

Does Rapid Guessing Prevent the Detection of the Effect of a Time Limit in Testing?

Karl Schweizer^{1,2}, Dorothea Krampen¹, Brian F. French³

[1] *Institute of Psychology, Goethe University Frankfurt, Frankfurt, Germany.* [2] *Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou, China.* [3] *Department of Kinesiology and Educational Psychology, Washington State University, Pullman, WA, USA.*

Methodology, 2021, Vol. 17(3), 168–188, <https://doi.org/10.5964/meth.4663>

Received: 2020-11-06 • Accepted: 2021-07-30 • Published (VoR): 2021-09-30

Corresponding Author: Karl Schweizer, Institute of Psychology, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 6, 60323 Frankfurt a. M., Germany. E-mail: K.Schweizer@psych.uni-frankfurt.de

Abstract

Rapid guessing is a test taking strategy recommended for increasing the probability of achieving a high score if a time limit prevents an examinee from responding to all items of a scale. The strategy requires responding quickly and without cognitively processing item details. Although there may be no omitted responses after participants' rapid guessing, an open question remains: do the data show unidimensionality, as is expected for data collected by a scale, or bi-dimensionality characterizing data collected with a time limit in testing, speeded data. To answer this question, we simulated speeded and rapid guessing data and performed confirmatory factor analysis using one-factor and two-factor models. The results revealed that speededness was detectable despite the presence of rapid guessing. However, detection may depend on the number of response options for a given set of items.

Keywords

rapid guessing, test taking behavior, speededness, effect of a time limit in testing, confirmatory factor analysis, fixed-links model

Rapid guessing is a test taking strategy that consists of responding to items fast and without attempting to solve the item properly (Wise, 2017). This strategy enables the completion of a set of items within a very short time span. This behavior is motivated by testing situations that impose a time limit, as is common with many achievement tests. In such an occasion, rapid guessing enables the completion of all so-far not-reached items shortly before the end of testing time. The use of this test taking strategy in combination with a time limit in testing avoids item response omissions that can impair



validity of measurement (Lu & Sireci, 2007). Therefore, it may appear that rapid guessing contributes to valid measurements. However, the presence of rapid guessing response behavior may actually introduce variance not related to the trait being measured. That is, variance due to the participants' intention to respond at random to complete the test. This irrelevant variance may be substantial enough to begin to alter the technical quality of the items and resulting scores.

Another possible consequence is that the irrelevant variance manifests itself as an additional factor in the latent structure of the test. Models for structural investigations (e.g., factor analysis, dimensionality assessment) mostly assume that there is only one latent source of responding that leads to systematic and relevant variation, which is captured by the latent variable included in the measurement model (Graham, 2006). The enlargement of such a model by integrating another latent variable for capturing systematic irrelevant variation due to processing speed as assumed source of omissions because of a time limit can provide an account of speeded data in the absence of rapid guessing (Schweizer, Troche et al., 2019). But it remains unclear whether such an enlarged model can account for data if responses due to participants' intent to respond at random replace omissions. This study reports on an investigation as to whether the influence of speeded data can be detected despite the replacement of omissions by random data. Such data can be expected in speeded testing as compared to power testing, and may even originate from power testing with an ample time limit that is yet insufficient for a subset of participants.

Rapid Guessing for Preventing Omissions

Although participants taking a test are expected to spend as much time as necessary on each item and to provide the best possible response, they may deviate from such behavior for various reasons. For example, there are situations, for example, where test scores will have major consequences (e.g., employment, education opportunities) that may lead participants to use inappropriate test taking strategies when completing items in order to increase the chance of reaching a high score (Stenlund et al., 2018). Furthermore, social desirability may play a role in responding where 1) a participant works to complete all items to be a "good" participant (Vogt & Johnson, 2015) or 2) a participant behaves according to the stereotype of a smart person by completing all items, even if guessing. Moreover, there may be the instruction or recommendation to make use of rapid guessing in test taking that is taught in test coaching courses. Also, the possibility exists that participants use rapid guessing in assessment environments where the consequences of the scores are low (i.e., low-stakes testing) or due to other reasons.

The advantage promised by rapid guessing is that a random response can be correct. If there are several response options and only one is correct, the probability that the random response is correct is one divided by the number of response options. That is, if there are four options and the participant guesses at random, there is a 25% chance

of a correct response compared to not responding at all and ensuring a 0% chance of a correct response. Smaller numbers of response options are associated with larger probabilities of a correct response and larger numbers of response options with smaller probabilities. This strategy could be an advantage for the examinee if the number of correctly completed items serves as measure of performance.

