

Does the Effect of a Time Limit for Testing Impair Structural Investigations by Means of Confirmatory Factor Models?

Educational and Psychological
Measurement

2019, Vol. 79(1) 40–64

© The Author(s) 2018

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0013164418770824

journals.sagepub.com/home/epm



Karl Schweizer¹, Siegbert Reiß¹, and Stefan Troche² 

Abstract

The article reports three simulation studies conducted to find out whether the effect of a time limit for testing impairs model fit in investigations of structural validity, whether the representation of the assumed source of the effect prevents impairment of model fit and whether it is possible to identify and discriminate this method effect from another method effect. Omissions due to the time limit for testing were not considered as missing data but as information on the participants' processing speed. In simulated data the presence of a time-limit effect impaired comparative fit index and nonnormed fit index whereas normed chi-square, root mean square error of approximation, and standardized root mean square residual indicated good model fit. The explicit consideration of the effect due to the time limit by an additional component of the model improved model fit. Effect-specific assumptions included in the model of measurement enabled the discrimination of the effect due to the time limit from another possible method effect.

Keywords

time-limit effect, speededness, structural validity, omissions, model of measurement, processing speed

The limitation of the time span for completing the items of psychological scales can prevent some or even all participants from processing all items. The effect of the

¹Goethe University Frankfurt, Frankfurt am Main, Germany

²University of Witten/Herdecke, Witten, Germany

Corresponding Author:

Karl Schweizer, Department of Psychology, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 6, 60323 Frankfurt am Main, Germany.

Email: k.schweizer@psych.uni-frankfurt.de

limitation is obvious in the increasing number of omissions that means incomplete data. In this article, this effect is considered as a method effect that has been suggested to impair the validity of measurement (Lu & Sireci, 2007). It is studied how this effect influences model fit in structural investigations and whether it is possible to control for its influence by including a representation of the source of this effect into the confirmatory factor model used for investigating the data. Furthermore, the possibility to identify this effect and to discriminate it from another possible method effect is investigated using simulated data.

The effect of a time limit for testing has for a long time been in the focus of scientific discussion in the measurement area (see, Gulliksen, 1950). For a time, there was the idea that such a limit simply means speeded testing that is the opposite of power testing (Lord & Novick, 1968). However, the investigation of the details of the effect of a time limit reveals a functional relationship to participants' characteristics and other influences and makes apparent that even the allowance of a very large time span may be insufficient for very slow participants (Oshima, 1994). Recent research on the effect of a time limit for testing has addressed parameter estimation (Bolt, Cohen, & Wolack, 2002; Goegebeur, DeBoeck, Wollack, & Cohen, 2008), the influence on validity (Lu & Sireci, 2007), and the possibilities of preventing it (van der Linden & Xiong, 2013). Within the framework of item-response theory, this effect is referred to as speededness (Oshima, 1994). Tailored testing and branched testing are proposed to overcome problems resulting from omissions due to the time limit (Kubinger, 2016; van der Linden & Xiong, 2013).

In investigations of the underlying structure of data using factor-analytic methods traditional ways of treating omissions are the elimination of cases showing omissions or their replacement by means of imputation methods (J. W. Graham, 2009). The elimination of cases, however, leads to a decrease of the sample size while imputation methods mean extrapolation on the basis of an assumed underlying model. A common characteristic of these approaches is that omissions are considered as missing data. In contrast, the present article regards omissions as information on the participants' processing speed. This means that a second source of responding is assumed besides the source associated with the construct that is represented by the scale.

The Effect due to the Time Limit

A time limit for testing would only be a minor problem if all participants of a sample would work at the same speed so that the construct captured by the scale was the only source of systematic responding. However, there are well-known interindividual differences in the speed of mental processing (Jensen, 2006; Roberts & Stankov, 1999), and the most recent model of cognitive abilities includes cognitive processing speed, decision and reaction speed, and also psychomotor speed as broad (stratum II) abilities (McGrew, 2009). Because of interindividual differences in these abilities, some participants may reach their highest possible scores within the time limit for testing

whereas others stay below what they could reach otherwise. Therefore, a time limit for testing is likely to mean distortion of measurement.

Processing speed as the result of the coaction of broad abilities may follow the normal distribution. Although the normal distribution is frequently assumed but not given (Micceri, 1989), it appears to be the best-possible assumption regarding the distribution of processing speed because of the large number of specific sources underlying the broad abilities contributing to what is observable. The idea regarding the contributions of the specific sources originates from the theory of cognitive abilities (McGrew, 2009) with three broad (stratum II) speed abilities (cognitive processing speed, decision and reaction speed, and, tentatively identified, psychomotor speed). Each of these broad abilities is built up on a number of specific (stratum I) abilities. Assuming that subsets of specific abilities are stimulated repeatedly during the completion of the various items of a reasoning test, the various contributions to processing speed in responding may create a normal distribution. This rationale is given by the central limit theorem (Fischer, 2011), which suggests that composites of independent contributions approximate a normal distribution. However, the details of the contributions of the specific sources are not known so that the normal distribution of processing speed in the population is a tentative assumption.

When starting from the assumption of the normal distribution, it can be expected that the number of omissions increases from the first item not completed by all participants to the last item according to the normal ogive. The use of the logistic function simplifies the computations regarding the course of the ogive; the logistic function is, for example, also preferred as basis of the item-characteristic curve of item-response models (Lord, 1965).

These considerations suggest that the probabilities of omissions in the columns of a data matrix due to a lack of processing speed are reflected by the logistic function. Assume the probability of entry X of the i th column of the data matrix to be an omission is given by

$$\Pr(X_i \text{ is omission}) = \frac{e^i}{1 + e^i}. \quad (1)$$

A disadvantage of this equation is that specific time limits are not reflected. Adaptability can be achieved by including a constant that fixes the .5-probability point of this function to where this point is expected in the sequence of items. This constant is addressed as turning point tp . To make the difference apparent, i is replaced by k ($k = 1, \dots, p$) in such a way that

$$\Pr(X_k \text{ is omission}) = \frac{e^{k-tp}}{1 + e^{k-tp}}. \quad (2)$$

Correspondence of k and tp yields $\Pr(X_k \text{ is omission}) = .5$.

Figure 1 illustrates how individual differences in processing speed may influence the percentage of omissions due to three different time limits addressed as strong,

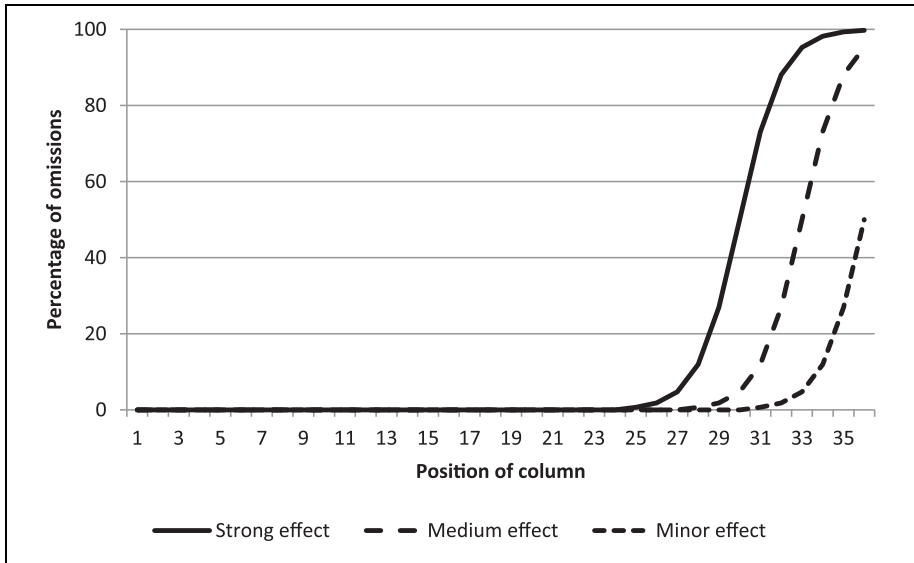


Figure 1. Curves reflecting the percentages of omissions for strong, medium, and weak effects due to a time limit.

medium, and weak effects due to a time limit for testing when the course of the curve follows the logistic function.

