. Furthermore, the investigation of the equality of the factor loadings by means of the  $F_{\max}$  test revealed no substantial deviation from the expected equal sizes ( $N = 300: F_{\max}(2,5) = 3.45, \text{ n.s.}; N = 1000: F_{\max}(2,5) = 2.96, \text{ n.s.}; N = 2000: F_{\max}(2,5) = 2.80, \text{ n.s.}$ ).

Finally, there were the results of investigating the correspondence of the observed and expected sizes across different sample sizes. The tetrachoric correlations in combination with the congeneric model of measurement (TetCon) led to the mean factor loadings of .55, .57 and .57 ( $SD = 0.16, 0.09, 0.04$ ) for the sample sizes of 300, 1000 and 2000 respectively. The replacement of the congeneric model by the model with constrained factor loadings (TetTau) yielded mean factor loadings of .59, .57 and .57 ( $SD = 0.07, 0.03, 0.02$ ) for the sample sizes of 300, 1000 and 2000 respectively. In probability-based covariances in combination with the congeneric model (PbCWCon) all mean factor loadings were .57 ( $SD = 0.09, 0.05, 0.03$ ). These covariances in combination with the model with constrained factor loadings (PbCWTau) led to .56 ( $SD = 0.03, 0.02, 0.01$ ) for all sample sizes.

In sum, according to the results presented in the Tables 5 and 6 the constrained factor loadings showed consistency instead of variability in seven out of twelve cases whereas the free factor loadings in only three out of twelve cases. In six out of twelve cases the link transformation by means of tetrachoric correlations led to consistent factor loadings; otherwise consistency was found in four out of twelve cases. Regarding sample size there was mainly variability in the smallest sample size and mostly consistency in the largest one.

Most of the mean factor loadings were quite close to the expected factor loadings. Only the mean factor loadings computed for the smallest sample size and on the basis of tetrachoric correlations showed major deviations from the expected value. Furthermore, it needs to be highlighted that there were considerable differences regarding variability. In most cases constrained factor loadings were associated with smaller standard deviations than free factor loadings. Furthermore, the standard deviations of factor loadings due to weights and probability-based covariances were smaller than those of factor loadings based on tetrachoric correlations.

### 3. Discussion

Regarding the main objective that is the comparison of constrained and free discriminability realized as constrained and free factor loadings in confirmatory factor analysis, the outcomes of the simulation study suggest no difference. Constrained and free discriminability do not differ according to the degrees of model fit, and the result of the investigation of the parameter estimates is also suggestive of similarity instead of difference. In a way these results are surprising since free factor loadings mean a high degree of adaptability to specificities of data whereas the other type of factor loadings does not (Graham, 2006). The reason for the good performance despite constrained discriminability is presumably the correctness of the model. In correct models the kind of discriminability does not count. The difference between high and low adaptability to specificities of data is presumably important if the model is only partly correct. Consequently, constrained discriminability is more than just an option that is useful for overcoming estimation problems, as has been suggested (Millsap, 2001). Furthermore, constrained discriminability already proved useful in research using models designed to represent precise structural expectations (Schweizer, 2006a, 2006b, 2008, 2009).

The comparison of the results observed for different sample sizes reveals differences between the two ways of conducting the link transformation. The effect of the sample size is minor when the transformation employs weights in combination with probability-based covariances. But it is large for the transformation by means of tetrachoric correlations. Presumably the use of robust maximum likelihood estimation (Bryant & Satorra, 2012; Satorra & Bentler, 1994) would have improved the results, as is recommended by Finney and DiStefano (2013). However, there is also reason for abstaining from robust estimation. The first reason is that the theoretical foundation of tetrachoric correlation is so conclusive that insufficiency is not expected. Another reason is provided by the results of the simulation study for the largest sample size that are really good.

Finally, there are the results of investigating the possible dependence on the item marginals (Kubinger, 2003; Torgerson, 1958) by means of the  $F_{\max}$  test. These results are important since the absence of dependence on the item marginals is indicative of the appropriateness of the link transformation. Without a link transformation the factor loadings of very easy and very difficult items are likely to differ considerably from those of the other items and give rise to a large variance. On the other hand a large variance assigned to the numerator of the  $F_{\max}$  test leads to a large  $F_{\max}$  value that means significance. However, significance needs to be interpreted with caution since the variance of the denominator of the  $F_{\max}$  test that is computed from continuous data is very small.

### Acknowledgments

The authors are grateful to the editor and anonymous reviewers for valuable comments.

