Psychological Test and Assessment Modeling, Volume 61, 2019 (3), 333-353

On modeling the ceiling effect observed in cognitive data in the framework of confirmatory factor analysis

Karl Schweizer[1], *Xuezhu Ren*[2] *& Florian Zeller*[3]

Abstract

The paper reports on the extension of a method of modeling the ceiling effect for preventing the impairment of model fit regarding the data type and the evaluation in a simulation study. This method uses weights reflecting the impact of the ceiling effect on variance for representing it in the confirmatory factor model. Whereas the original method applies to binomially distributed data, the extension is applicable to continuous data. Furthermore, both the original method and its extension are applied to simulated data. The results show that the modeling of the ceiling effect reduces the impact of this effect on model fit. Assuming a ceiling effect size of 100 regarding model fit for the original confirmatory factor model, the consideration of the method of modeling the ceiling effect effectuates a reduction to an effect size between 5 and 14. Weights assuming continuous data prove to be generally more efficient than other weights.

Keywords: ceiling effect, confirmatory factor analysis, model of the covariance matrix, cognitive data, constrained factor loadings

---

[1] *Correspondence concerning this article should be addressed to:* Karl Schweizer, PhD, Department of Psychology, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 6, 60323 Frankfurt a. M., Germany; K.Schweizer@psych.uni-frankfurt.de

[2] Huazhong University of Science and Technology, China

[3] Goethe University Frankfurt, Germany

The ceiling effect denotes a distributional distortion that can be found particularly in data observed by means of measures capturing easy cognitive processes. A major characteristic of this effect is that the variance of the affected data is smaller than expected. Data showing this characteristic are not suited for an investigation by means of confirmatory factor analysis. However, the reduction in variance can provide the outset for modeling this effect. If appropriately modeled, it may no longer impair model fit. A method that incorporates such modeling of the ceiling effect was introduced recently and demonstrated in an empirical example (Schweizer, 2016). This paper reports a major advancement of this method of modeling the ceiling effect by extending it to a more general data type and provides a systematic evaluation by means of a simulation study.

On the ceiling effect

In social science methodology the term ceiling effect is used for indicating that a large number of participants reaches maximum score or achieve scores close to it (Hessling, Traxel, & Schmidt, 2004; Vogt, 2005). Since in statistical investigations a distribution similar to the normal distribution is mostly expected, a distribution with the peak close to the maximum score means that there is a distortion of the data because of the limitation to the scale range. In order to prevent the ceiling effect, textbooks on test construction recommend the limitation of the range for the item difficulties and the avoidance of extremely easy items (e.g., Allen & Yen, 2001). Another way of avoiding the ceiling effect is the adaptation of the item difficulty to the participant's previous performance in considering a selection strategy like the strategy of achieving maximum information (Eggen, 1999; Thompson, 2009). Given such a selection strategy and a large item base with a very broad range of item difficulties, the observation of a ceiling effect should be a very rare event.

However, avoidance of the ceiling effect is sometimes hardly possible. It depends on the kind of mental information processing that is in the focus of the investigation since some kinds are more prone to lead to the ceiling effect than others. Especially if the processing is mainly perceptual and basically cognitive processing, as is for example in the attentional blink task (Raymond, Shapiro, & Arnell, 1992), the memory scanning task (Sternberg, 1966) or Posner's task (Posner & Mitchell, 1967), the probability of making an error is low. Such processing shows a high degree of automation and can be performed very efficiently since the mental system has been optimized to master the large amount of information that is permanently stimulating the individual. As a consequence of this special situation, such tasks include a large number of trials that are usually treated as items, and the error rate in applying such tasks is usually quite low. This implies that the probability of such data to show a ceiling effect is high.

Furthermore, the ceiling effect can be a consequence of the change of a characteristic of the population. A specific age cohort may constitute the population of interest, and data may be collected from a sample drawn from this population at different time points. While the data collected at the first testing occasion may show no ceiling effect, the investigation of the data originating from later testing occasions may reveal the ceiling effect because of developmental changes. Such an incidence of a ceiling effect due to developmental processes has been reported in longitudinal research (Wang, Zhang, McArdle, & Salthouse, 2009).

The ceiling effect impairs statistical investigations by means of confirmatory factor analysis since data serving as input are expected to follow the normal distribution. Deviations from the normal distribution have been found to lead to model misfit (Fan & Hancock, 2012; West, Finch, & Curran, 1995). Although the data distribution does not directly play a role in confirmatory analysis, there is an indirect effect since deviation from normality mostly means a reduction of the variance and covariances included in the covariance matrix that serves as input to confirmatory factor analysis (Jöreskog, 1970). It is the reduction of variance that turns the ceiling effect into a source of model misfit since the reduction of variance is a major characteristic of the ceiling effect (Hessling et al., 2004; Vogt, 2005; Uttl, 2005).

The ceiling effect as the source of model misfit

In order to get insight into how the ceiling effect influences model fit in confirmatory factor analysis, the model of the covariance matrix must be taken into consideration: in confirmatory factor analysis the pxp empirical covariance matrix **S** that shows the ceiling effect in one or a few observed variables is reproduced by means of the model of the pxp covariance matrix **Σ**:

$$\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Theta} \tag{1}$$

with **Λ** as the pxq matrix of factor loadings, **Φ** as the qxq matrix of the variances and covariances of the q latent variables and **Θ** as the pxp diagonal matrix of error variances. The p observed variables contributing to **S** are reflected by the p manifest variables of **Σ**. The first summand to the right of the equality sign represents the systematic variances and covariances and the second summand the error variances. In most applications the model assumes one systematic source giving rise to one latent variable (q=1) that systematically influences the p manifest variables reflecting the p observed variables of $\{X_1, \ldots, X_p\}$.

Parameter estimation is expected to lead to a high degree of similarity between the elements of **S** and **Σ** if the model is correct. This expectation implies that after parameter estimation the variance of each manifest variable $X^{(m)}$ of **Σ** (= $\lambda\phi\lambda + \theta$) shows a size similar to the one of the corresponding observed variable $X^{(o)} \in \{X_1, \ldots, X_p\}$ [=var($X^{(o)}$)]:

$$\mathrm{var}\left(X^{(o)}\right) \approx \lambda\varphi\lambda + \theta \tag{2}$$

where λ is the factor loading, ϕ the variance of the latent variable und θ the error variance.