The solid curve illustrated the effect of a time limit that allows all participants to complete about two thirds of a set of 36 items of a scale but virtually no one to complete all items. The time limit leading to the dashed curve gives some participants the opportunity to complete all items. In the remaining curve, 50% of the participants can complete all items.

Omissions due to a time limit for testing mean a decrease of the number of valid responses in some columns of the data matrix. The effect of the time limit can be so strong that the remaining number of valid responses does no more satisfy the requirements for factor-analytic investigations (Hogarty, Hines, Kromrey, Ferron, & Mumford, 2005; MacCallum, Widaman, Zhang, & Hong, 1999). The percentage of remaining entries can be so small that the estimated probability of a specific entry can no longer be considered as an accurate reflection of the probability for the same column of entries when there is no time limit. Figure 2 illustrates how the three effects due to a time limit change the percentage of the entries considered as correct responses.

The dotted line that is merged with the other lines for most of its course represents the percentage of correct responses due to the hypothetical increasing degree of item difficulty that is similar to what can be found in many established scales with fixed orders of items. Furthermore, the solid line departs from the dotted line in the last

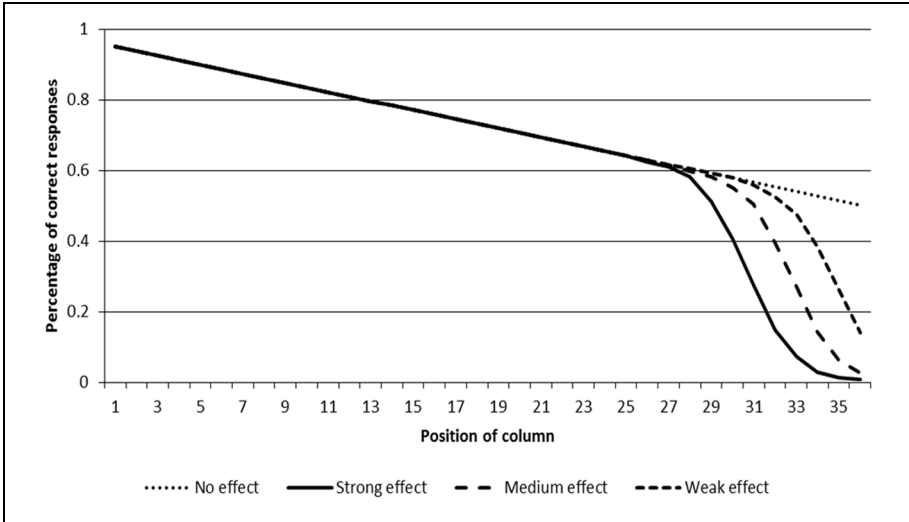


Figure 2. Curves reflecting the percentage of correct responses resulting from the dichotomization of continuous data and effects due to a time limit.

third of the sequence of items. Its separation from the other lines is because of the strong effect. Furthermore, there are lines describing the medium and weak effects.

Before closing this section, it needs to be mentioned that in real data the separation between the areas of valid responses and of omissions due to the effect of a time limit may not be so strict, as is suggested by the figures. A strict separation presupposes that all participants complete all items carefully and there is no checking of response options at random or without sufficient consideration. However, research regarding the attitude in test taking indicates that participants are increasingly aware of the chance of reaching a higher score when checking the remaining items at random before the time limit is reached (Must & Must, 2013) and behave accordingly.

The Treatment of Omissions

Nowadays imputation methods are the preferred means for replacing omissions. Most of the imputation methods assume only one source of systematic responding (e.g., De Ayala, Plake, & Impara, 2001; Finch, 2008; Hohensinn & Kubinger, 2011) whereas a few model-based imputation methods additionally assume another unobserved and unknown source (Holman & Glas, 2005; O'Muircheartaigh & Moustaki, 1999). All these methods are concerned with the replacement of omissions by means of imputation in the sense of extrapolation on the basis of the available data (J. W. Graham, 2009).

The approach taken in this article differs from these imputation methods in that it is assumed that omissions are because of a known specific source, which is (lack of)

processing speed. The information on this source is included in the shape of the distribution of omissions. It needs to be transferred into the covariance pattern of the covariance matrix that serves as input to confirmatory factor analysis (Jöreskog, 1970). According to the previous reasoning the logistic function provides this information so that it is necessary to consider how this function influences the covariance pattern. The probability-based covariance that does not require knowledge of the individual entries of a data matrix serves this purpose. The covariance of two binary random variables X_i and X_j ($i, j = 1, \dots, p$) with zero and one as values $\text{cov}(X_i, X_j)$ is defined as

$$\text{cov}(X_i, X_j) = \Pr(X_i = 1 \wedge X_j = 1) - \Pr(X_i = 1) \Pr(X_j = 1), \quad (3)$$

where $\Pr(X_i = 1)$ and $\Pr(X_j = 1)$ represent the probabilities that X_i and X_j are equal to one (i.e., a correct response) and $\Pr(X_i = 1 \wedge X_j = 1)$ represents the probability of the combination of them.

Each one of the three probabilities of Equation (3) depends on the construct and processing speed, as is especially obvious in the case of the probability that X_i is 1:

$$\Pr(X_i = 1) = \Pr(X_i = 1 | \xi) \times [1 - \Pr(X_i \text{ is omission})]. \quad (4)$$

The first component of the product is the probability of one due to the influence of the primary source that is construct ξ . The second one is the probability that an entry of the i th column is *no* omission. It is obtained by means of the logistic function (see Equation 2) that reflects processing speed. Since the logistic function varies between 0 and 1, the construct influence is assumed to be constant and the not explicitly considered difficulty may also be constant or vary to a minor degree only; it is mainly the logistic function that finds its expression in the probability within the range of columns with omissions. Similar considerations regarding the other components of the product give rise to the expectation that covariances of X_i and X_j in the range of columns with omissions mainly reflect the different values of the logistic function for different columns. Using these covariances, omissions should not be replaced by ones.

