

A Weighted Version of the Tau-equivalent Model of Measurement for Items with Ordered Response Categories

Karl Schweizer¹

¹ Department of Psychology, Goethe University Frankfurt, Frankfurt a. M., Germany

Correspondence: Karl Schweizer, Department of Psychology, Goethe University Frankfurt, 60054 Frankfurt a. M., Germany. Tel: 49-69-7982-2081. E-mail: K.Schweizer@psych.uni-frankfurt.de

Received: August 29, 2012 Accepted: September 11, 2012 Online Published: October 17, 2012

doi:10.5539/ijsp.v1n2p151 URL: <http://dx.doi.org/10.5539/ijsp.v1n2p151>

Abstract

A weighted version of the tau-equivalent model of measurement is proposed for overcoming a shortcoming of the original and essential versions of this model resulting from the assumption of equal true scores. Because of this assumption the original version cannot be expected to do well in items with considerably varying variances. The proposed version includes weights reflecting the easiness of items. In a small simulation study the original and weighted versions of the model were applied to items characterized by variances of either equal or varying sizes for demonstrating the usefulness of the weighted version. As expected, only the weighted version of the tau-equivalent model did well in all item sets. Furthermore, the results obtained by the weighted version in combination with maximum likelihood estimation appeared to be closer to the expectations than the results obtained by asymptotic distribution-free estimation and robust estimation.

Keywords: confirmatory factor analysis, congeneric model, tau-equivalent model, binary data

1. Introduction

The tau-equivalent model of measurement was proposed in combination with the model of parallel measurements in the context of true-score theory (Lord & Novick, 1968, p. 47-50). Both models describe the observed score as the composite of true and error scores. However, whereas parallel measurements are assumed to include corresponding true and error scores, in the tau-equivalent model correspondence is restricted to the true scores of measurements. The tau-equivalent model is primarily known as the possible theoretical basis of Cronbach's coefficient alpha that is the most commonly used measure of internal consistency (Raykov, 1997a; 1997b; 2001). In this paper a weighted version of this model is proposed in order to increase its usefulness for other applications.

Although the tau-equivalent model of measurement was considered for confirmatory factor analysis (Alwin & Jackson, 1980), it has not developed into the standard model of measurement of this procedure. There is another, less restrictive model which has gained this status: the congeneric model (Jöreskog, 1971). This model even omits the assumption of equal true scores. As a consequence, it displays a considerable capability in adapting to the peculiarities of data. The downside of the capability is that it is not especially sensitive for method effects and the presence of an internal structure. Both method effects and internal structure can mean partial overlaps of underlying dimensions which may be difficult to detect in considering the congeneric model and correlated errors (Reddy, 1992).

A major reason for the neglect of the tau-equivalent model is its insufficient performance in sets of items that considerably differ from each other according to their observed variances since this model implicitly assumes that the true variances of all items are the same. So the true variance of the item showing the largest observed variance cannot surmount the true variance of the item with the smallest observed variance that may be quite small. In many cases the variability of observed variances can be suspected to be due to the skewness respectively nonnormality of items that is considered as a severe problem in confirmatory factor analysis (Curran, West, & Finch, 1997; Fan & Hancock, 2012). The skewness can frequently be ascribed to popular response formats that can be found in many paper-and-pencil measures. The rather limited number of ordered response categories of such measures often leads to ceiling or floor effects which become apparent as skewness. The weighted version of the tau-equivalent model can be instrumental in overcoming this shortcoming and by the way increase the range of possible applications.

Furthermore, an improved version of the tau-equivalent model can prove to be useful in the context of recent

developments in the representation and investigation of measures. Starting from the congeneric model of measurement the switch to the bifactor model that can presently be observed (e. g., Chen, West, & Sousa, 2006; Mulaik & Quartelli, 1997; Reise, Morizot, & Hays, 2007) is an important step since it acknowledges that many measures stimulate method effects in addition to the ability or trait of interest. This model proved to be especially useful in the investigation of the item wording effect (DiStefano, & Motl, 2006; Vautier, Raufaste, & Cariou, 2003). However, the bifactor model is insufficient if there is a considerable overlap of the method effect and ability or trait of interest. The position effect is an example of such an overlap since all the items are affected and even worse there is an increasing effect from the first to last items (Hartig, Hölzel, & Moosbrugger, 2007; Schweizer, 2012). In this situation the constraint of the factor loadings associated with the true scores can be very useful in separating the components due to the method effect and due to the ability or trait of interest.

2. The Problems of the Tau-equivalent Model of Measurement and Its Modification

2.1 The Tau-equivalent and Congeneric Models of Measurement

This section provides descriptions of the relevant models of measurement that are the original and essentially tau-equivalent and congeneric models. Following Bollen (1989, p. 18) a model of measurement is considered as a testable set of equations relating manifest to latent variables. For ease in communication the notation of the paper by Graham (2006) that concentrates on one of these equations only is adapted. The equation that characterizes the original tau-equivalent model of measurement relates the observed score of the i th ($i = 1, \dots, I$) individual responding to the j th ($j = 1, \dots, N$) item X_{ij} to the latent true score characterizing this individual with respect to the ability or trait of interest τ_i and an error score ε_{ij} such that

$$X_{ij} = \tau_i + \varepsilon_{ij} \quad (1)$$

where the true and error scores are assumed to be independent of each other. This model implies that the observed variance of the j th item $\sigma_{X_{ij}}$ is composed of the true variance σ_τ and the error variance $\sigma_{\varepsilon_{ij}}$ such that

$$\sigma_{X_{ij}} = \sigma_\tau + \sigma_{\varepsilon_{ij}} \quad (2)$$

The size of $\sigma_{\varepsilon_{ij}}$ is item-specific whereas σ_τ is the same for all the items of the item set.