However, in the case of a ceiling effect in the observed variable there is a reduction of the observed variance due to the distortion of the distribution of this variable. This distortion means skewness and is signified by the subscript C. Usually the ceiling effect means a large reduction of the observed variance as compared to the variance of another variable X from the set of observed variables $\{X_1, \ldots, X_p\}$. Both X and X_C are assumed to be due to the same source and originate from the same measurement procedure:

$$\mathrm{var}\left(X_C^{(o)}\right) << \mathrm{var}\left(X^{(o)}\right) \tag{3}$$

In parameter estimation without considering the ceiling effect explicitly it is implicitly assumed that $\mathrm{var}\left(X_C^{(o)}\right)$ is actually $\mathrm{var}\left(X^{(o)}\right)$. Parameter estimation according to Equation 2 yields estimates of the parameters λ and θ if ϕ is constrained or alternatively of θ and to lower degree of ϕ if the λs are constrained. Whereas λ and θ only apply to the parts of the model of the covariance matrix that involve the corresponding manifest variable, ϕ concerns to all manifest variables of the set $\{X_1, \ldots, X_p\}$ in the same way. At this point it is useful to distinguish between two types of models: models with free factor loadings and models with constrained factor loadings. The first type shows the structure of the congeneric model of measurement (Jöreskog, 1971). Parameter estimation in considering this model is mostly conducted in assuming that

$$\mathrm{var}\left(X_C^{(o)}\right) \approx \hat{\lambda}\varphi\hat{\lambda} + \hat{\theta} \tag{4}$$

where $\hat{\lambda}$ and $\hat{\theta}$ are estimates of λ and θ regarding $X_C^{(o)}$. Since $\mathrm{var}\left(X_C^{(o)}\right)$ is smaller than $\mathrm{var}\left(X^{(o)}\right)$, $\hat{\lambda}$ and $\hat{\theta}$ of Equation 4 are likely to be smaller than the $\hat{\lambda}$s and $\hat{\theta}$s associated with the other manifest variables. The small size of $\hat{\lambda}$ can lead to problems in reproducing the covariances of the observed variable showing the ceiling effect with the other observed variables. Furthermore, it can also happen that $\hat{\lambda}$ of the variable showing the ceiling effect is overestimated because of a large influence of the fit of the covariances with the other observed variables in parameter estimation. In this case $\hat{\theta}$ may even become negative. Therefore, there is reason for expecting model misfit if an observed variable shows the ceiling effect.

Furthermore, there is the other type of models that is characterized by the constraint of all factor loadings. Since factor loadings are assumed to represent discriminability (Lucke, 2005), constrained factor loadings mean constrained discriminability. Constrained discriminability is not that rare in assessment. It characterizes the Rasch (1960) model respectively the corresponding one-parameter model (Birnbaum, 1968) and the linear logistic test model (see Kubinger, 2009). Furthermore, the tau-equivalent model (Lord & Novick, 1968) shows constrained discriminability. A convenient way of integrating constrained factor loadings into the model of measurement is fixing all the factor loadings to the same size and to estimate the size of the variance parameter of the corresponding latent variable. Subsequent standardization adjusts the sizes of the factor loadings to the amount of variance explained by the latent variable. In this case it is mainly $\hat{\theta}$ that has to compensate for the reduced variance of the observed variable showing the ceiling effect. In contrast, the variance of the latent variable $\hat{\phi}$ that is estimated instead of the factor loadings has to fit to the variances of all observed variables and not only of $X_{\text{C}}^{(\text{o})}$:

$$\text{var}\left(X_{\text{C}}^{(\text{o})}\right) \approx \lambda\hat{\varphi}\lambda + \hat{\theta} \tag{5}$$

In this case it is very likely that $\hat{\phi}$ is too large for the observed variance of $X_{\text{C}}^{(\text{o})}$ because of the dominating influence of the other unimpaired observed variables of the set $\{X_1, \ldots, X_p\}$ in parameter estimation. Furthermore, there are the constrained factor loadings that largely can be expected to fit to the unimpaired observed variables. In the case of same-sized constraints the factor loading of the manifest variable associated with the observed variable showing the ceiling effect is likely to be too large for reproducing the covariances with the other observed variables well. As a consequence, model misfit is very likely to be observed. In this type of factor loadings the likelihood of observing a negative $\hat{\theta}$ is even larger than in free factor loadings since there is not the possibility of a compensation by adapting the sizes of the various factor loadings individually.

Modeling the ceiling effect

This section describes the method for overcoming the negative consequences of the ceiling effect for model-data fit in confirmatory factor analysis. It consists in the modeling of the effect regarding the reduction of the variance: the variance that is reduced due to the ceiling effect is related to the full variance that is the variance observable in the absence of any distributional deviation. Assume that there is the random variable Y_i (i=1, …, p) showing the reduced variance var(Y_i) > 0 that may represent a score obtained from the binary outcomes of the trials respectively items of a cognitive test. Another assumption is that there is a reference variable Y_{u} originating from the same source and measurement procedure and showing the full variance var(Y_{u}) > 0. The variances of these variables are related to each other by means of the weight w_i ($0 < w_i \leq 1$) such that

$$\text{var}\left(Y_i\right) = w_i^2 \text{var}\left(Y_{\text{u}}\right) \tag{6}$$

In order to isolate the weight for modeling the ceiling effect, it is necessary to reorder the components of Equation 6 such that

$$w_i = \sqrt{\frac{\text{var}\left(Y_i\right)}{\text{var}\left(Y_{\text{u}}\right)}} \tag{7}$$

If it is assumed that Y_i and Y_{u} are scores following a binomial distribution, this implies that Y_i and Y_{u} are obtained from binary random variables X_{ij} and $X_{\text{u}j}$ (i=1, …, p; j=1, …, r) of {0,1}. Furthermore, it is implicitly assumed that the binary random variables originate from the same cognitive task and share the same properties. In this case Equation 7 needs to be updated in the following way:

$$w_i = \sqrt{\frac{r\Pr\left(X_{i.}=1\right)\left[1\text{-}\Pr\left(X_{i.}=1\right)\right]}{r\Pr\left(X_{\text{u.}}=1\right)\left[1\text{-}\Pr\left(X_{\text{u.}}=1\right)\right]}} \tag{8}$$

where r is the number of binary outcomes contributing to each score, $\Pr(X_{i.} = 1)$ is the average probability that the trial is correctly completed, and $\Pr(X_{\text{u.}} =1)=0.5$ is the probability of a correct outcome of the binary random variables that constitute Y_{u}. Since r can be eliminated and the value of the remaining denominator is always 0.25, Equation 8 can be simplified as

$$w_i = \sqrt{\frac{\Pr\left(X_{i.}=1\right)\left[1\text{-}\Pr\left(X_{i.}=1\right)\right]}{0.25}} \tag{9}$$

The denominator is always 0.25 since in the binomial distribution the probability associated with the reference variable is always 0.5. It needs to be added that different data levels may necessitate the consideration of different denominators for different levels (Schweizer, 2016).