As a consequence of considering a second source of responding, the appropriateness of the one-factor model of measurement needs to be questioned. There are a number of one-factor models of measurement used in confirmatory factor analysis (J. M. Graham, 2006). Structural investigations are mostly conducted by means of the congeneric model (Jöreskog, 1971). This model is described by the following equation:

$$\mathbf{x} = \boldsymbol{\lambda}\xi + \boldsymbol{\delta}, \quad (5)$$

where the $p \times 1$ vector \mathbf{x} represents the centered data, the $p \times 1$ vector $\boldsymbol{\lambda}$ the factor loadings, ξ the latent variable and the $p \times 1$ vector $\boldsymbol{\delta}$ the error components. It implicitly assumes that ξ is the only source of systematic responding, and a good model fit is signified if there is actually only one such source. An additional source influencing the pattern of entries of the data matrix that is roughly reflected by the logistic

function of Equation (2) is unlikely to be represented by ξ unless the primary source is missing.

The Representation of the Effect due to a Time Limit

The representation of the effect due to a time limit can be accomplished in different ways if there is another source that is the primary source of responding. It is convenient to concentrate on processing speed as source since the effect is due to the coaction of the time limit for testing that can be assumed to be constant within a sample and processing speed. First, processing speed can be considered as moderator since processing speed determines in how many items the primary source of responding is effective. It is thought that in these items the response mainly depends on the primary source but not on processing speed. In the remaining items any further response is prevented by (lack of) processing speed. Available statistical programs for using the traditional models of measurement do not support such a kind of moderation.

Second, processing speed can be perceived as an independent source of variation in data. This treatment is justified by theory suggesting speed as source of cognitive performance and as component of models of cognitive ability (Jensen, 2006; McGrew, 2009; Roberts & Stankov, 1999). The consideration of this source requires the transformation of the one-factor model of measurement into the two-factor model. In this model and the corresponding model of the covariance matrix the latent variable representing this additional source can be expected to account for the variation due to the number of omissions from column to column of the data matrix.

The model of measurement considering this possibility includes two systematic sources of responding: the source reflecting the construct and the source reflecting processing speed. The latter one is identified by the subscript p-speed. A formal representation of this model of measurement is given by

$$\mathbf{x} = \boldsymbol{\lambda}_{\text{construct}}\xi_{\text{construct}} + \boldsymbol{\lambda}_{\text{p-speed}}\xi_{\text{p-speed}} + \boldsymbol{\delta} \quad (6)$$

where the $p \times 1$ vector \mathbf{x} represents the centered data, the $p \times 1$ vectors $\boldsymbol{\lambda}_{\text{construct}}$ and $\boldsymbol{\lambda}_{\text{p-speed}}$ the factor loadings, $\xi_{\text{construct}}$ and $\xi_{\text{p-speed}}$ the latent variables and the $p \times 1$ vector $\boldsymbol{\delta}$ the error components.

Since it is assumed that the source representing the construct contributes to all items equally, the factor loadings included in $\boldsymbol{\lambda}_{\text{construct}}$ are constrained to show equal size. It is convenient to select the number one as factor loading so that

$$\boldsymbol{\lambda}_{\text{construct}} = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ 1 \end{bmatrix} \quad (7)$$

and to estimate the variance of the latent variable representing the construct. The factor loadings of Equation (7) can be considered as realization of the tau-equivalent model of measurement (J. M. Graham, 2006). The numbers of Equation (7) undergo an additional link transformation before they are applied in confirmatory factor analysis (see method section of Study 1).

Processing speed can be reflected by the factor loadings in two different ways. First, there is the possibility to design the representation according to the cumulative distribution of omissions. This representation is based on the assumptions that the increase of the number of omissions follows the normal ogive and that the normal ogive is approximated by the logistic function. Therefore, the k th factor loading λ_k is in the first step defined as function $f_{\text{logistic}}(k-tp)$ of the difference between k referring to the k th column of the matrix and the turning point of the logistic function tp :

$$\lambda_k = f_{\text{logistic}}(k - tp). \quad (8)$$

The turning point of the logistic function is the point k that yields the .5-probability according to Equation (2). Next $f_{\text{logistic}}(k-tp)$ is defined with respect to the logistic function:

$$f_{\text{logistic}}(k - tp) = \frac{e^{k-tp}}{1 + e^{k-tp}}. \quad (9)$$

This definition assures that the constrained factor loadings reflect the effect due to the time limit. Finally, the vector for representing processing speed $\lambda_{\text{p-speed}}$ is specified as

$$\lambda_{\text{p-speed}} = \begin{bmatrix} f_{\text{logistic}}(1 - tp) \\ f_{\text{logistic}}(2 - tp) \\ \vdots \\ f_{\text{logistic}}(p - tp) \end{bmatrix}. \quad (10)$$

Such a specification of the factor loadings already served well for representing processing speed in an empirical investigation (Schweizer & Ren, 2013).

The two-factor model of measurement including constraints of factor loadings according to Equations (7) and (10) is referred to as *tp-specific model*. It is a very specific model since it can be expected to indicate good model fit only if the turning point of the model corresponds with the turning point of the curve characterizing the omissions of the data. If the turning point underlying the cumulative distribution of omissions is not known, it is necessary to investigate a set of possible turning points and to find the best-fitting one.

Second, the need to search for the best-fitting turning point when using the *tp-specific model* can be overcome by selecting a less specific way of

representing processing speed. The characterization as less specific refers to a combination of free and fixed factor loadings. Zero is selected as factor loading for the g items that can be assumed to fall into the range of items which can be completed by all participants whereas the other factor loadings are free for estimation. In this case, the vector for representing processing speed $\lambda_{p\text{-speed}}$ is specified as follows:

$$\lambda_{p\text{-speed}} = \begin{bmatrix} 0 \\ \cdot \\ 0 \\ \lambda_{g+1} \\ \lambda_{g+2} \\ \cdot \\ \lambda_p \end{bmatrix}. \quad (11)$$

Uncertainty about g should not be considered as a disadvantage since the factor loadings that are especially important for the identification of the factor reflecting processing speed are expected to show the larger sizes at the end of the sequence of items. Furthermore, it is important that a few factor loadings are fixed to zero to retain the confirmatory character of the model. This model is referred to as *hybrid model*.

Considerations Regarding the Identification of the Type of Effect

This section addresses the question whether it is possible to identify a specific type of method effect without drawing conclusions from the inspection of factor loadings. There is a variety of types of method effects that may be found in a data set, as for example, the ceiling effect (Wang, Zhang, McArdle, & Salthouse, 2009), the item-position effect (Zeller, Krampen, Reiß, & Schweizer, 2016), the item-wording effect (DiStefano & Motl, 2012), the effect of speededness (Oshima, 1994), and so on. Distinguishing between effects on the basis of factor loadings can be difficult, as for example, in the case of the item-position effect and the effect due to a time limit. In both cases, the largest factor loadings are expected for the last items of the sequence of items.

The identification of a specific method effect by means of model fit requires correctness of the model of measurement. In simulation studies correctness implies that there is correspondence of the assumptions used for data generation and the model designed for data analysis. Using a model of measurement with fixed factor loadings demands exact correspondence. In the case of data showing the effect due to a time limit, it is necessary that the vector of values for creating omission and the vector including the constraints for the factor loadings on the p -speed latent variable: $\lambda_{p\text{-speed}(\text{data_generation})}$ and $\lambda_{p\text{-speed}(\text{data_analysis})}$, correspond:

$$\lambda_{p\text{-speed}(\text{data_generation})} = \begin{bmatrix} f_{\log\text{istic}}(1 - tp_{\text{data_generation}}) \\ f_{\log\text{istic}}(2 - tp_{\text{data_generation}}) \\ \vdots \\ f_{\log\text{istic}}(p - tp_{\text{data_generation}}) \end{bmatrix} = \begin{bmatrix} f_{\log\text{istic}}(1 - tp_{\text{data_analysis}}) \\ f_{\log\text{istic}}(2 - tp_{\text{data_analysis}}) \\ \vdots \\ f_{\log\text{istic}}(p - tp_{\text{data_analysis}}) \end{bmatrix} = \lambda_{p\text{-speed}(\text{data_analysis})}. \quad (12)$$

Most important, the turning points must correspond. Any deviation of the turning points can be expected to cause an impairment of model fit.