Structural Investigation of Speeded Data

A popular way of investigating the internal structure of a scale to support a scoring inference for validity is with confirmatory factor analysis (CFA). A common assumption in item response theory is unidimensionality, which can, in part, be demonstrated by a one-factor CFA model (Graham, 2006). Such a demonstration confirms that the data are due to one latent source of systematic relevant variation. In this case, a latent variable specified in the CFA measurement model captures the systematic relevant variation due to the latent trait or ability measured by the test or assessment. These measurement models in a CFA framework can be specified with different types of indicators or items (e.g., continuous or dichotomous variables) of the trait measured. For didactical reasons we separate the discussion regarding the factor structure from the discussion regarding modeling different data types. For convenience, we discuss models under the assumption of continuous data in this section for the discussion regarding structure.

A measurement model specifies the influences that are assumed to determine the participants' responses to a given item. A one-factor CFA model assumes one latent source of systematic responding that is reflected by latent variable ξ . The contribution of ξ to completing the i th item ($i = 1, \dots, p$) is quantified by factor loading λ_i . Additionally, assumed contributions are those of random influences that are represented by δ_i without further specification (e.g., no correlated residuals). Such a model relates the $p \times 1$ vector of manifest variables \mathbf{x} to the sum composed of the product of the $p \times 1$ vector of factor loadings λ of the manifest variables on the latent variable and latent variable ξ on one hand and of the $p \times 1$ vector δ of random variables on the other hand:

$$\mathbf{x} = \lambda_{\text{latent_source}} \xi_{\text{latent_source}} + \delta. \quad (1)$$

A scale is said to show structural validity if this model accounts for the item covariance matrix. However, this is not general or all validity but validity restricted to major characteristics of the circumstances of data collection. One major characteristic is the time span for completing the items of the scale, as time limits in testing can alter the validity of the data (Lu & Sireci, 2007). In the case of a time limit in testing that prevents participants from completing all items, the data are not only due to the latent source and random influences but also due to latent processing speed (Partchev et al., 2013). The influence of latent speed even appears to increase with increasing age including adulthood (Borter et al., 2020). A lack of latent processing speed can lead to omissions, as is demonstrated by

comparing the outcomes for models representing different latent sources of responding in investigating reasoning data (Schweizer, Reiß et al., 2019).

A modified CFA model of Equation 1 is necessary in order to account for systematic variation that is due to latent processing speed. Since latent processing speed is to be considered as another latent source, the variation due to latent processing speed needs to be captured by another latent variable. The necessity to consider a second latent variable creates circumstances comparable to the circumstances leading to multitrait-multimethod models (Byrne, 2016), and this irrelevant variance may be seen as a method effect. While in the standard CFA measurement model a manifest variable shows one factor loadings only, manifest variables of multitrait-multimethod models have cross-loadings. The modified model is a two-factor model with two latent variables, which we address as the primary latent variable and an additional latent variable ξ_{primary} and $\xi_{\text{additional}}$, respectively. Together they explain the manifest variables of the $p \times 1$ vector \mathbf{x} :

$$\mathbf{x} = \lambda_{\text{primary}}\xi_{\text{primary}} + \lambda_{\text{additional}}\xi_{\text{additional}} + \delta \quad (2)$$

(Schweizer, Troche et al., 2019). The two $p \times 1$ vectors λ_{primary} and $\lambda_{\text{additional}}$ include the factor loadings and the $p \times 1$ vector δ the error variables. The first component of the sum represents the construct measured by the scale and the second one the influence of processing speed.

There are two different types of two-factor models. The first type combines two one-factor models into a whole. The major characteristic of this type is that each manifest variable (e.g., item) loads on one latent variable only so that there are no cross-loadings (Kline, 2005, p. 175). The other type of a two-factor model allows the manifest variables (e.g., items) to cross-load. One version of this model is a bifactor model that comprises a general latent variable and a specific latent variable (Reise, 2012). The specific latent variable differs from the general latent variable in that it receives factor loadings from a subset of items only, and general and specific latent variables are orthogonal. This model enables the capturing of the systematic variation due to a general source and a specific source. The other version of this type of model is the multitrait-multimethod model (Byrne, 2016). This version has been proposed for investigating data that were collected according to a multitrait-multimethod design.