However, the observed data may not follow the binomial distribution. Cognitive data as the major field of application are mostly treated as continuous data. There are usually many trials with binary outcomes, and there are research results justifying the treatment of such data as continuous data (Dolan, 1994). Because of the continuous scale of such data the method of computing variances of continuous data is considered additionally for computing weights:

$$w_i = \sqrt{\frac{\text{E}\left[Y_i - \text{E}\left(Y_i\right)\right]^2}{\text{E}\left[Y_{\text{u}} - \text{E}\left(Y_{\text{u}}\right)\right]^2}} \tag{10}$$

where E( ) indicates that the parameter or mathematical expression included in parentheses is or amounts to an expected value. Weights according to Equations 9 and 10 can be computed for every observed random variable as long as the ceiling effect is not entire and an estimate of the full variance is available. In the case that there is no ceiling effect, the Equations yield the weight of one that does not change anything as part of the model of the covariance matrix (see Equation 1).

For integrating the weights obtained for the scores $Y_1, \ldots, Y_p$ into the model of the covariance matrix of Equation 1 they are assigned to the main diagonal of the pxp diagonal matrix of weights $\mathbf{W}$:

$$\mathbf{W} = \begin{bmatrix} w_1 & . & . & . & 0 \\ . & . & & & . \\ . & & w_i & & . \\ . & & & . & . \\ 0 & . & . & . & w_p \end{bmatrix} \tag{11}$$

Since the ceiling effect is a systematic effect regarding the distribution of items, i.e. it does not occur at random, the weights are not applied to the whole model of the covariance matrix. Instead they are applied to the part of the model comprising the systematic variances and covariances:

$$\mathbf{\Sigma} = \mathbf{W}\left(\mathbf{\Lambda\Phi\Lambda'}\right)\mathbf{W'} + \mathbf{\Theta} \tag{12}$$

The systematic part of the model of the covariance matrix according to Equation 12 includes three matrices: **W**, **Λ** and **Φ**. **W** is composed of fixed parameters only that are estimated in a preceding step. In the case of constrained discriminability **Λ** does not include parameters that need to be estimated. Only the parameters of **Φ** require estimation. In this case it is convenient and without further consequences to rearrange the matrices as follows:

$$\mathbf{\Sigma} = \left(\mathbf{W\Lambda}\right)\mathbf{\Phi}\left(\mathbf{W\Lambda}\right)' + \mathbf{\Theta} \tag{13}$$

This weighted model can be expected to take the reduction of variance due to the ceiling effect into consideration. As a consequence, there should be a noticeable reduction in the impairment in model-data fit due to the ceiling effect at the least, and there should be no more an impairment at the best.

In order to demonstrate the added value of this method for controlling the ceiling effect, 500 matrices with 350 rows and 5 columns of simulated data, that included one data column showing the ceiling effect, were investigated by means of confirmatory factor analysis. In this investigation the weight was not determined by means of one of the Equations but systematically varied between zero and one. Figure 1 includes the CFI, chi-square and RMSEA results achieved in this investigation as curves. Since chi-square values can become rather large, the observed values were rescaled in such a way that the largest observed value was one.

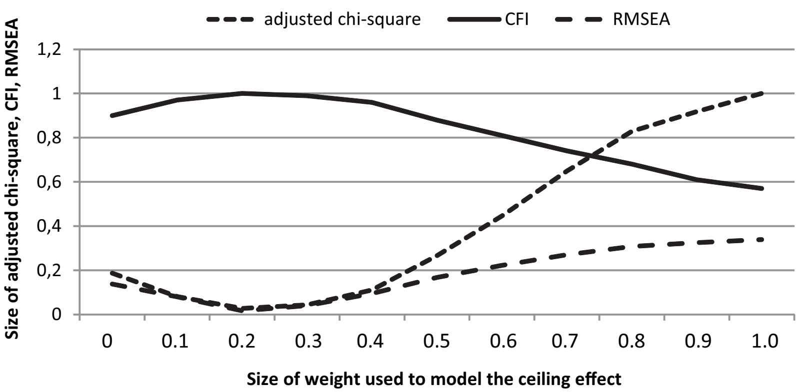

Figure 1:

Courses of adjusted chi-square, CFI and RMSEA when the weight for modeling the ceiling effect was increased from zero to one in data showing the ceiling effect in one variable

Almost zero in the horizontal axis ($w_i \sim 0.0$) meant that the assumption of an extremely strong ceiling effect was represented by the weight whereas the value of one ($w_i = 1.0$) was indicative of the situation that no ceiling effect was assumed. As is obvious from each one of the curves, starting from the assumption of no ceiling effect ($w_i = 1.0$) a decrease of w_i (from right to left) leads to an increase in model fit until the minimum (see chi-square and RMSEA) or the maximum (see CFI), resp., of the fit curves is reached. A further decrease of w_i gradually reverses the improvement in model fit.

Research goals

One major goal of the present study was to provide an evaluation of the method of modeling the ceiling effect as part of the confirmatory factor model in a simulation study. The method was also applied to an empirical example that is reported in the Appendix. Another major goal was to compare the method of computing weights under the assumption of binomially distributed data (Equation 9) and under the assumption of continuous data (Equation 10). Only confirmatory factor models with constrained factor loadings that could be expected to be especially vulnerable to the ceiling effect were considered in the simulation study.

Simulation study

The simulation study was expected to provide information on the efficiency of the methods of modelling the ceiling effect in data showing a variety of different characteristics. In this study the observed variables provided justification for assuming the same full variance for all of them with the exception of the variable selected to show the ceiling effect. Furthermore, data simulation enabled the selection of a specific value for the full variance so that it was possible to consider weights binary and continuous data.

Method

Three 60x60 covariance patterns were designed using factor loadings of .25, .35 and .45 and setting all elements of the main diagonal to 1.00. These patterns constituted the three *levels of systematic variance* design factor of the simulation study. These patterns served the generation of continuous data that were dichotomized to obtain binary data of {0, 1}. Finally subsets of the binary data were combined to arrive at five scores, as they typically could be found in cognitive research.