A good model fit can also be achieved by means of a model of measurement with free factor loadings. Equation (13) provides a description of the situation resulting from the use of free factor loadings for data analysis by means of the congeneric model of measurement, $\lambda_{\text{congeneric_model}(\text{data_analysis})}$:

$$\lambda_{p\text{-speed}(\text{data_generation})} = \begin{bmatrix} f_{\log}(1 - tp) \\ f_{\log}(2 - tp) \\ \vdots \\ f_{\log}(p - tp) \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_p \end{bmatrix} = \lambda_{\text{congeneric_model}(\text{data_analysis})}. \quad (13)$$

Parameter estimation can lead to estimates of the factor loadings corresponding to the values obtained by means of the logistic function. This characteristic of the congeneric model is due to the high degree of adaptability of free parameters to the characteristics of data. Using free factor loadings, it is not necessary to search for the turning point. But it may also mean that there is no sensitivity for the characteristics of what exactly is underlying the data.

In sum, only models including specific fixed factor loadings are likely to identify a specific structural effect whereas a good model fit appears to be possible by means of both free and fixed factor loadings.

Objectives

We report three partly overlapping studies. The main objective of the first study was to investigate the impairment in model fit due to the effect of a time limit. Three different degrees of the effect were considered: strong, medium, and weak effects. The effects were inserted in simulated data assuming one underlying source of responding. Another objective of this study was the identification of the fit index that was most sensitive to the effect due to a time limit. The objective of the second study was the investigation of the improvement in model fit that was expected due to the use of two-factor models including a representation of the assumed source of the effect. The tp -specific and hybrid models served this purpose. They were compared with one-factor models using the fit index identified as especially sensitive to the effect due to a time limit in the first study. In the third study the objective was to investigate how

well models with fixed and free factor loadings discriminated between different method effects.

Study I: The Effect due to a Time Limit on Structural Investigations

This study served the investigation of the impairment of model fit due to the effect of the time limit when conducting structural investigations. Although different degrees of the effect due to a time limit were realized in generating the data, for this investigation it was supposed that the data would show one underlying source of responding only. This supposition suggested the selection of a one-factor model of measurement (J. M. Graham, 2006) for analyzing the data. It was expected that an increasing degree of the effect would be associated with decreasing model fit. Furthermore, information about the tolerable degree of the effect was expected.

Because of the specificity of the effect due to a time limit, the various fit indices developed to reflect different characteristics of data (DiStefano, 2016) could be expected to show different degrees of sensitivity regarding this effect. It was of interest to classify the indices into two types: fit indices that were not sensitive to the effect and fit indices that discriminated between the levels of the effect. The second type of fit indices was of particular interest for the second and third studies reported in this article.

Method

Data Generation. In this and the following studies, the investigated matrices of structured random data included 36 columns and 500 rows to have data showing major characteristics of data sets originating from assessment using Advanced Progressive Matrices, a frequently used scale of cognitive ability. Most data were generated by means of a uniform relational pattern reflecting the assumption that there was only one source of responding. In the last study, another relational pattern was additionally used. The relational pattern was achieved by means of factor loadings of .40, and it was also assumed that the contribution of random error would lead to variances of 1.0. The data generation was expected to yield matrices with columns composed of ones and zeros as entries whereby the probabilities of ones decreased starting from the first column ($p = .95$) to the last column ($p = .50$). The effect due to a time limit was realized by randomly selecting an increasing number of rows using the logistic function when moving from the 25th column to the last column. In the selected rows, the entries were changed into omissions starting from the column where the row was selected to the last column.

Five hundred 500×36 matrices of random data were generated and specified to show four levels of the effect due to the time limit. At first each column consisted of continuous and normally distributed data ($N(0, 1)$). The underlying structure according to the relational pattern was induced by the procedure proposed by Jöreskog and Sörbom (2001) that included the computation of regressions weights and their use for

computing weighted composites. Subsequently, there was dichotomization into zeros and ones to achieve the probabilities reported in the previous paragraph. Afterward, the omissions were included as described to obtain matrices showing three levels of omissions: omissions due to weak, medium, and strong effects in addition to no omissions. These levels were realized by setting the turning point of the logistic function to the 36th, 33th, and 30th items for the weak, medium, and strong effects. The turning point for the strong effect was selected so that virtually all rows of the last columns of a data matrix showed omissions. The selection of the other turning points was guided by the idea that there should be equal distances between the turning points so that increasing degrees of good model fit could be expected.

Practical problems led to a minor deviation from the described procedure in realizing the strong effect. The last columns of some generated matrices included not a single one so that the corresponding covariance matrices were not positive definite. Therefore, the procedure was adjusted in such a way that in no column of a data matrix the percentage of omission surmounted 98%.

Data Analysis. The data were investigated by means of confirmatory factor models with one factor only to represent the single underlying source of responding. The aim was to examine the impairment of model fit due to the effect of a time limit. One model of measurement was the model with free factor loadings according to Equation (5). The tau-equivalent model of measurement was the second model. It differed from the first one in the type of factor loadings. The factor loadings were constrained according to Equation (7). Both models could be considered as correct in the absence of the effect since the relational pattern used in data generation assumed factor loadings of equal size. In the model with free factor loadings the variance parameter was set to one, and in the model with fixed factor loadings it was set free for estimation.

Because of the binary nature of the simulated data it was necessary to select an approach for overcoming the difference between the data included in the matrices and the data expected by the model of measurement. The simulated data were binary and followed the binomial distribution whereas continuous data following the normal distribution were expected. The established way of overcoming the difference was to use tetrachoric correlations as input to confirmatory factor analysis (Muthén, 1984). However, such correlations could not be expected to do well because of the very low probabilities of ones if there was a strong or medium effect due to a time limit. The thresholds implicitly estimated as an intermediary step in computing such correlations show a high degree of accuracy close to $p = .5$ but the accuracy decreases as the probability approaches either the upper or lower boundary. Furthermore, correlations did not show additivity. An alternative way of overcoming the difference between simulated and expected data was provided by the threshold-free approach (Schweizer, 2013; Schweizer, Ren & Wang, 2015). Characteristic features of this approach were the computation and use of probability-based covariances and the use of a link transformation (see, McCullagh & Nelder, 1985) in investigating the data. The link

transformation was accomplished by assigning weights w_i ($i = 1, \dots, p$) to the constraints for factor loadings. These weights were defined as

$$w_i = \sqrt{\Pr(X_i = 1)[1 - \Pr(X_i = 1)]}, \quad (14)$$

where $\Pr(X_i = 1)$ is the probability of a correct response. Since the weights for accomplishing the link transformation comprise the probability of a correct response as ingredients, they are specific for each item. This implicitly means that the relationships among the constraints for the factor loadings are modified by performing the link transformations. As a consequence, it is unlikely that, for example, after the link transformation the constraints of Equation (7) still show equal sizes. In simulation studies, probability-based covariances led to good fit results in matrices with 200 rows only. This alternative approach was selected for the study.