The essentially tau-equivalent model of measurement assumes that there is an additional score c_j reflecting the easiness of the j th item. So this version of the tau-equivalent model differs from the original version given by Equation (1) such that

$$X_{ij} = (c_j + \tau_i) + \varepsilon_{ij} \quad (3)$$

An interesting property of this model is that the true variance is not modified because of the consideration of the easiness of the item since c_j is not a source of individual differences. So there is correspondence between the original and essentially tau-equivalent models concerning the variance; i.e. Equation (2) also applies to the essentially tau-equivalent model of measurement.

The congeneric model (Jöreskog, 1971) additionally assumes that the true score changes as a function of characteristics of the item. In the representation of the j th item according to Graham (2006) there is additionally the item-specific function $g_j(\cdot)$ so that the following model is achieved:

$$X_{ij} = [c_j + g_j(\tau_i)] + \varepsilon_{ij} \quad (4)$$

Since function $g_j(\cdot)$ transforms the true score in a systematic way, a modified representation of the variance $\sigma_{X_{ij}}$ is necessary. It is given by

$$\sigma_{X_{ij}} = \sigma_{g_j(\tau_i)} + \sigma_{\varepsilon_{ij}} \quad (5)$$

where $\sigma_{g_j(\tau_i)}$ represents the item-specific true-variance. The consequence of modifying the true component by means of function $g_j(\cdot)$ is that the true variance is different for different items. Consequently, if the functions $g_j(\cdot)$ and $g_k(\cdot)$ of the j th and k th items differ from each other, there is no correspondence of the corresponding true variances of these items although they originate from the same item set and are expected to depend on the same underlying source:

$$\sigma_{g_j(\tau)} \neq \sigma_{g_k(\tau)} \quad (6)$$

and the difference is not predictable by a general function since $g_j(\cdot)$ and $g_k(\cdot)$ are item-specific. The steps from Equation (1) to Equation (4) reveal the hierarchical relationships among the models (Graham, 2006).

2.2 The Disadvantageous Consequences of the Constancy of True Variance

Apparently, there is a major difference between the versions of the tau-equivalent model of measurement on one hand and the congeneric one on the other hand: the true variance is either assumed to be constant or to differ depending on the relatedness of the individual item and the ability or trait. This difference is especially important when applying these models of measurement as part of confirmatory factor analysis.

The model of the $N \times N$ covariance matrix of confirmatory factor analysis Σ is composed of a true part and an error part (Brown, 2006, p. 19). The true part includes the $N \times M$ matrix of factor loadings Λ and the $M \times M$ covariance matrix of factors Ψ and the error part the $N \times N$ diagonal matrix of error variances Θ . The model is defined as

$$\Sigma = \Lambda\Psi\Lambda' + \Theta \quad (7)$$

The specific perspective of the j th item ($j=1, \dots, N$) that is taken in considering and comparing the variances of individual items provides the following equation:

$$\sigma_j = \lambda_j\psi\lambda_j + \theta_j \quad (8)$$

where σ_j corresponds to $\sigma_{X_{ij}}$ of either Equations (2) or Equation (5). Equation (8) reveals that the true variance is represented by the product of the factor loading and variance of the factor.

In the congeneric model of measurement the differences between the observed variances of different items are reflected by the factor loadings (λ_j) and the error variances (θ_j). In contrast, in the versions of the tau-equivalent model the factor loadings are constant so that differences between the observed variances of items can only be ascribed to differences between error variances (Alwin & Jackson, 1980). An important implication of this restriction is that the true variance must always be smaller (or equal) than the smallest observed variance:

$$\min(\sigma_1, \dots, \sigma_N) \geq \lambda_1\psi\lambda_1 = \dots = \lambda_N\psi\lambda_N \text{ respectively } \min(\sigma_{X_{i1}}, \dots, \sigma_{X_{iN}}) \geq \sigma_\tau = \dots = \sigma_\tau \quad (9)$$

It is obvious from this equation that in observed variances showing a high variability the true variance can only be small whereas some error variances must be large. So the tau-equivalent model can only be expected to do well in items with equal (or rather similar) variances since in this case

$$\begin{aligned} \sigma_1 = \dots = \sigma_N = \min(\sigma_1, \dots, \sigma_N) &\geq \lambda_1\psi\lambda_1 = \dots = \lambda_N\psi\lambda_N \text{ respectively} \\ \sigma_{X_{i1}} = \dots = \sigma_{X_{iN}} \min(\sigma_{X_{i1}}, \dots, \sigma_{X_{iN}}) &\geq \sigma_\tau = \dots = \sigma_\tau \end{aligned} \quad (10)$$

Given the described assumption this equation demonstrates that the restriction to the true variances of the items does not mean a disadvantage. However, since this situation is rare in real data, the congeneric model is more likely to lead to a good model fit in 'normal' applications than the versions of the tau-equivalent model.