The ceiling effect was created by dichotomizing the continuous data in an extreme way. The selected continuous random data (i. e. data of one selected subset of columns of a data matrix) were dichotomized so that 2 percent, 4 percent or 6 percent of the values were zero and the other values were one. Since such proportions of zeros and ones meant skewness, these percentages were addressed as *levels of skewness*. The remaining data

(i. e. data of the remaining columns) were split so that 50 percent of the values were zero and the other values were one.

Furthermore, the *isolated* ceiling effect described in the previous paragraph was also one condition with respect to another design factor denoted *types of ceiling effect*. In the complementary condition of this factor one subset of columns showed a strong ceiling effect, another one no ceiling effect and the remaining columns of the data matrix in between percentages of values of zero (16 percent, 27 percent, 38 percent) while the other values were one. This condition of the types of ceiling effect factor is addressed as *embedded* ceiling effect.

Moreover, the *sample size* was varied between 250, 350 and 450. These sample sizes were selected with respect to the field of cognitive research. In this field the data were usually collected in laboratory investigations. Such investigations were usually costly in term of time and the necessary resources. Therefore in this field the sample sizes were normally quite limited.

Furthermore, all columns of all subsets of continuous data were additionally split up evenly in order to be able to estimate the model-data fit in the absence of the ceiling effect. These data were necessary for determining the degree of impairment of the model-data fit due to the ceiling effect.

The simulated data were generated by means of the procedure proposed by Jöreskog and Sörbom (2001). The generation yielded sets of 250 x 60, 350 x 60 and 450 x 60 matrices of continuous and normally distributed random data ($N(0,1)$). Each set included 500 matrices. The numbers of the columns of these matrices were dichotomized according to the described schemes to achieve matrices composed of zeros and ones. Subsequently the zeros and ones of twelve neighboring columns that were considered as subsets were added up to obtain scores. In the end there were sets of 250 x 5, 350 x 5 and 450 x 5 matrices including scores between zero and twelve.

Five hundred matrices of each type were generated. These matrices were investigated by means of different models (see paragraph below). Each result reported in the results section was achieved by investigating 500 matrices. In investigating the matrices the focus was on model fit instead of the accuracy in estimating parameters since model misfit could be expected if the model did not account for the data showing a ceiling effect.

The three confirmatory factor models for investigating these data included one latent variable and five manifest variables. As already indicated, the factor loadings were constrained (Schweizer, 2006, 2008). The number one served as constraint since the same source was assumed to contribute to each item in the same way unless there was a ceiling effect. Furthermore, the variance of the latent variable was set free for estimation. In the confirmatory factor model with weights for preventing the ceiling effect (CFMwW) the factor loading of the manifest variable expected to show the ceiling effect was set equal to the weight computed either according to Equations 9 or 10 depending on the assumption on which the investigation was focused. Furthermore, there was the confirmatory factor model with all factor loadings set equal to one that was applied to the same simulated data. Since this model could be expected to yield bad results, it is addressed as wrong confirmatory factor model (wCFM). Additionally, the confirmatory factor model with all factor loadings set equal to one was applied to the data without a ceiling effect. This model was expected to yield the best results and is referred to as cCFM.

Since cognitive scores are usually treated as continuous, parameter estimation was conducted using maximum likelihood estimation in LISREL (Jöreskog & Sörbom, 2006). The following fit statistics were used for evaluating model-data fit: chi-square, RMSEA, SRMR, CFI, TLI and GFI. The results were evaluated using the following cut-offs: RMSEA .06, SRMR .08, CFI .95, TLI .95, GFI .95 (see DiStefano, 2016).

The results for the sets of 500 matrices were summarized by computing means. The means were included in the Tables 1 to 4. Furthermore, the remaining ceiling effect coefficient (RCE) was computed on the basis of the means. It was defined as

$$RCE = \frac{\left(\chi^2_{\text{CFMwW}} - \chi^2_{\text{cCFM}}\right)}{\left(\chi^2_{\text{wCFM}} - \chi^2_{\text{cCFM}}\right)} \times 100 \tag{14}$$

The RCE varied between 0 and 100. A value of 100 meant that the impact of the ceiling effect on model fit was not reduced in any way whereas a value of 0 meant that the impact of the ceiling effect was completely removed. A small RCE coefficient was desirable.

Results

All fit results are included in tables that are structured in the same way. These tables comprise bundles of four rows of results. The first row provides the results for data showing the ceiling effect that, however, was investigated by the confirmatory factor model without weights (wCFM). The second row comprises the results for data with no ceiling effect, i. e. binary data split evenly (cCFM). Whereas the first row is expected to comprise the worst results, the second row should show the best results. The results reported in the third and fourth rows are based on the use of weights referring to the binomial distribution (b-weights) and continuous level (c-weights) and are expected to be closer to the best (cCFM) than the worst results (wCFM).

Although several fit statistics are reported, the discussion concentrates on the chi-square since in this statistic the impact of the ceiling effect and the consequences of using weights are especially obvious.

The variation of the systematic variance combined with the embedded ceiling effect. The variation of the proportion of systematic variance (small, medium, large) by means of different sizes of the factor loadings led to the fit results reported in Table 1 when the sample size was 350 and there was an embedded ceiling effect.

The chi-square statistic - inflated due to the ceiling effect (rows 1, 5 and 9) - varied between 102.79 and 513.83. The increase in systematic variance (small, medium, large) was associated by a virtually linear increase of the chi-square. In contrast, in the absence of the ceiling effect the chi-square results only varied between 8.48 and 8.72 (rows 2, 6

Table 1:
Fit Results for Data Showing the Embedded Ceiling Effect and Different Proportions of Systematic Variance (Each statistic is based on 500 matrices)

Systematic variance	Ceiling effect	Use of weights	χ^2	df	RMSEA	SRMR	CFI	NNFI	GFI
Small	Yes	No	102.79	9	.172	.150	.68	.64	.89
	No	No	8.68	9	.012	.039	1.00	1.00	.99
	Yes	b-weight	21.20	9	.058	.070	.95	.95	.97
	Yes	c-weight	15.68	9	.041	.059	.97	.97	.98
Medium	Yes	No	276.26	9	.293	.261	.67	.63	.76
	No	No	8.72	9	.012	.037	1.00	1.00	1.00
	Yes	b-weight	39.88	9	.097	.106	.95	.95	.96
	Yes	c-weight	20.74	9	.058	.074	.98	.98	.98
Large	Yes	No	513.89	9	.402	.373	.64	.59	.63
	No	No	8.48	9	.011	.033	1.00	1.00	.99
	Yes	b-weight	63.70	9	.131	.134	.95	.94	.93
	Yes	c-weight	26.14	9	.073	.081	.98	.98	.97

and 10) and were close to the expected chi-square of 9.0. The consideration of weights according to the binomial distribution (b-weights) led to chi-squares between 21.20 and 63.74 whereas weights according to the continuous level (c-weights) led to chi-squares between 15.68 and 26.14. In both cases the increase of the proportion of systematic variance was associated with an increase of the chi-square.