The parameter estimation in investigating the data was conducted using the maximum likelihood estimation method by means of LISREL (Jöreskog & Sörbom, 2006). The following fit indices were recorded for evaluating model–data fit: normed χ^2 , root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), normed fit index (NFI), nonnormed fit index (NNFI), and comparative fit index (CFI). Cutoffs provided by Kline (2005) and Hu and Bentler (1999) served the evaluation of the results (normed $\chi^2 < 3$, RMSEA $< .06$, SRMR $< .08$, GFI $> .90$, AGFI $> .90$, NFI $> .95$, NNFI $> .95$, CFI $> .95$; see also, DiStefano, 2016). The comparisons between the levels were conducted by means of the CFI difference (see, Chen, 2007; Cheung & Rensvold, 2002).

Results

The fit results observed in investigating sets of 500 matrices with no effect, weak, medium, and strong effects due to a time limit were averaged. Table 1 presents the means and standard deviations obtained by means of the one-factor model with free factor loadings.

The means of the normed χ^2 , RMSEA, and SRMR statistics indicated a good model fit for all levels of the effect. According to GFI and AGFI, the mean model fit was good with the exception of the level showing the strongest effect. NNFI and CFI only indicated a good model fit for the level showing no effect due to a time limit. For not even one level, the fit was good according to NFI. The comparisons of the levels by means of the CFI difference (see last column) revealed that each pair of neighboring levels differed from each other. The means and standard deviations are included in Table 2. The same pattern of means signifying good or bad results was observed for the model with fixed factor loadings as for the model with free factor loadings with one exception. The AGFI results suggested a good model fit for the first and second levels of the effect only instead of for the first to third levels. The

Table 1. Means and Standard Deviations of the Fit Results for Different Effect Levels (No Effect, Weak Effect, Medium Effect, Strong Effect) Obtained by the Congeneric Model of Measurement ($N = 500$).

Effect level	Fit index								
	Normed χ^2	RMSEA	SRMR	GFI	AGFI	NFI	NNFI	CFI	Δ CFI
No effect									
Mean	1.11	.014	.041	.932	.923	.765	.961	.963	—
SD	0.06	.00	.00	.00	.00	.03	.02	.02	—
Weak effect									
Mean	1.22	.021	.043	.925	.916	.724	.925	.929	.03*
SD	0.07	.00	.00	.00	.00	.03	.02	.02	
Moderate effect									
Mean	1.47	.031	.048	.911	.901	.668	.856	.864	.07*
SD	0.10	.00	.00	.01	.01	.04	.03	.03	
Strong effect									
Mean	2.11	.047	.058	.877	.863	.558	.691	.709	.16*
SD	0.29	.01	.00	.01	.02	.07	.09	.08	

Note. SD = 0.00 means $SD < 0.005$. SD = standard deviation. RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; CFI = comparative fit index; GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; NFI = normed fit index; NNFI = nonnormed fit index.

* $p > .05$.

comparison of the levels by means of the CFI difference revealed substantial differences between all pairs of neighboring levels.

In both models, CFI and NNFI showed the largest range of values (.25 and .27, respectively) and included all categories of model fit from good to bad. All differences between the models with free and fixed factor loadings were smaller than .01 for NNFI. In contrast, only for no effect and the weak effect the CFI differences of the two models were smaller than .01. For medium and large effects, the differences were .016 indicating that the model with fixed factor loadings fitted the data worse than the other one.

Discussion

Although the close inspection of the mean results revealed an effect on model fit in all indices, there was considerable variation regarding the degree of impairment. No model misfit was indicated by normed χ^2 , RMSEA, and SRMR. NNFI and CFI appeared to be most sensitive for the effect due to a time limit. They signified good model fit only when the models were applied to data showing no effect due to a time limit. The comparison of the CFIs for the different levels of the effect revealed a significant difference between each pair of neighboring levels.

Table 2. Means and Standard Deviations of the Fit Results for Different Effect Levels (No Effect, Weak Effect, Medium Effect, Strong Effect) Obtained by the Model of Measurement With Fixed Factor Loadings ($N = 500$).

Effect level	Fit index								Δ CFI
	Normed χ^2	RMSEA	SRMR	GFI	AGFI	NFI	NNFI	CFI	
No effect									
Mean	1.13	.016	.047	.927	.922	.747	.955	.955	—
SD	0.06	.00	.00	.00	.00	.03	.02	.02	
Weak effect									
Mean	1.24	.022	.048	.921	.915	.706	.920	.921	.03*
SD	0.07	.00	.00	.00	.00	.04	.02	.02	
Moderate effect									
Mean	1.52	.032	.054	.904	.898	.642	.848	.848	.07*
SD	0.10	.00	.00	.01	.01	.04	.03	.03	
Strong effect									
Mean	2.00	.048	.062	.870	.862	.533	.688	.692	.16*
SD	0.26	.01	.00	.01	.01	.07	.09	.08	

Note. SD = 0.00 means $SD < 0.005$. SD = standard deviation. RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; CFI = comparative fit index; GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; NFI = normed fit index; NNFI = nonnormed fit index.

* $p > .05$.

Study 2: The Influence of the Representation of the Effect on Model Fit

This study served the investigation of whether the negative consequences of the effect due to a time limit on model fit could be avoided by including a representation of the source of this effect into the model of measurement. In the simulation study, the negative consequences of the omissions created according to the logistic function were assumed to reflect the influence of processing speed that could be traced back to broad cognitive abilities (McGrew, 2009). There was good reason for expecting that a second latent variable would capture the influence of processing speed on the pattern of covariances, as outlined in the theoretical section, and would eliminate the negative consequences. To investigate this idea, one-factor models neglecting the effect due to a time limit were compared with two-factor models representing both the construct and the source. Because of its particular sensitivity to the effect due to a time limit and the possibility to check the statistical significance of differences, only CFI results were considered.

Method

Four models of measurement were used in confirmatory factor analysis. There were the one-factor model with free factor loadings according to Equation (5), the one-

Table 3. Means and Standard Deviations of the CFI Results for Different Effect Levels (No Effect, Weak Effect, Medium Effect, Strong Effect) Obtained by One- and Two-Factor Models with Free and Fixed Factor Loadings ($N = 500$).

Effect level	Type of model						Overall comparison ^a
	Free, one factor	Fixed, one factor	Comparison ^b	tp-specific model	Hybrid model	Comparison ^c	
	Mean (SD)	Mean (SD)		Mean (SD)	Mean (SD)		
No effect	.963 (.02)	.955 (.02)	.008	.955 (.02)	.960 (.02)	-.005	-.002
Weak effect	.929 (.02)	.920 (.02)	.009	.953 (.02)	.948 (.02)	.005	.026*
Medium effect	.864 (.03)	.848 (.03)	.016*	.934 (.02)	.926 (.02)	.008	.074*
Strong effect	.709 (.08)	.692 (.08)	.017*	.876 (.03)	.873 (.03)	.003	.174*

Note. CFI = comparative fit index.