The constancy of the true variance that characterizes the versions of the tau-equivalent model is a disadvantage and limits the applicability. This property is especially disadvantageous in items with skewed distributions since such items usually show smaller variances than items lacking skewness, as it is obvious for example from the simulation study by Curran et al. (1997). In empirical data skewness is mainly due to one of two reasons: it can be due to the response format or to response bias. For example, the acquiescence tendency is a kind of response bias that is likely to lead to skewness since it can shift the mean of the response distribution close to one of the limitations of the range of possible responses. More importantly, there is the response format. Small numbers of ordered response categories are quite popular. Between four and seven categories are recommended (Lozano, Garcia-Cueto, & Muniz, 2008). Such numbers of categories are prone to yield skewness because a deviation from what is considered as the mean of the range is likely to lead to a ceiling or floor effect resulting in a reduction of the observed variance. This consequence is especially likely in items with two ordered response categories.

2.3 The Weighted Version of the Tau-equivalent Model of Measurement

In order to overcome the restriction to the applicability of the tau-equivalent model, a new version of the model is proposed. It is addressed as *weighted version* and is especially proposed for data obtained by items that are characterized by a small number of ordered response categories and test scores with a limited range. The equation characterizing the weighted version is given by

$$X_{ij} = [c_j + g(c_j)\tau_i] + \varepsilon_{ij} \quad (11)$$

where X_{ij} represents the observed score of the i th ($i = 1, \dots, I$) individual with respect to the j th ($j = 1, \dots, N$) item, c_j the easiness of the j th item, τ_i the latent true score of this individual with respect to the ability or trait of interest, $g(\cdot)$ the function reflecting the easiness of the j th item and ε_{ij} the error score. The function can be defined differently with respect to different response formats, and a proposal is made in the following section. Based on Equation (11) the variance of the j th item is given by

$$\sigma_{X_{ij}} = g(c_j)^2 \sigma_\tau + \sigma_{\varepsilon_{ij}} \quad (12)$$

The Equations (2) and (12) only differ according to the square of the function $g(\cdot)$ serving as multiplier to the true variance.

The weighted version of the tau-equivalent model and the congeneric model differ in two major points: both points are concerning the true score. First, in the weighted version the true-score is left unchanged whereas in the congeneric model it is changed in an item-specific way. Second, whereas in the weighted version skewness respectively the related change in variance is represented, there is no such representation in the congeneric model. The differences in the sizes of loadings obtained by the congeneric model can originate from various effects.

In further elaborating on the difference between the models it needs to be highlighted that the computer implementations of the true scores usually differ. In the congeneric model each factor loading λ_j that reflects the size of a true score is a parameter that is estimated whereas in the weighted version of the tau-equivalent model the factor loadings are constrained to correspond to the outcome of the general function serving as multiplier to the true score:

$$\lambda_j = g(c_j) \quad (13)$$

This equation is a reflection of c_j , and it is obvious that the factor loading changes its size as a function of the easiness of the item.

2.4 The Adaptation to the Response Format

Function $g(\cdot)$ can be considered as a link function in the sense of the generalized linear model since it establishes a link between a latent variable following the Normal distribution and a manifest variable characterized by another distribution (Nelder & Wedderburn, 1972; McCullagh & Nelder, 1985). Various specific distributions can be considered as the other distribution. Skewed distributions fall within the scope of other distributions. Furthermore, the function can be defined differently in order to apply to different response formats.

In this section $g(\cdot)$ is considered with respect to the distribution resulting from a response format including two ordered response categories and scores obtained by summing up items characterized by two ordered response categories. Since in the case of one item with two ordered response categories the variance and skewness of such an item and the probability of a correct response are closely linked, the probability of a correct response to the j th ($j = 1, \dots, N$) item π_j and the corresponding variance $\sigma_{X_{ij}}$ provide the outset. The variance is given by

$$\sigma_{X_{ij}} = \pi_j(1 - \pi_j) \quad (14)$$

In the case that score X_k ($k = 1, \dots, K$) is obtained by summing up p items showing the same π_j the definition of the variance is provided by

$$\sigma_{X_k} = p\pi_k(1 - \pi_k) \quad (15)$$

where $\pi_k (= \pi_j)$ is the probability of a correct response. Equation 15 is the generalized version of Equation 14 since setting p ($1 \leq p \leq N$) equal to one transforms the latter equation into the former one.

According to Equation 11 the constant c_k ($k = 1, \dots, K$) is the argument of $g(\cdot)$. Now c_k is defined as follows:

$$c_k = p\pi_k \quad (16)$$

with respect to the binomial distribution of Equation 15. Furthermore, $g(\cdot)$ can be defined as a real-valued function such that

$$g(c_k) = \sqrt{p\pi_k(1 - \pi_k)} \quad (17)$$

where $0 \leq \pi_k \leq 1$ describes the domain of arguments. This function serving as weight respectively multiplier reflects differences in variance due to skewness. The definition assures that $\pi_k=0.5$ leads to the largest value and $\pi_k=0$ and also $\pi_k=1$ to the smallest one. This function mimics the skewness due to the small number of response categories if π deviates from 0.5.

3. A Demonstration

In the following sections it is demonstrated that the original version of the tau-equivalent model of measurement does not perform well in skewed data whereas the weighted version does. Skewed data are considered because skewness is likely to result from the described limitation of the number of response categories and is normally associated with a decrease of the size of the variance. Furthermore, a robust estimation method is applied additionally for considering an alternative approach to investigating skewed data. Moreover, non-skewed data are included in the investigation of the models for evaluating performance under ideal conditions. It is a small simulation study only since the outcomes should be very obvious.