The investigation of the efficiency in reducing the impact of the ceiling effect by means of the Remaining Ceiling Effect coefficient (RCE) regarding the weights according to the binomial distribution (b-weights) revealed RCEs of 13.3, 11.6 and 10.9, with respect to the three levels of systematic variance. In weights according to the continuous level (c-weights) the RCEs were 7.4, 4.5 and 3.5 in corresponding order.

The variation of the proportion of systematic variance combined with the isolated ceiling effect. The fit results for the isolated ceiling effect in combination with the variation of the proportion of systematic variance (small, medium, large) and the sample size of 350 is reported in Table 2.

When the chi-square statistics were inflated by the ceiling effect, the coefficients varied between 144.61 and 631.11, and there was an association of the increase in systematic variance and the increase in chi-square. In the absence of the ceiling effect the chi-square results were close to the expected chi-square that was 9.0 and corresponded to the results of Table 1. The correspondence was due to the fact that the same continuous data were simply split in the same way. Weights according to the binomial distribution led to chi-squares between 25.96 and 78.35 and weights according to the continuous level to chi-

Table 2:
Fit Results for Data Showing the Ceiling Effect in One Variable only and Different Proportions of Systematic Variance (Each statistic is based on 500 matrices)

Systematic variance	Ceiling effect	Use of weights	χ^2	df	RMSEA	SRMR	CFI	NNFI	GFI
Small	Yes	No	144.61	9	.208	.176	.62	.58	.86
	No	No	8.68	9	.012	.039	1.00	1.00	.99
	Yes	b-weight	25.96	9	.070	.080	.94	.94	.97
	Yes	c-weight	18.74	9	.051	.067	.97	.96	.98
Medium	Yes	No	368.11	9	.339	.300	.57	.53	.70
	No	No	8.72	9	.012	.037	1.00	1.00	1.00
	Yes	b-weight	50.58	9	.114	.124	.94	.94	.94
	Yes	c-weight	26.02	9	.072	.088	.98	.98	.97
Large	Yes	No	631.67	9	.446	.407	.55	.50	.58
	No	No	8.48	9	.011	.033	1.00	1.00	.99
	Yes	b-weight	78.35	9	.148	.161	.94	.93	.92
	Yes	c-weight	32.83	9	.086	.105	.98	.98	.96

squares between 18.74 and 32.83. The increase of the proportion of systematic variance appeared to reflect the increase in chi-square.

The investigation of the efficiency in reducing the impact of the ceiling effect yielded RCEc of 12.7, 11.6 and 11.2 for weights according to the binomial distribution and of 7.4, 4.8 and 3.9 for weights according to the continuous level.

The variation of the type of ceiling effect. The results reported in the previous two subsections were to be related to each other in order to compare the types of the ceiling effect: the embedded and isolated types. The results reported in the first and second Tables revealed that there was less inflation of the chi-square due to the ceiling effect if there was one major ceiling effect combined with several minor ones (embedded type) instead of only one major ceiling effect (isolated type) in combination with other variables showing no ceiling effect. In the first case the overall mean chi-square was 297.64 and in the second case 381.46. However, there was no difference regarding the efficiency of the modeling by means of the two types of weights. The RCE results were virtually the same for both types of ceiling effect.

The variation of the sample size. The variation of the sample size (250, 350, 450) led to the fit results reported in Table 3 when the proportion of systematic variance was medium.

As is obvious from Table 3, the chi-square statistic inflated by the ceiling effect varied between 201.15 and 508.27. The increase of the sample size was associated with an increase in chi-square. If there was no ceiling effect, the chi-square results (between 8.68

Table 3:
Fit Results for Data Showing the Embedded Ceiling Effect and Different Sample Sizes (N=250, 350, 450) (Each statistic is based on 500 matrices)

Systematic size	Ceiling effect	Use of weights	χ^2	df	RMSEA	SRMR	CFI	NNFI	GFI
250	Yes	No	201.15	9	.291	.261	.66	.63	.76
	No	No	8.78	9	.012	.039	1.00	1.00	.99
	Yes	b-weight	31.13	9	.096	.107	.95	.95	.95
	Yes	c-weight	17.10	9	.055	.076	.98	.98	.97
350	Yes	No	276.26	9	.293	.261	.67	.63	.76
	No	No	8.70	9	.012	.037	1.00	1.00	1.00
	Yes	b-weight	39.88	9	.097	.106	.95	.95	.96
	Yes	c-weight	20.74	9	.058	.074	.98	.98	.98
450	Yes	No	631.67	9	.446	.407	.55	.50	.58
	No	No	8.48	9	.011	.033	1.00	1.00	.99
	Yes	b-weight	78.35	9	.148	.161	.94	.93	.92
	Yes	c-weight	32.83	9	.086	.105	.98	.98	.96

and 8.78) were quite stable and close to the expected chi-square that was 9.0. For the weights according to the binomial distribution chi-squares between 31.13 and 63.02 were observed, and for the weights according to the continuous level the results varied between 17.10 and 25.81. In both cases the increase of the sample size was reflected by an increase in chi-square.

The efficiency in reducing the impact of the ceiling effect represented by the RCE coefficient was also investigated. The results for the weights according to the binomial distribution were 11.6 percent in the sample size of 250, 11.6 percent in the sample size of 350 and 10.9 percent in the sample size of 450. The results for the weights according to the continuous level were 4.3, 4.5 and 3.4 percent in the sample sizes of 250, 350 and 450.

The variation of skewness. This subsection reports the chi-square results obtained in varying the degree of skewness characterizing the column showing the ceiling effect while the other columns were split evenly. There were three different levels (high: 6 percent zeros, very high: 4 percent zeros and extremely high: 2 percent zeros) that were realized in the sample size of 350. The fit results are reported in Table 4.