### References

- Bandalos, D. L., & Gagné, P. (2012). Simulation methods in structural equation modeling. In R. H. Rick (Ed.), *Handbook of structural equation modeling* (pp. 92-108). New York, NY: Guilford Press.
- Bergh, D. (2015). Chi-squared test of fit and sample size - a comparison between a random sample approach and a chi-square value adjustment method. *Journal of Applied Measurement*, 16, 204-217.
- 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.
- Bryant, F.B., & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling*, 19, 373-398. <http://dx.doi.org/10.1080/10705511.2012.687671>
- Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 439-492). Charlotte, NC: Information Age Publishing, Inc.
- Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability. *Educational and Psychological Measurement*, 66, 930-944. <http://dx.doi.org/10.1177/0013164406288165>
- Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M., & Mumford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination. *Educational and Psychological Measurement*, 65, 202-226. <http://dx.doi.org/10.1177/0013164404267287>
- 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. <http://dx.doi.org/10.1080/10705519909540118>
- Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109-133. <http://dx.doi.org/10.1007/BF02291393>
- 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: Guilford Press.
- Kubinger, K.D. (2003). On artificial results due to using factor analysis for dichotomous variables. *Psychology Science*, 45, 106-110.
- Kubinger, K. D. (2008). On the revival of the Rasch model-based LLTM: From constructing tests using item generating rules to measuring item administration effects. *Psychology Science Quarterly*, 50, 311-327.
- Lord, F. M. & Novick, M. R. (Eds.) (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Lucke, J. F. (2005). The  $\alpha$  and  $\omega$  of congeneric test theory: an extension of reliability and internal consistency to heterogeneous tests. *Applied Psychological Measurement*, 29, 65-81. <http://dx.doi.org/10.1177/0146621604270882>
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84-99. <http://dx.doi.org/10.1037/1082-989X.4.1.84>
- McCullagh, P., & Nelder, J.A. (1985). *Generalized Linear Models*. London: Chapman and Hall.
- McDonald, R. P., & Ahlswat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*, 27, 82-99. <http://dx.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. [http://dx.doi.org/10.1207/S15328007SEM0801\\_1](http://dx.doi.org/10.1207/S15328007SEM0801_1)
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous variable indicators. *Psychometrika*, 49, 115–132. <http://dx.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.), *Testing 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. <http://dx.doi.org/10.2307/2344614>
- Pearson, K. (1900). Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London Series A*, 195, 1–47. <http://dx.doi.org/10.1098/rsta.1900.0022>
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Satorra, A., & Bentler, P.M. (1994). Corrections to the test statistics and standard errors on covariance structure analysis. In A. von Eye & C. C. Glogg (Eds.), *Latent variable analysis* (pp. 399–419). Thousand Oaks, CA: Sage.
- 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. (2006a). The fixed-links model in combination with the polynomial function as tool for investigating choice reaction time data. *Structural Equation Modeling*, 13, 403–419. [http://dx.doi.org/10.1207/s15328007sem1303\\_4](http://dx.doi.org/10.1207/s15328007sem1303_4)
- Schweizer, K. (2006b). The fixed-links model for investigating the effects of general and specific processes on intelligence. *Methodology*, 2, 149–160. <http://dx.doi.org/10.1027/1614-2241.2.4.149>
- Schweizer, K. (2008). Investigating experimental effects within the framework of structural equation modeling: an example with effects on both error scores and reaction times. *Structural Equation Modeling*, 15, 327–345. <http://dx.doi.org/10.1080/10705510801922621>
- Schweizer, K. (2009). Fixed-links models for investigating experimental effects combined with processing strategies in repeated measures designs: a cognitive task as example. *British Journal of Mathematical and Statistical Psychology*, 62, 217–232. <http://dx.doi.org/10.1348/000711007X268558>
- Schweizer, K. (2012). A weighted version of the tau-equivalent model of measurement for items with ordered response categories. *International Journal of Statistics and Probability*, 1, 151–163. <http://dx.doi.org/10.5539/ijsp.v1n2p151>
- 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. <http://dx.doi.org/10.5539/ijsp.v2n2p67>
- Schweizer, K., & Reiß, S. (2014). The structural validity of the FPI Neuroticism Scale revisited in the framework of the generalized linear model. *Psychological Test and Assessment Modeling*, 56, 320–335.
- 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.
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modelling: multilevel, longitudinal and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC.
- Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *testing structural equation models*, (pp. 10–40). Thousand Oaks, CA: Sage.
- Torgerson, W.S. (1958). *Theory and Method of Scaling*. New York: Wiley.

## 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/3.0/>).