The investigation focuses on model fit and the sizes of the standardized factor loadings since both these outcomes can show whether there is dependency on varying variances resulting from different degrees of skewness. Such dependency should be obvious in the degree of good model fit in the first place because varying variances can mean a small true variance in combination with a large error variance. Furthermore, it should be apparent from the comparison of standardized and non-standardized factor loadings. Whereas the non-standardized factor loadings are constant by definition, the process of standardization can lead to varying sizes because of varying denominators in standardization.

Data that were continuous and binary were constructed for this study. Skewness resulting from the limitation of the number of response categories could be expected to be especially strong in data designed according to items with two ordered response categories only. For the purpose of this study skewness was defined as the deviation of the proportion of endorsements of the two ordered categories from equality. The binary data were subdivided into two types: skewed and non-skewed data. Differences between the performances of the versions of the tau-equivalent model of measurement were expected in skewed data but not in non-skewed or continuous data.

3.1 Data Generation

Thirty 500×10 matrices of continuous random data with columns following the normal distribution were generated on the basis of a one-factor model. Sets of 10 matrices were constructed such that the standardized factor loadings included in the matrices of factor loadings Λ corresponded to .5, .4 and .3:

$$\Lambda = \begin{bmatrix} .5 \\ .5 \\ .5 \\ .5 \\ .5 \\ .5 \\ .5 \\ .5 \\ .5 \\ .5 \end{bmatrix} \quad \Lambda = \begin{bmatrix} .4 \\ .4 \\ .4 \\ .4 \\ .4 \\ .4 \\ .4 \\ .4 \\ .4 \\ .4 \end{bmatrix} \quad \Lambda = \begin{bmatrix} .3 \\ .3 \\ .3 \\ .3 \\ .3 \\ .3 \\ .3 \\ .3 \\ .3 \\ .3 \end{bmatrix}$$

Since these three sets differed according to the salience of an assumed underlying source, they are addressed as salience sets.

Next the numbers included in the columns of the matrices were transformed from continuous into dichotomous for simulating the outcomes of two ordered categories. Thirty matrices of such data with a symmetric distribution were obtained by means of median split. After having the numbers of the columns ordered according to their sizes those below the median were transformed into zeros and the others into ones. These matrices are labeled as *non-skewed* in the following sections.

The achievement of data simulating the outcomes of items with two ordered categories and showing a skewed distribution required splits in considering specific percentages. In each matrix the columns were transformed in the following way: the numbers were ordered according to size and subsequently subdivided into two groups according to a specific percentage. The smaller numbers up to this percentage were transformed into zeros and the larger numbers into ones. The percentages selected for the first to tenth columns were 10, 20, 30, 40, 50, 50, 60, 70, 80, and 90 in corresponding order. So each column of a matrix was modified according to another percentage with the exception of the fifth and sixth columns. The matrices achieved this way are labeled as *skewed*.

Because of the way of generation special relationships between the types of matrices could be assumed. Consequently, virtually the same results should be achieved in skewed and non-skewed binary data by means of com-

binations of model and estimation method that are insensitive for the described type of nonnormality. SPSS was used for generating the continuous random data.

3.2 Statistical Investigation

The tau-equivalent model of measurement (Lord & Novick, 1968, p. 50) demanded the constraint of the factor loadings. All the factor loadings had to be linked to each other such that their sizes corresponded. Technically this could be easily achieved by setting all the factor loadings equal to one. However, this provision prevented the estimation of the average size of factor loadings. Setting free the variance of the latent variable compensated for the constraint of the factor loadings, as it is obvious from Equation (8): the true variance was the product of the factor loading and the variance of the factor ($\lambda\psi\lambda$); one of them needed to be fixed in order to be able to estimate the other one. In order to avoid arbitrariness in the estimation of the variance of the latent variable, scaling is recommended in a subsequent step (Schweizer, 2011). The weighted version was realized in very much the same way as the tau-equivalent model. It was just necessary to consider weights according to Equation (16) additionally.

The confirmatory factor model for the investigation of the simulated data included one latent variable since only one underlying source was assumed, and ten manifest variables respectively items. The error scores of the items were not allowed to correlate with each other. Confirmatory factor analyses according to this model were performed by means of LISREL (Jöreskog & Sörbom, 2006) with ML estimation on the basis of the covariance matrices. Furthermore, the asymptotic distribution-free estimation method suggested by Browne (1984) was applied for investigating the skewed data since this method was recommended for large data sets and also the robust estimation method proposed by Satorra and Bentler (1994).

The evaluation of model fit was conducted by means of CFI, RMSEA, and SRMR that were recommended by Kline (2005). The chi-squares which were also recommended were omitted since they proved not to be very useful in large samples. Furthermore, TLI was considered. Hu and Bentler (1999) suggested the following cut-off values for these fit statistics: RMSEA < .05, SRMR < .08, CFI > .95 and TLI > .95. Since it was necessary to perform comparisons between models which were not nested, CFI which has recently become popular for this purpose was of special interest. According to Cheung and Rensvold (2002) a CFI difference of .01 or larger indicated a substantial difference at the 5% level.