In the model without a representation of the ceiling effect the chi-square statistic varied between 298.53 and 440.21. There was an almost linear increase from the high to the extremely high degree. The reduction of the proportion of zeros in one of the five columns of a matrix from 6 to 4 percent caused an increase of the chi-square by 23 percent. The further reduction from 4 to 2 percent led to a further increase of the chi-square by 20

Table 4:
Fit Results for Data Showing the Ceiling Effect in One Variable only Characterized by Different Degrees of Skewness (Each statistic is based on 500 matrices)

Degree of skewness	Ceiling effect	Use of weights	χ^2	df	RMSEA	SRMR	CFI	NNFI	GFI
High	Yes	No	298.53	9	.304	.275	.70	.70	.74
	No	No	8.72	9	.012	.037	1.00	1.00	1.00
	Yes	b-weight	37.16	9	.092	.104	.96	.96	.96
	Yes	c-weight	19.86	9	.056	.074	.98	.98	.98
Very high	Yes	No	368.11	9	.339	.300	.57	.53	.70
	No	No	8.72	9	.012	.037	1.00	1.00	1.00
	Yes	b-weight	50.58	9	.114	.124	.94	.94	.94
	Yes	c-weight	26.02	9	.072	.088	.98	.98	.97
Extremely high	Yes	No	440.21	9	.372	.311	.49	.42	.66
	No	No	8.48	9	.011	.033	1.00	1.00	.99
	Yes	b-weight	69.70	9	.138	.148	.91	.90	.92
	Yes	c-weight	35.29	9	.091	.107	.96	.96	.96

percent. The consideration of weights according to the binomial distribution led to chi-squares between 37.16 and 69.70, and in weights according to the continuous level they varied between 19.86 and 35.29.

The RCE coefficients obtained for the weights according to the binomial distribution were 9.8, 11.6 and 14.1. In weights according to the continuous level the coefficients were 3.8, 4.8 and 6.2 in corresponding order. In both cases there was a small increase of the percentage related to the increase of skewness.

The comparison of the types of weights. Finally the two types of weights were compared with respect to the results reported in the first and second Tables. The overall chi-square for the weights according to the binomial distribution (b-weight) was 46.61 and the weights according to the normal distribution (c-weight) 23.36. Furthermore, the RCEs were computed. The overall RCE for the weights according to the binomial distribution was 11.9 and for the weights according to the continuous level 5.3. These results suggested that weights based on the assumption of a continuous data level did much better than weights according to the binomial distribution.

Concluding discussion

The reported investigation provides insight into the efficiency of the methods of modeling the ceiling effect in confirmatory factor analysis and also into the characteristics of the ceiling effect with respect to model-data fit. Both ways of computing weights prove to be very efficient in reducing the impact of the ceiling effect on model-data fit; there is a reduction of at least 80 percent of what is due to the ceiling effect in the deviation from good model fit. However, the computation of weights in assuming data showing the continuous level even excels the other way of computing weights. On average the remaining deviation from the state of no ceiling effect is only about five percent from the original 100 percent in weights for data showing the continuous level. The advantage of the weights based on the assumption of the data showing the continuous level over the other weights is probably due to the employed variance coefficient. The variance coefficient for data showing the continuous level seems to reflect the actual distribution of the data to a higher degree than the other variance coefficient. Analogically in parameter estimation with diagonal weighting the use of variances of the same type is reported to be especially efficient (Bandalos, 2014). Furthermore, there seems to be a systematic difference between the two types of weights: the weights for the continuous level mostly are smaller than other weights.

Although the results of this investigation advocate the use of weights assuming a continuous level in the modeling of the ceiling effect, the actual use of such weights is restricted. This kind of weight presupposes the availability of an estimate of the full variance. Such a variance characterizes a random variable assumed to be due to the same systematic source as the variable showing the ceiling effect and to result from the same measurement procedure. In an empirical investigation it may not be known what the full variance is, and it may be difficult to find a replacement that should be characterized by normal skewness and kurtosis. In contrast, the full variance according to the binomial distribution is always known.

The investigated method for preventing the ceiling effect to impair model fit differs from the other methods used to deal with this effect: listwise deletion, the treatment as missing data and Tobit regression (1958) that is proposed by Muthen (1989) but found to give reason for doubts (Van den Oord & Van der Ark, 1997). The proposed method minimizes the loss of information and the use of additional assumptions as compared to the other methods. Furthermore, in models with constrained factor loadings, there is the possibility to set the affected factor loading free for estimation (Schweizer, Altmeyer, Ren, & Schreiner, 2015).

Lastly it needs to be indicated that the results show that the proportion of systematic variance and the skewness of the variable showing the ceiling effect influence the impact of the ceiling effect on model-data fit. In contrast, there seems to be virtually no effect due to the sample size and the type of ceiling effect if weights are used in the investigation. Furthermore, a very interesting observation of the present study is that the impact of one variable showing the ceiling effect on model fit is stronger than the impact of several variables showing different degrees of the ceiling effect.

References

Allen, M. J., & Yen, W. M. (2001). *Introduction to measurement theory* (2nd ed.). Monterey CA: Brooks/Cole.

Bandalos, D.L. (2014). Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation. *Structural Equation Modeling, 21*, 102-116. doi: 10.1080/10705511.2014.859510

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.

DiStefano, C. (2016). Examining fit with structural equation models. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction. Standards and recent advances* (pp. 166–193). Göttingen, Berne: Hogrefe.

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5, and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology, 47*, 309–326. doi: 10.1111/j.2044-8317.1994. tb01039.x

Eggen, T. J. H. M. (1999). Item selection in adaptive testing with the sequential probability ratio test. *Applied Psychological Measurement, 23*, 249-261. doi: 10.1177/014662199 22031365

Fan, W., & Hancock, G.R. (2012). Robust means modeling: an alternative for hypothesis testing of independent means under variance heterogeneity and nonnormality. *Journal of Educational and Behavioral Statistics, 37*, 137- 156. doi: 10.3102/1076998610396897

Hessling, R. M., Traxel, N. M., & Schmidt, T. J. (2004). Ceiling effect. In Lewis-Beck, M. S., A. Bryman, & Liao, T. F. (Eds.), *The SAGE encyclopedia of social science research methods* (p. 107). Thousand Oaks: Sage.