^aCFI difference between the means of the results for the two-factor models (tp-specific model and hybrid model) and the one-factor models. ^bCFI difference between the results for the one-factor models. ^cCFI difference between the results for the two-factor models (tp-specific model and hybrid model).

* $p > .05$.

factor model with fixed factor loadings according to Equation (7), the two-factor model according to Equations (6), (7), and (10) (the tp-specific model), and the two-factor model with fixed and free factor loadings (the hybrid model). The first latent variable of the hybrid model showed factor loadings according to Equation (7) and the other one according to Equation (11). Since the turning point of the logistic function used in data generation was assumed to be unknown, the tp-specific model had to be applied for different turning points, and the best-fitting model had to be selected.

The data generated for the first study were also used in the second study. CFI differences larger than .01 were considered as significant differences (Cheung & Rensvold, 2002).

Results

The means and standard deviations of the CFI results for the combinations of the four models and the four levels of the effect due to a time limit are given in Table 3.

The results for the levels reflected the degrees of the effect and also the absence of an effect. The left-hand part of Table 3 provided the results for the one-factor models and the right-hand part for the two-factor models. For the one-factor models, the CFI means monotonously decreased from values slightly larger than .95 (no effect) to values of about .70 (strong effect). For the two-factor models, a similar pattern was observed but the decrease ended up with values of about .87.

Table 3 also reports the CFI differences between the one-factor models with free and fixed factor loadings on one hand (see column 5) and between the tp-specific

and hybrid models on the other hand (see column 10). All differences between the CFI results of the two-factor models were negligible. In contrast, the comparison of the results for the one-factor models revealed substantial difference for medium and strong effects in favor of the model with free factor loadings.

The last column of this table contains the CFI differences between the means for the two one-factor models and the two two-factor models. In the case of no speed effect, the one-factor models described the data equally well as the two-factor models. In contrast, substantial differences between the one-factor models and the two-factor models were observed in the cases of weak, medium, or strong effects. In the weak effect, the model fit was not good for the one-factor models but could be considered as still acceptable and good or marginally good for the two-factor models. In the medium effect the one-factor models showed model misfit whereas the model fit of the two-factor models was still acceptable. In the strong effect, the results indicated a bad model fit for all models despite a large improvement due to the representation of the effect due to a time limit.

Discussion

The representation of the source of the effect due to a time limit by means of an additional latent variable as part of the model of measurement in confirmatory factor analysis led to a large improvement of the mean CFI results. A good or an acceptable model fit was indicated if there was a weak or even a medium effect. However, the improvement by integrating a second latent variable in the confirmatory factor model was not large enough to compensate for the impairment due to the strong effect.

The two two-factor models of measurement considered in this study did not substantially differ according to the CFI means. There was no advantage for the tp-specific model because of the more accurate representation of the effect. Consequently, the expense due to the search for the best-fitting turning point in the application of the tp-specific model could be avoided if the aim was simply to minimize the consequences of the effect for model fit.

Study 3: The Identification of the Effect due to a Time Limit

To deal appropriately with a method effect that found its expression in bad model fit when conducting a structural investigation, it was important to know about the nature of this effect. In impairment due to a subset of especially homogenous items, for example, the removal of some items might be appropriate (Bandalos & Gerstner, 2016). If the impairment was due to a time limit that enabled processing speed to contribute to performance, on the contrary, the preferable option for dealing with the misfit could be the representation of this source as an additional latent variable in the confirmatory factor model.

One problem in the identification of a method effect, when using factor-analytic methods with free factor loadings, was that various sources could find their

expression in the sizes of factor loadings and also in model fit. A way out of this problem was using precise expectations for constraining the factor loadings since in this case the suspected effect would find its main expression in model fit. Expectations regarding the effect due to a time limit could be based on the assumption that processing speed would follow the normal distribution. Furthermore, there was the dependency of model fit on the correctness of the turning point. Because of this dependency, the fit results for factor loadings fixed according to the effect due to a time limit might show the inverse of a u-shaped course for a sequence of neighboring turning points.

The study reported in the following section investigated whether such a course could be observed when the tp-specific model was applied to data showing a strong effect due to a time limit. Furthermore, it was investigated how this model would perform when there was a uniform effect in a subset of the columns of the data matrix.

Method

A first set of five hundred 500×36 matrices was generated in basically the same way as the data used in the previous study. These data showed a strong effect due to a time limit similar to the one of the other studies. However, the maximum percentage of omissions per column of a matrix was modified so that CFIs larger than .90 could be expected. This modification led to a minimum of 4% of the rows of a matrix that did not show omissions.

A second set of five hundred 500×36 matrices was generated according to a relational pattern that showed another method effect, as could be expected due to a second source equally contributing to a subset of columns. The data generated by means of this relational pattern were referred to as *uniform-effect data*. The relational pattern used for data generation was realized by means of two latent variables. The first latent variable received factor loadings from all 36 manifest variables of the model and the second one from the 25th to 36th manifest variables only. All factor loadings showed the same size that was 0.2828. As a consequence, the off-diagonal entries of the relational pattern were 0.08 if only one latent variable contributed and 0.16 if two latent variables contributed. The selection of the exact sizes as part of a pilot study was guided by the aim to achieve data matrices that could be expected to lead to the same mean CFI when investigated by the hybrid model as when this model was applied to the first set of matrices. Furthermore, the continuous data generated this way were dichotomized into zeros and ones in such a way that the entries of the columns of these matrices showed the same probabilities of one as the columns of the matrices of the first set referred to as *time-limit data*.

Both sets of matrices were investigated by means of the tp-specific model and the hybrid model that were also used in the second study. The tp-specific model was applied repeatedly using different turning points of the logistic function. The CFI results were recorded and averaged for each turning point.

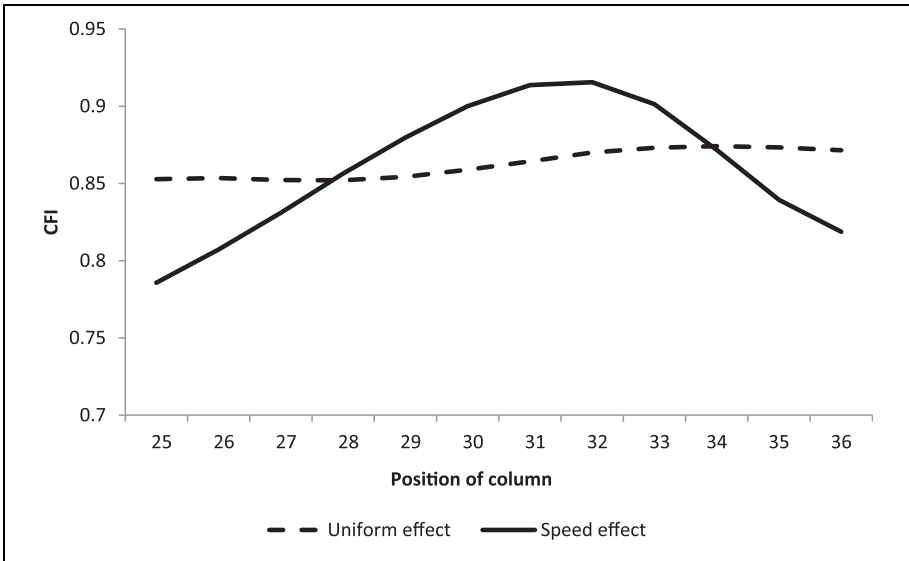


Figure 3. Curves based on the comparative fit index (CFI) results observed in simulated data showing a strong time-limit effect and data showing a uniform effect by means of the tp-specific model ($N = 500$).