The effect of skewness on standardized factor loadings was investigated by means of the t statistic and a non-parametric test (Median Test). For this investigation the factor loadings were assigned to two groups: the group associated with the first, second, ninth and tenth columns (extreme group) and the group composed of the factor loadings of the fourth, fifth, sixth and seventh columns (moderate group). Consequently, the first group included the items resulting from a split into rather disproportional numbers of zeros and ones in skewed data whereas the numbers of zeros and ones of the other group did not differ considerably.

3.3 Evaluation of the Original Tau-equivalent Model with Respect to Model Fit

Fit results obtained by means of the original tau-equivalent model are reported separately for each type of data and each salience set. These CFI, TLI, RMSEA and SRMR results are provided in separate parts of Table 1.

Each part includes one row for the results achieved in skewed data, one row for the results in non-skewed data and a third one for the results in continuous data. Furthermore, there is one column for each salience set and a column for the average. A '†' attached to a number as a superscript indicates an inappropriate fit.

The first part of Table 1 gives the CFI results. Apparently, a less than optimal model fit was found for the skewed data in all salience sets and the average but for no one of the other types of data. Furthermore, the difference between the averages of continuous data and non-skewed data was below the selected limit of .01. In contrast, the other differences were considerably larger than this limit (continuous data - skewed data: .105; non-skewed data - skewed data: .097) and thus indicated an effect of skewness.

The TLI results are given in the second part. Whereas a less than optimal model fit characterized all the salience sets and the average of the skewed data, a good degree of model fit was observed for continuous data and non-skewed data. The difference between the averages of continuous and non-skewed data (.015) was minor in comparison to the other differences (continuous data - skewed data: .090; non-skewed data - skewed data: .105).

The third and fourth parts of Table 1 include the RMSEA and SRMR results. Although the averages seemed to signify that there were different degrees of model fit for skewed data on one hand and continuous and non-skewed data on the other hand, all fit statistic indicated an optimal degree of model fit.

Apparently, the CFI and TLI results showed the expected difference between the types of data respectively items. They signified that the original version of the tau-equivalent model of measurement was appropriate for continuous and non-skewed data but not for skewed data. In contrast, the model fit was good according to almost all the RMSEA and SRMR results. This finding simply indicated a low degree of random variation that probably resulted from the large number of data points. Since the various fit statistics possess different properties, the latter results should not invalidate the former results.

Table 1. Fit results for the original version of the tau-equivalent model in continuous, non-skewed and skewed items for three degrees of salience

Characteristics of data	Salience level			Average
	0.5	0.4	0.3	
CFI				
Skewed	0.865 [†]	0.900 [†]	0.905 [†]	0.890 [†]
Non-skewed	0.987	0.988	0.987	0.987
Continuous	0.995	0.997	0.992	0.995
TLI				
Skewed	0.860 [†]	0.900 [†]	0.933 [†]	0.897 [†]
Non-skewed	0.995	1.005	1.060	1.002
Continuous	1.002	0.926	1.035	0.987
RMSEA				
Skewed	0.038	0.021	0.010	0.023
Non-skewed	0.009	0.005	0.003	0.006
Continuous	0.006	0.003	0.003	0.004
SRMR				
Skewed	0.061	0.047	0.042	0.050
Non-skewed	0.038	0.039	0.038	0.038
Continuous	0.039	0.038	0.038	0.038

[†] A less than optimal degree of model fit is indicated.

3.4 Comparison of the Versions of the Tau-equivalent Model with Respect to Model Fit

This section reports the fit results obtained by the weighted and original tau-equivalent models of measurement separately for each salience level of skewed data. Furthermore, the distribution-free estimation method is considered in combination with the original version. The CFI, TLI, RMSEA and SRMR results are provided in separate parts of Table 2.

Table 2. Fit results for the original and weighted tau-equivalent models achieved for skewed items

Version of the model	Salience level			Average
	0.5	0.4	0.3	
CFI				
Original	0.865 [†]	0.900 [†]	0.905 [†]	0.890 [†]
Original ADF	-	-	-	-
Original SB	0.847 [†]	0.896 [†]	0.904 [†]	0.882 [†]
Weighted	0.984	0.983	0.977	0.981
TLI				
Original	0.860 [†]	0.900 [†]	0.933 [†]	0.897 [†]
Original ADF	-	-	-	-
Original SB	0.843 [†]	0.894 [†]	0.930 [†]	0.889 [†]
Weighted	1.002	1.030	1.093	1.041
RMSEA				
Original	0.038	0.021	0.010	0.023
Original ADF	0.000	0.000	0.000	0.000
Original SB	0.000	0.000	0.000	0.000
Weighted	0.008	0.005	0.003	0.005
SRMR				
Original	0.061	0.047	0.042	0.050
Original ADF	0.078	0.054	0.044	0.058
Original SB	0.071	0.047	0.042	0.053
Weighted	0.041	0.037	0.039	0.039

Note. Original ADF means original tau-equivalent model with asymptotic distribution-free estimation and original SB with robust estimation according to Satorra-Bentler. † A less than optimal degree of model fit is indicated.

This Table shows the same structure as Table 1 with four rows included in each part referring to different combinations of models and estimation methods: the original version of the tau-equivalent model in combination with maximum likelihood estimation (original), this version in combination with distribution-free estimation (original ADF), this version in combination with robust estimation (original SB) and the weighted version of the tau-equivalent model in combination with maximum likelihood estimation (weighted). The fit results for the original version and maximum likelihood estimation are also part of Table 1.