Jöreskog, K. G. (1970). A general method for analysis of covariance structure. *Biometrika, 57*, 239-257. doi: 10.1093/biomet/57.2.239

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

Kubinger, K. D. (2009). Applications of the linear logistic test model in psychometric research. *Educational and Psychological Measurement, 69*, 232-244. doi: 10.1177/001316 4408322021

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MASS: Addison-Wesley.

Lucke, J. F. (2005). The α and ω of congeneric test theory: an extension of reliability and internal consistency to heterogeneous tests. *Applied Psychological Measurement, 29*, 65–81. doi: 10.1177/0146621604270882

Muthen, B. (1989). TOBIT factor analysis. *British Journal of Mathematical and Statistical Psychology, 42*, 241–250. doi 10.1111/j.2044-8317.1989.tb00913.x

Posner M. I., & Mitchell, R. F. (1967). Chronometric analysis of classification. *Psychological Rewiew, 74*, 392-409. doi:10.1037/h0024913

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18*, 849-860. doi: 10.1037/0096-1523.18.3.849

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

Schweizer, K. (1998). Complexity of information processing and the speed-ability relationship. *The Journal of General Psychology, 125*, 89-102. https://doi.org/10.1080/00221309809595578

Schweizer, K. (2006). The fixed-links model for investigating the effects of general and specific processes on intelligence. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 2,* 149-160. doi: 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. doi: 10.1080/10705510801922621

Schweizer, K. (2007). Investigating the relationship of working memory tasks and fluid intelligence tests by means of the fixed-links model in considering the impurity problem. *Intelligence, 35*, 591-604. doi: 10.1016/j.intell.2006.11.004

Schweizer, K. (2011). Scaling variances of latent variables by standardizing loadings: applications to working memory and the position effect. *Multivariate Behavioral Research, 46*, 938-955. doi: 10.1080/00273171.2011.625312

Schweizer, K. (2016). A confirmatory factor model for the investigation of cognitive data showing a ceiling effect: an example. In L. A. van der Ark, D. M. Bolt, W.-C. Wang, J. Douglas & M. Wiberg (Eds.), *Quantitative Psychology Research*. Heidelberg: Springer.

Schweizer, K., Altmeyer, M., Ren, X., & Schreiner, M. (2015). Models for the detection of deviations from the expected processing strategy in completing the items of cognitive measures. *Multivariate Behavioral Research, 50*, 544-554. doi: 10.1080/00273171.2015.1032397

Schweizer, K., Troche, S., & DiStefano, C. (2019). Scaling the variance of a latent variable while assuring constancy of the model. *Frontiers in Psychology* (Section Quantitative Psychology and Measurement), *10*, ArtID 887. doi: 10.3389/fpsyg.2019.00887

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*, 652–654. doi: 10.1126/science.153.3736.652

Thompson, N. (2009). Item selection in computerized classification testing. *Educational and Psychological Measurement, 69*,778-793. doi: 10.1177/0013164408324460

Tobit, J. (1958). Estimation of relationships for limited dependent variables. *Econometria, 26*, 24-36. doi: 10.2307/1907382

Uttl, B. (2005). Measurements of individual differences: lessons from memory assessment in research and clinical practice. *Psychological Science, 16*, 460–467. doi: 10.1111/j.0956-7976.2005.01557.x

Van den Oord, E. J. C. G., & Van der Ark, L. A. (1997). A note on the use of Tobit approach for tests scores with floor or ceiling effects. *British Journal of Mathematical and Statistical Psychology, 50*, 351–364. doi: 10.1111/j.2044-8317.1997.tb01150.x

Vogt, W. P. (2005). Dictionary of statistics & methodology: a nontechnical guide for the social sciences (3rd ed.). Thousand Oaks: Sage.

Wang, L., Zhang, Z., McArdle, J. J., & Salthouse, T. A. (2009). Investigating ceiling effects in longitudinal data analysis. *Multivariate Behavioral Research, 43*, 476-496. doi: 10.1080%2F00273170802285941

Wechsler, D. (2006). In M. von Aster (Ed.), *Wechsler Intelligenztest für Erwachsene (WIE). Adaption of the WAIS-III* (Wechsler Intelligence Test For Adults). Frankfurt: Hartcourt Test Services.

West, S.G., Finch, J.F., & Curran, P.J. (1995). Structural equation models with nonnormal variables: problems and remedies. In R. H. Hoyle. (ed.), *Structural equation modeling: concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA: Sage.

Appendix: Empirical example

A confirmatory factor model of measurement adapted to the representation of working memory served as example. Other models of measurement were additionally considered for demonstrating the consequences of the various assumptions represented by these models. The manifest variables of this model had to reflect the three different treatment levels of the Backword Counting scale. The scores of one of these levels were likely to show the ceiling effect since only as few as two numbers had to be recount whereas the scores requiring more complex processing (Schweizer, 1998) of the other levels were not. The modeling of the ceiling effect was expected to reverse the model misfit due to this effect.

Method

The Backward Counting scale that was an adaptation of Wechsler's (2006) "digit span" subtest was completed by 211 university students. The scale required the recounting of 2, 4 or 6 numbers that were auditory presented in the reversed order. There were three treatment levels due to the three sets of numbers.

The confirmatory factor analysis model of measurement included one latent variable for representing working memory and three manifest variables. The accuracy scores observed in the recounting of 2, 4 or 6 numbers served as manifest variables. This model of measurement was realized as congeneric model (Jöreskog, 1971), as tau-equivalent model (Lord & Novick, 1968) and as fixed-links model considering measurement impurity (Schweizer, 2007; Schweizer, Altmeyer, Ren, & Schreiner, 2015).

Since cognitive scores were usually treated as continuous, parameter estimation was conducted using maximum likelihood estimation in LISREL (Jöreskog & Sörbom, 2006). The following fit statistics were used for evaluating model-data fit: chi-square, RMSEA, SRMR, CFI, TLI and GFI. The results were evaluated in using the following cut-offs: RMSEA .06, SRMR .08, CFI .95, TLI .95, GFI .95 (see DiStefano, 2016).

Results

The mean accuracies as probabilities to respond correctly varied between .99, .94 and .58 for the first to third treatment levels. The corresponding standard deviations were 0.04, 0.10 and 0.26. The probability of the first level was very close to the upper limit that was 1.00 and of the second level in the near range of this limit. Especially the first one was considered as being due to the ceiling effect. Note. So far there was no proposal of a boundary for differentiating between the ceiling effect and no ceiling effect.