Results

The CFI results obtained for the tp-specific model varied as a function of the turning point. The largest CFIs surmounted the .90 barrier for acceptable model fit.

Figure 3 illustrates these results as curves. The solid line resulting from the investigation of the time-limit data (first set of matrices) showed the expected inverse of a u-shaped course. It should be noted, however, that the largest CFIs were found for the turning points of 31 and 32 but not for 30, as was expected because of the turning point for generating data showing the strong effect.

The application of the tp-specific model to the uniform-effect data (second set of matrices) led to the dashed line. This line was not completely flat but did also not show a distinct maximum so that it could not be interpreted as course reflecting the effect of a time limit. The two lines showed distinctly different shapes.

Furthermore, the largest CFIs obtained by means of the tp-specific model differed. It was .92 ($SD = .02$) in the case of the time-limit data and .87 ($SD = .05$) in the case of the data showing the uniform effect due to a common source. There was a CFI difference of .04 that indicated significance.

The hybrid model led to a CFI of .91 ($SD = .02$) in the case of the time-limit data (first set of matrices) and a CFI of .92 ($SD = .05$) in the case of the uniform-effect data (second set of matrices). The CFI difference was marginally substantial, but the CFI distributions showed a large overlap.

Discussion

The discrimination between the effect due to a time limit and the uniform effect due to a source equally contributing to a subset of columns was in the focus of this study. As expected, the *tp*-specific model discriminated between the effects, as was obvious in the CFI difference and the different shapes of the CFI courses obtained by systematically changing the turning point of the logistic function. These results suggested that the *tp*-specific model was suitable for the identification of the effect due to a time limit if the effect was strong.

The hybrid model was not expected to discriminate between the effects. Contrary to this expectation the CFI difference between the CFIs for the two types of matrices reached marginal significance in favor of the matrices showing the uniform effect. This CFI difference simply indicated that the two relational patterns used for data generation were not completely equivalent with respect to CFI although it was the aim to generate two sets of data leading to the same CFI result when investigated by means of the hybrid model.

General Discussion

A time limit for testing leads to a typical method effect caused by the assessment procedure. The consequences of this effect have mainly been investigated within the framework of item response theory. A major finding of this research is that a time limit impairs the validity of measurement (Lu & Sireci, 2007). Such impairment should also be apparent in factor-analytic investigations of the data if there is a similar sensitivity for the effect. The observed results are partly in line with this expectation. All considered fit indices change as a consequence of an increase of the effect due to a time limit. However, only CFI (Bentler, 1990) and NNFI (Bentler & Bonett, 1980) show sensitivity in the sense that model fit changes from good to bad. Other fit indices indicate good model fit even if there is a strong effect due to a time limit or bad model fit although there is no such effect, as in the case of NFI.

The way of treating omissions due to a time limit is guided by the idea that omissions are the result of a second underlying source of responding. The idea of a second source of responding systematically influencing the outcome of assessment is already well established. For example, the multitrait–multimethod approach proposed by Campbell and Fiske (1959) suggests an influence of the observational method on responding besides the influence due to the construct represented by the scale. Furthermore, there is the approach of the model-based imputation methods that considers a second source of responding (Holman & Glas, 2005; O’Muircheartaigh & Moustaki, 1999). Since this second source is considered as unknown, the approach shows broad applicability. In contrast, the way of treating omissions in the three reported studies includes the assumption of a very specific source. The omissions are assumed to originate from the coaction of an insufficient time limit and processing speed. It enables the design of a representation that is especially sensitive to the effect due to the time limit.

An advantage of this representation is that it can be adapted to the actual degree of the effect. Such adaptability is an advantage since it is not a stable effect. Instead the effect depends on several influences, and some of them are independent of the time limit (Oshima, 1994). An important influence appears to be the participants' processing speed. A stronger effect can be expected in an overall slow sample than in an overall fast sample. As a consequence, in older participants the effect may be larger than in younger participants because older participants tend to be slower (Salthouse, 2000). The special adaptability to the actual degree of the effect contributes to the usefulness of this way of investigating the effect due to a time limit.

The integration of the representation of the source of the effect due to a time limit into the model of measurement gives rise to the expectation that the impairment in model fit disappears. The results of the second and third study are only partly in line with this expectation. Despite large improvements of model fit, the reached CFIs are below the CFIs for data without an effect. Since the models of measurement reflect the model used for data generation, the models of measurement can be ruled out as reason for the insufficient improvement. Another possible reason is the fit index selected for the second and third studies, the CFI (Bentler, 1990). It appears to be useful because of its sensitivity for the effect due to a time limit. But it may also own a characteristic that is disadvantageous for the present purpose, as is suggested by the comparison of the results of the second and third studies. The maximum percentage of omissions due to the strong effect was 98% in the second study and 96% in the third study. The change of the percentage was accompanied by the change from bad model fit to acceptable but not yet good model fit. The association of the percentages and fit results suggests that CFI may also be sensitive to very high or very low probabilities of binary responses.

Although the tp-specific and hybrid models perform equally well regarding model fit in data showing the effect due to a time limit, they differ regarding the identification of this effect. The tp-specific model shows a considerably better fit to data showing this effect than to uniform-effect data. In contrast, a small preference in favor of detecting the uniform effect is obvious for the hybrid model that, however, may simply reflect a difference originating from data generation rather than different sensitivities. The virtual correspondence of the fit results for the two models suggests that the hybrid model may be the option for data analysis if the correct identification of the turning point of the function underlying the effect is not necessary.

Finally, the generalizability of the results from simulated data to real data needs to be addressed. Assumptions guided the generation and investigation of simulated data. Although these assumptions are based on empirical grounds, demonstrations using real data are necessary to assure generalizability. There is the possibility that in real data the normality assumption regarding the distribution of processing speed is violated because the sample may not be representative of the population that is assumed to show the normal distribution. It is an open question whether the logistic function used for generating the primary constraints for factor loadings performs in samples that are very homogeneous or quite heterogeneous as well as in samples that show the normal distribution.

So far there seem to be only two empirical studies using the described method. In the first study, the representation of the effect by means of the logistic function improved model fit in investigating reasoning data (Schweizer & Ren, 2013). In the second study, the representation of the effect was used for demonstrating that processing speed contributed to the relationship of cognitive ability and working memory (Ren, Wang, Sun, Deng, & Schweizer, 2017). More studies are needed to overcome this limitation.

Another limitation is the design of the generated data matrices. The design of these matrices mirrors a specific well-known reasoning scale. This design includes enough items for both reasoning and processing speed to unfold so that there was a good chance to be identified. The lack of the variation of the design is justified by the focus on the effect that was varied and on the three objectives. However, there are achievement scales including smaller numbers of items where it may be more difficult to separate the influences of reasoning and processing speed. Furthermore, the impact of reasoning on performance may show a moderating influence on the detectability of the contribution of processing speed to performance. These are limitations that need to be addressed in future research.