In the first part a less than optimal model fit is indicated by the CFI results for the original version of the tau-equivalent model with maximum likelihood and robust estimation and a good degree of fit for the weighted version of the tau-equivalent model. No result is provided for the original model in combination with distribution-free estimation because the program did not provide this statistic. Furthermore, the differences between the average CFI results for the original and weighted versions (.091 and .099) were considerably larger than the selected limit of .01 and indicated that the weighted version did considerably better than the other one.

The TLI results of the second part shows a similar structure: a less than optimal model fit was indicated by the TLI results of the original version of the tau-equivalent model with maximum likelihood and robust estimation and a good degree of fit by the TLI results of the weighted version of this model. Since all the TLI results obtained by the distribution-free estimation method were very small and also negative, they are omitted. Again, the differences between the averages were rather large (.144 and 0.152). However, it was not clear whether these differences were substantial since there was no established limit associated with a probability level for this fit statistic.

All the results reported in the third and fourth parts of Table 2 indicated a good model fit. All RMSEA results obtained by the distribution-free estimation method and the robust estimation method signified a perfect model fit. In contrast, the SRMR results obtained by the distribution-free estimation method and also the robust estimation method were a bit less favorable than the results achieved by maximum likelihood estimation method although they still indicated a good model fit. Again it needs to be emphasized that the overall good RMSEA and SRMR results are presumably due to a low degree of random variation resulting from the large number of data points.

3.5 Investigating the Constancy of Factor Loadings for the Original Tau-equivalent Model

The results of investigating the constancy of factor loadings after standardization are provided in Table 3 for the original version of the tau-equivalent model of measurement.

Table 3. Results of investigating the difference between the standardized factor loadings of skewed and non-skewed items achieved by the original tau-equivalent model

Characteristics of data	Arithmetic means for		<i>t</i> statistic			Median
	extreme items	moderate items	<i>t</i>	<i>df</i>	<i>p</i>	test <i>p</i>
	First salience level (average size of original loading: 0.5)					
Skewed	0.394	0.284	3.961 ¹	3.02	.02	.03
Non-skewed	0.396	0.394	1.441	6.00	.20	.49
Continuous	0.503	0.498	0.975	6.00	.37	.49
	Second salience level (average size of original loading: 0.4)					
Skewed	0.324	0.229	3.741 ¹	3.02	.03	.03
Non-skewed	0.327	0.325	2.530	6.00	.05	.49
Continuous	0.405	0.403	2.554	6.00	.04	.14
	Third salience level (average size of original loading: 0.3)					
Skewed	0.240	0.167	3.896 ¹	3.01	.03	.03
Non-skewed	0.243	0.242	1.732	6.00	.13	.43
Continuous	0.300	0.299	0.715	6.00	.50	1.0

¹ *t* statistic for unequal variances.

This Table includes the means for the groups of extreme and moderate factor loadings (see the section on the statistical investigation) that were investigated and the outcomes of the comparisons (*t* Test and Median Test). It comprises three parts corresponding to the three salience levels (upper, medium, lower). Each part is composed of three rows providing the results obtained for skewed, non-skewed and continuous data. The first column of Table 3 gives the means of factor loadings for the extreme items and the second column for the moderate items. The third to fifth columns include the *t* Test results and the sixth column the results obtained by the Median Test.

The means obtained for the extreme and moderate items in skewed data differ considerably (salience level 1: 0.100, salience level 2: 0.095, salience level 3: 0.073). The corresponding *t* statistics indicated a substantial difference for all salience sets. In each case the error probability was .03 or smaller. The investigation of the non-skewed data revealed no difference with the exception of the second salience level although in this case the difference of means was rather small (0.002). Similar results were observed for the continuous data (0.002). In both cases the *t* statistic indicated a substantial difference. Since these results were unexpected and were probably due the extremely low degree of variability of factor loadings, the Median Test was applied additionally. This Test could be expected to be less sensitive to the low degree of variability of factor loadings. As it was expected and is obvious from the last column of Table 1, a significant difference was only indicated for skewed data but not for non-skewed and continuous data.

In sum, the standardized factor loadings obtained by means of the original version of the tau-equivalent model of measurement showed the expected lack of constancy in skewed data but not in other data.

3.6 Comparison of the Factor Loadings Obtained by the Original and Weighted Versions

The means of the standardized factor loadings for the extreme and moderate items obtained by the original and weighted versions of the tau-equivalent model in combination with maximum likelihood estimation and by the original version in combination with distribution-free estimation and robust estimation were compared with each other with respect to skewed data. The results are presented separately for the three salience levels together with the results of the comparisons (*t* Test and Median Test) in Table 4.