The congeneric model was realized within the framework illustrated by Figure 2A.

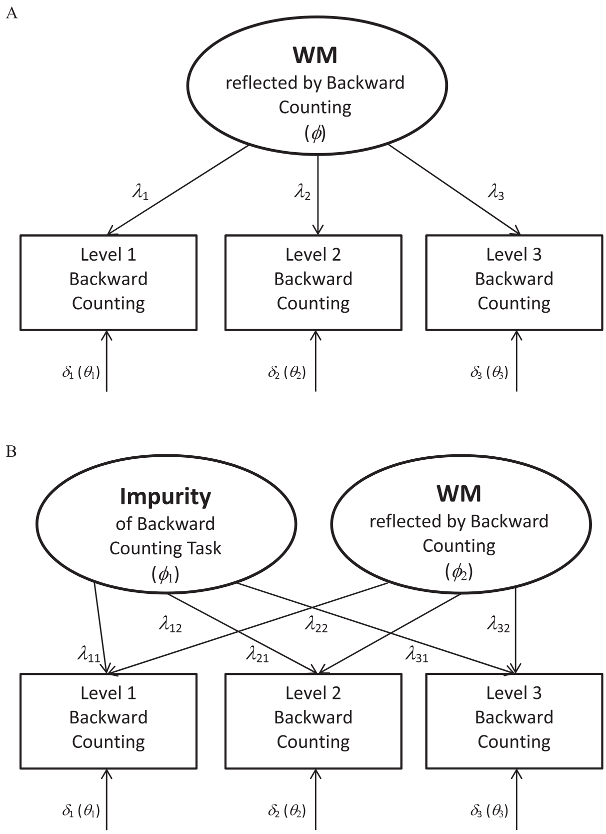

Figure 2:
Illustrations of the general structure specified as congeneric or tau-equivalent models (A) and as fixed-links model considering measurement impurity (B)

Two of the seven parameters needed to be constrained to have one degree of freedom. Most options for constraining led to model misfit. The only good fitting model was achieved for setting $\phi = 1$ and $\delta_3 = 0$. The model specified this way yielded the following fit results: χ^2 (1) = 0.64, RMSEA = .000, SRMR = .021, CFI = 1.00, TLI = 1.04, and GFI = 1.00. Although the model fit was good, the outcome was not acceptable because of $\lambda_1 = 0.00$ and $\delta_1 = 1.00$ indicating that the 3-indicator model was actually a 2-indicator model. This outcome contradicted the aim to construct a model of measurement that links the working memory latent variable to the three indicators. Therefore, no further consideration was given to this model.

In order to have the tau-equivalent model, the factor loadings (λ_1, λ_2, λ_3) were set to equal size. The following fit results were achieved for this model: χ^2 (2) = 20.46, RMSEA = .210, SRMR = .123, CFI = .36, TLI = .35 and GFI = .94. These statistics indicated model misfit whereas the factor loadings and error variances were within the acceptable ranges. Because of the large size of the chi-square and the small number of degrees of freedom this model was also not given further consideration.

For investigating the fixed-links model considering measurement impurity a second latent variable was considered, and the following constraints were selected: $\lambda_{11} = \lambda_{21} = \lambda_{31} = 1$, $\lambda_{12} = 1$, $\lambda_{22} = 2$, $\lambda_{32} = 3$, which is illustrated in Figure 2B. The model with these constraints yielded the following fit results: χ^2 (1) = 3.04, RMSEA = .099, SRMR = .043, CFI = .92, TLI = .76, and GFI = .99. These results indicated in some statistics a good and in some others a bad model fit.

Modeling the ceiling effect in this example was especially difficult since the tree levels (recounting 2, 4 or 6 numbers) could be assumed to relate to different full variances instead of only one, and these variances were not known. Because of the unknown variances the modeling variant assuming the binomial distribution was selected (see Equation 8). Because of the inappropriateness of using the same full variance for each one of the three levels (for more information on this problem see Schweizer, 2016), it was necessary to select a specific probability of a correct response for each one of these levels that could provide the core for the computation of the full variance. We assumed that the probability of a correct response leading to the full variance in the third level was .5 and that there should be equal distances between the probabilities leading to full variances in the other levels. Based on this assumption the missing probabilities for the first and second levels were .833 and .666 [p_1=.833 (=.166+.166+.5), p_2=.666 (=.166+.5), p_3=.5]. Using these probabilities for computing the denominator of Equation 8 and the observed probabilities to respond correctly reported in the beginning of the results section for the numerator yielded 0.267 as weight for the first level and .504 for the second level. The third weight was close to 1.0 and, therefore, omitted.

Including these weights into the fixed-links model considering measurement impurity led to a good model fit (χ^2 (1) = 1.76, RMSEA = .060, SRMR = .039, CFI = .97, TLI = .91, and GFI = .99). However, the estimates of the variance parameters of the matrix of the variances and covariances of latent variables did no more reach the level of significance according to the asymptotic t statistic ($t_1 = 0.34$, ns; $t_2 = 0.70$, ns). The sizes of the scaled variances in assuming average factor loadings of .3 (Schweizer, 2011; Schweizer, Troche, & DiStefano 2019) were .90 and 2.86 for the impurity and working memory latent variables. After the elimination of the latent variable showing the smaller variance and representing impurity, the model fit was good (χ^2 (2) = 2.13, RMSEA = .018, SRMR = .037, CFI = .99, TLI = .99, and GFI = .99), and the estimate of the remaining parameter reached the level of significance ($t_{(2)}$ = 4.95). Furthermore, all completely standardized factor loadings were larger than zero (λ_1 = .09, λ_2 = .36, λ_{32} = .99), and all error variances were smaller than one and larger than zero (θ_1 = .99, θ_2 = .87, θ_3 = .02).

Finally it turned out that the final model was even the best model according to AIC: $AIC_{fixed\text{-}links\ model\ considering\ measurement\ impurity}$ = 10.13, $AIC_{congeneric\_model}$ = 10.64, $AIC_{tau\text{-}equivalent\_model}$ = 14.40 (with weights included).

Discussion

The use of weights enabled the construction of a model of measurement that included all three indicators for working memory measured by the Backward Counting scale. The estimates of the standardized factor loadings reflected the different contributions of the indicators to the representation of working memory, as they were obvious from the descriptions of the treatment levels of the Backward Counting scale. In this empirical example the difficulty in computing weights that eventually might have to be overcome was well illustrated.