Regarding directions for future research we like to highlight that the effect due to a time limit in testing is an important issue of assessment. It is insufficient just to flag tests as speeded or unspeeded since speededness is the result of several sources (Oshima, 1994) and the influence of these sources may vary from one application to another one. This means that there is virtually always some chance that in investigating reasoning data and eventually also some other performance data the model fit is impaired due to a time limit in testing. Therefore, it appears to be advisable to check data originating from this content area in applied research for the presence of an effect due to a time limit in testing. Another possible direction for future research is working out guidelines for tolerable impairments of model fit due to the effect due to a time limit in testing because the observed impairment is the result of the effect but not because of an impaired representation of the construct underlying the scale. Furthermore, it might be explored whether the type of models used for investigating the data showing the effect due to a time limit can be useful for obtaining estimates of the participants' reasoning abilities that are unimpaired by this effect.


Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Stefan Troche  <https://orcid.org/0000-0002-0961-1081>

References

- Bandalos, D. L., & Gerstner, J. J. (2016). Using factor analysis in test construction. In C. DiStefano (Eds.), *Principles and methods of test construction* (pp. 26-51). Göttingen, Germany: Hogrefe.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238-246. doi:10.1037/0033-2909.107.2.238
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588-606. doi:10.1037/0033-2909.88.3.588
- Bolt, D. M., Cohen, A. S., & Wolack, J. A. (2002). Item parameter estimation under conditions of test speededness: Application of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement*, *39*, 331-348.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105. doi:10.1037/h0046016
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*, 464. doi:10.1080/10705510701301834
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*, 233-255. doi:10.1207/S15328007SEM0902_5
- De Ayala, R. J., Plake, B. S., & Impara, J. C. (2001). The impact of omitted responses on the accuracy of ability estimation in item response theory. *Journal of Educational Measurement*, *38*, 213-234. doi:10.1111/j.1745-3984.2001.tb01124.x
- DiStefano, C. (2016). Examining fit with structural equation models. In C. DiStefano (Eds.), *Principles and methods of test construction* (pp. 26-51). Göttingen, Germany: Hogrefe.
- DiStefano, C., & Motl, R. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, *13*, 440-464. doi:10.1207/s15328007sem1303_6
- Finch, H. (2008). Estimation of item response theory parameters in the presence of missing data. *Journal of Educational Measurement*, *45*, 225-245. doi:10.1111/j.1745-3984.2008.00062.x
- Fischer, H. (2011). *A history of the central limit theorem*. New York, NY: Springer.
- Goegebeur, Y., De Boeck, P., Wollack, J. A., & Cohen, A. S. (2008). A speeded item response model with gradual process change. *Psychometrika*, *73*, 65-87. doi:10.1007/s11336-007-9031-2
- Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability. *Educational and Psychological Measurement*, *66*, 930-944. doi:10.1177/0013164406288165
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549-576. doi:10.1146/annurev.psych.58.110405.085530
- Gulliksen, H. (1950). Speed versus power tests. In H. Gulliksen (Ed.), *Theory of mental tests* (pp. 230-244). New York, NY: Wiley.
- Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M., & Mumford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communalities, and overdetermination. *Educational and Psychological Measurement*, *65*, 202-226. doi:10.1177/0013164404267287
- Hohensinn, C., & Kubinger, K. D. (2011). Applying item response theory methods to examine the impact of different response formats. *Educational and Psychological Measurement*, *71*, 732-746. doi:10.1177/0013164410390032

- Holman, R., & Glas, C. A. W. (2005). Modelling non-ignorable missing-data mechanisms with item response theory models. *British Journal of Mathematical and Statistical Psychology*, 58, 1-17. doi:10.1111/j.2044-8317.2005.tb00312.x
- 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. doi: 10.1080/10705519909540118
- Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Oxford, England: Elsevier.
- Jöreskog, K. G. (1970). A general method for analysis of covariance structure. *Biometrika*, 57, 239-257. doi:10.2307/2334833
- 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.
- Jöreskog, K. G., & Sörbom, D. (2006). *LISREL 8.80*. Lincolnwood, IL: Scientific Software International.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.
- Kubinger, K. (2016). Adaptive testing. In C. DiStefano (Eds.), *Principles and methods of test construction. Standards and recent advances* (pp. 104-119). Göttingen, Germany: Hogrefe.
- Lord, F. M. (1965). A note on the normal ogive or logistic curve in item analysis. *Psychometrika*, 30, 371-372. doi:10.1007/BF02289500
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Lu, Y., & Sireci, S. G. (2007). Validity issues in test speededness. *Educational Measurement*, 26, 29-37. doi:10.1111/j.1745-3992.2007.00106.x
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84-99. doi:10.1037/1082-989X.4.1.84
- McCullagh, P., & Nelder, J. A. (1985). *Generalized linear models*. London, England: Chapman & Hall.
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37, 1-10. doi: 10.1016/j.intell.2008.08.004
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105, 156-166. doi:10.1037/0033-2909.105.1.156
- Must, O., & Must, A. (2013). Changes in test-taking patterns over time. *Intelligence*, 41, 780-790. doi:10.1016/j.intell.2013.04.005
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered, categorical, and continuous latent variables indicators. *Psychometrika*, 32, 1-13. doi: 10.1007/BF02294210
- O'Muircheartaigh, C., & Moustaki, I. (1999). Symmetric pattern models: A latent variable approach to item non-response in attitude scales. *Journal of the Royal Statistical Society, Series A: Statistics in Society*, 162, 177-194. doi:10.1111/1467-985X.00129
- Oshima, T. C. (1994). The effect of speededness on parameter estimation in item response theory. *Journal of Educational Measurement*, 31, 200-219. doi:10.1111/j.1745-3984.1994.tb00443.x
- Schweizer, K. (2013). A threshold-free approach to the study of the structure of binary data. *International Journal of Statistics and Probability*, 2, 67-75. doi: 10.5539/ijsp.v2n2p67

- Schweizer, K., & Ren, X. (2013). The position effect in tests with a time limit: The consideration of interruption and working speed. *Psychological Test and Assessment Modeling*, 55, 62-78.
- Schweizer, K., Ren, X., & Wang, T. (2015). A comparison of confirmatory factor analysis of binary data on the basis of tetrachoric correlations and of probability-based covariances: A simulation study. In R. E. Millsap, D. M. Bolt, L. A. van der Ark, & W.-C. Wang (Eds.), *Quantitative psychology research* (pp. 273-292). Heidelberg, Germany: Springer.
- Ren, X., Wang, T., Sun, S., Deng, M., & Schweizer, K. (2017). Speeded testing in the assessment of intelligence gives rise to a speed factor. *Intelligence*, 66, 64-71. doi: 10.1016/j.intell.2017.11004
- Roberts, R. D., & Stankov, L. (1999). Individual differences in speed of mental processing and human cognitive abilities: Towards a taxonomic model. *Learning and Individual Differences*, 11, 1-120. doi:10.1016/S1041-6080(00)80007-2
- Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology*, 54, 35-54. doi:10.1016/S0301-0511(00)00052-1
- van der Linden, W. J., & Xiong, X. (2013). Speededness and adaptive testing. *Journal of Educational and Behavioral Statistics*, 38, 418-438. doi:10.3102/1076998612466143
- 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
- Zeller, F., Krampen, D., Reiß, S., & Schweizer, K. (2016). Do adaptive representations of the item-position effect in APM improve model fit? A simulation study. *Educational and Psychological Measurement*. doi:10.1177/0013164416654946