Table 4. Results of investigating the difference between the standardized factor loadings of skewed and non-skewed items achieved by the original and weighted tau-equivalent models

Version of the model	Arithmetic means for		<i>t</i> statistic			Median test <i>p</i>
	extreme items	moderate items	<i>t</i>	<i>df</i>	<i>p</i>	
	First salience level (average size of original loading: 0.5)					
Original	0.394	0.284	3.961 ¹	3.02	.02	.03
Original ADF	0.497	0.340	3.720 ¹	3.02	.03	.03
Original SB	0.399	0.293	4.009 ¹	3.02	.03	.03
Weighted	0.360	0.365	1.967	6.00	.10	1.0
	Second salience level (average size of original loading: 0.4)					
Original	0.324	0.229	3.741 ¹	3.02	.03	.03
Original ADF	0.387	0.267	3.733 ¹	3.02	.03	.03
Original SB	0.318	0.224	3.812 ¹	3.00	.03	.03
Weighted	0.287	0.288	0.926	6.00	.39	.43
	Third salience level (average size of original loading: 0.3)					
Original	0.240	0.167	3.896 ¹	3.01	.03	.03
Original ADF	0.295	0.205	3.786 ¹	3.02	.03	.03
Original SB	0.249	0.174	3.847 ¹	3.15	.03	.03
Weighted	0.211	0.211	0.000	6.00	1.00	1.0

Note. *Original ADF* means original tau-equivalent model with asymptotic distribution-free estimation and *original SB* with robust estimation according to Satorra-Bentler. ¹ *t* statistic for unequal variances.

In the first salience level a difference of 0.100 between the standardized factor loadings was observed for the original version with maximum likelihood estimation, of 0.157 with distribution-free estimation and of 0.106 with robust estimation. In contrast, in the weighted version the difference was only 0.005. The comparisons with statistical tests indicated a substantial difference for the original version and all estimation methods but no difference for the weighted version.

In the second salience level differences of 0.095, 0.120 and 0.094 characterized the original version of the tau-equivalence model whereas there was virtually no difference (0.001) in the weighted version. The statistical tests signified significant differences for the original version but not for the weighted version.

In the third salience level the computation yielded differences of 0.073, 0.090 and 0.075 for the original version of the tau-equivalent model whereas the difference for the weighted version was smaller than 0.001. The statistical tests indicated substantial differences for the original version, and there was no such difference for the weighted version.

In a nutshell, after standardization the factor loadings obtained by the original version of the tau-equivalent model of measurement reflected the different degrees of skewness of the items respectively the different variances of the items. This effect appeared to be independent of the estimation method. In contrast, there were no corresponding differences if the standardized factor loadings were obtained by the weighted version.

4. Discussion

So far the tau-equivalent model of measurement has not played a major role in psychological research with the exception of its role as the possible theoretical basis of Cronbach's coefficient alpha (Raykov, 1997a; 1997b; 2001). However, even this application has been questioned, and a promising alternative has been proposed meanwhile (McDonald, 1999; Sijtsma, 2009). In an attempt to improve the usefulness of this model it was complemented by the essentially tau-equivalent version that takes the degree of easiness into account. This modification integrated a basic property of IRT models into the tau-equivalent model and thus improved its appropriateness for scaling. However, it did not change the situation concerning its application as part of confirmatory factor analysis. As a model that appears to be especially sensitive to the variability of the variances of items, it is not really well prepared for serving as part of a procedure that is known for its sensitivity to deviations from normality (Curran et al., 1997; Fan & Hancock, 2012). The consideration of weights can be expected to change this situation. Weights can potentially create robustness against such deviations.

In the reported study the tau-equivalent model of measurement was applied for investigating the structure of simulated data. According to the results of this study and also of the theoretical investigation of the consequences of

assuming equal-sized true scores the original version of the tau-equivalent model can only be expected to do well if the items show equal or close to equal variances. This precondition is a severe restriction since in questionnaires as the major source of data to be investigated by means of this and other models there are usually items with a few ordered response categories (Blais & Gerondin, 2011; Lozano et al., 2008). In such items skewness and related variability of variances are very likely. Therefore, robustness to this kind of deviation from the ideal of items with a symmetric distribution is a precondition for achieving favorable results. The weighted version of the tau-equivalent model can include weights that adapt the model to this precondition.

It is interesting to find that the results obtained for the weighted version of the tau-equivalent model in skewed data are quite similar to the results obtained for the original version of the tau-equivalent model in non-skewed data if the maximum likelihood estimation method is applied in both cases. The results obtained by means of the original version in combination with the asymptotic distribution-free estimation method and also with the robust estimation method in skewed data do not show such a degree of similarity with the results for the original version of the model in non-skewed data. This difference is an important finding since in the reported study the similarity with the results obtained for non-skewed data counts most. The favorable results for the weighted version suggest the following interpretation: the weights applied as part of the model simulate the skewness that characterizes the data and, thus, create similarity of data and model. As a consequence of this simulation, it appears to be no more necessary to recommend a robust estimation method.

The weighted version of the tau-equivalent model of measurement can be expected to provide a contribution to the advancement of the investigation of structure. Since the detection of method effects as a source of systematic variance (Campbell & Fiske, 1959) it should be clear that the internal structure of measures is frequently not restricted to one underlying dimension only. Based on the multitrait-multimethod design a number of specific procedures were proposed for structural investigations in considering method effects (e.g., Eid, Lischetzke, Nussbeck, & Triemerweiler, 2003; Zhang & Browne, 2007). The weighted version of the tau-equivalent model of measurement is one step further in the direction of an increased sensitivity for closely linked or overlapping latent dimensions and method effects.

Although the weighted version has been discussed with respect to the skewness resulting from small numbers of response categories, there is considerable flexibility in the selection of the weights. For example, there is the change of item reliability in series of similar items which has become known as position effect (Knowles, 1988). Such a change of true variance can well be represented by a weighted version of the tau-equivalent model of measurement. Furthermore, there are various repeated-measures designs of longitudinal and experimental studies where a systematic change of the variance of the measure is expected (e.g., McArdle, 2009; Schweizer, 2008). The tau-equivalent model can be adapted to such expectations whereas the congeneric model cannot. Consequently, it would not be surprising if the weighted version of the tau-equivalent model would develop into a useful tool for investigating data achieved in the mentioned designs.

References

- Alwin, D. F., & Jackson, D. J. (1980). Measurement models for response errors in surveys: issues and applications. In K. F. Schuessler (Ed.), *Ethical standards of psychologists* (p. 69-119). San Francisco: Jossey-Bass.
- Blais, J.-G., & Gerondin, J. (2011). The influence of labels associated with anchor points of Likert-type response scales in survey questionnaires. *Journal of Applied Measurement, 12*, 370-386.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62-83. <http://dx.doi.org/10.1111/j.2044-8317.1984.tb00789.x>
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105. <http://dx.doi.org/10.1037/h0046016>
- Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41*, 189-225. http://dx.doi.org/10.1207/s15327906mbr4102_5
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255. http://dx.doi.org/10.1207/S15328007SEM0902_5

- Cheung, G. W., & Lau, R. S. (2011). A direct comparison approach for testing measurement invariance. *Organizational Research Methods*.
- Curran, P. J., West, S. G., & Finch, J. F. (1997). The robustness of test statistics to nonnormality and specification errors in confirmatory factor analysis. *Psychological Methods*, 1, 16-29. <http://dx.doi.org/10.1037/1082-989X.1.1.16>
- DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, 13, 440-464. http://dx.doi.org/10.1207/s15328007sem1303_6
- Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod analysis: A multiple indicator CTC(M-1) model. *Psychological Methods*, 8, 38-60. <http://dx.doi.org/10.1037/1082-989X.8.1.38>
- 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. <http://dx.doi.org/10.3102/1076998610396897>
- Ferrando, P. J., & Condon, L. (2006). Assessing acquiescence in binary responses: IRT-related item-factor-analytic procedures. *Structural Equation Modeling*, 13, 420-439. http://dx.doi.org/10.1207/s15328007sem1303_5
- 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>
- Hartig, J., Hölzel, B., & Moosbrugger, H. (2007). A confirmatory analysis of item reliability trends (CAIRT): differentiating true score and error variance in the analysis of item context effects. *Multivariate Behavioral Research*, 42, 157-183. <http://dx.doi.org/10.1080/00273170701341266>
- 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. (2006). *LISREL 8.80*. Lincolnwood, IL: Scientific Software International Inc.
- Kline, R. B. (2005). *Principles and practices of structural equation modeling*. New York: The Guilford Press.
- Knowles, E. S. (1988). Item context effects on personality scales: measuring changes the measure. *Journal of Personality and Social Psychology*, 55, 312-320. <http://dx.doi.org/10.1037/0022-3514.55.2.312>
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental scores*. Reading, Mass.: Addison-Wesley.
- Lozano, L. M., Garcia-Cueto, E., & Muniz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 4, 73-79. <http://dx.doi.org/10.1027/1614-2241.4.2.73>
- McArdle, J. J. (2009). Latent variable modeling of differences and change with longitudinal data. *Annual Review of Psychology*, 60, 577-605. <http://dx.doi.org/10.1146/annurev.psych.60.110707.163612>
- McCullagh, P., & Nelder, J. A. (1985). *Generalized linear models*. London: Chapman and Hall.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mulaik, S. A., & Quartelli, D. A. (1997). First order or higher order general factor? *Structural Equation Modeling*, 4, 193-211. <http://dx.doi.org/10.1080/10705519709540071>
- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, A*, 135, 370-384.
- Raykov, T. (1997a). Scale reliability, Cronbach's coefficient alpha, and violations of the essentially tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research*, 32, 329-353. http://dx.doi.org/10.1207/s15327906mbr3204_2
- Raykov, T. (1997b). Estimation of composite reliability for congeneric measures. *Applied Psychological Measure-*

- ment, 21, 173-184. <http://dx.doi.org/10.1177/01466216970212006>
- Raykov, T. (2001). Bias of coefficient alpha for fixed congeneric measures with correlated errors. *Applied Psychological Measurement, 25*, 69-76. <http://dx.doi.org/10.1177/01466216010251005>
- Reddy, S. K. (1992). Effects of ignoring correlated measurement error in structural equation models. *Educational and Psychological Measurement, 52*, 549-570. <http://dx.doi.org/10.1177/0013164492052003005>
- Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 16*, 19-31.
- 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.
- 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. (2011). Scaling variances of latent variables by standardizing loadings: applications to working memory and the position effect. *Multivariate Behavioral Research, 46*, 938-955. <http://dx.doi.org/10.1080/00273171.2011.625312>
- Schweizer, K. (2012). The position effect in reasoning items considered from the CFA perspective. *International Journal of Educational and Psychological Assessment, 11*, 44-58.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*, 107-120. <http://dx.doi.org/10.1007/s11336-008-9101-0>
- Vautier, S., Raufaste, E., & Cariou, M. (2003). Dimensionality of the Revised Life Orientation Test and the status of filler items. *International Journal of Psychology, 38*, 390-400. <http://dx.doi.org/10.1080/00207590344000222>
- Zhang, G., & Browne, M. W. (2007). Design and analysis of incomplete multitrait-multimethod studies from a multiplicative perspective. *Psychometrika, 72*, 361-375. <http://dx.doi.org/10.1007/s11336-004-1224-3>