It is the first version of the second type of model that is suitable for data originating from two latent sources that simultaneously contribute to at least a few items. More specifically, since one source can be assumed to be active in completing all items whereas the second source is only active in some items, it is a bifactor model that is required for investigating data collected with a time limit in testing. This means that all entries of λ_{primary} are either free for estimation or constrained to correspond to expected values whereas some entries of $\lambda_{\text{additional}}$ are fixed to zero. These are the entries regarding items that are not influenced by processing speed, that is, show no omissions:

$$\Lambda = [\lambda_{\text{primary}}, \lambda_{\text{additional}}] = \begin{bmatrix} \lambda_{\text{primary}_1} & 0 \\ \lambda_{\text{primary}_2} & 0 \\ \cdot & 0 \\ \cdot & \lambda_{\text{additional_onset}} \\ \lambda_{\text{primary}_i} & \lambda_{\text{additional}_i} \\ \cdot & \cdot \\ \cdot & \cdot \\ \lambda_{\text{primary}_p} & \lambda_{\text{additional}_p} \end{bmatrix} \quad (3)$$

where the matrix of factor loadings Λ is defined to include λ_{primary} and $\lambda_{\text{additional}}$.

Free factor loadings and factor loadings fixed to correspond to expected values have been shown to perform virtually equally well in simulated data if the expected values are adapted to the assumed latent source and the number of participants selecting the strategy (Schweizer et al., 2020; Schweizer, Troche et al., 2019). Using fixed factor loadings, it is necessary to free the associated variance parameter $\varphi_{\text{additional}}$ for estimation. These types of factor loadings have different properties. Free factor loadings can accommodate all kinds of effects so that there is hardly any impairment in model fit. This means that the factor loadings on the latent variable account for the systematic variation due to the intended latent source and in addition to some degree accommodate systematic variation due to other sources. In contrast, fixed factor loadings can only account for the systematic variation due to the intended latent source. If there is further systematic variation that may be due to a method effect, this leads to model misfit. The greater probability of model misfit may be considered as a downside of fixed factor loadings but there is also an advantage: good model fit indicates that the model captures exactly what it is expected to capture and nothing else.

Values for serving as factor loadings in order to capture systematic variation due to processing speed can be obtained by the cumulative normal distribution function that is approximated by the logistic function. The cumulative normal distribution function is obtained from the normal distribution function that is assumed to characterize the density distribution of latent processing speed. Using the logistic function, the factor loading of the i th item ($i = 1, \dots, p$) on the latent variable representing latent processing speed λ_i is defined as follows:

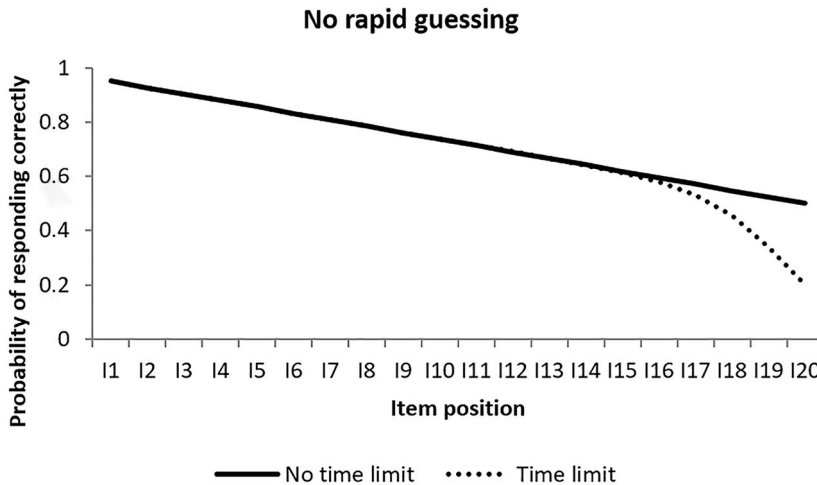
$$\lambda_{\text{additional}_i} = \frac{e^{i-tp}}{1 + e^{i-tp}} \quad (4)$$

where tp refers to the turning point of the logistic function. It depends on the time limit used in data collection and needs to be adapted to the distribution of omissions, that is, the time limit.

Figure 1 illustrates how a time limit in testing modifies the probability of responding correctly.

Figure 1

Illustration of the Probabilities of Responding Correctly Without and With a Time Limit in Testing



The curve printed as a solid line illustrates the assumed probabilities of a correct response if there is no time limit in testing. This curve suggests that the items are arranged according to their difficulty levels. The curve printed as a dashed line represents the assumed probabilities of a correct response originating from testing with a time limit. The assumed gradual drop-off of participants causes an increasing degree of deviation toward the end of the sequence of items.

Effect of Rapid Guessing on Investigating the Latent Structure

There may be consequences of rapid guessing different than leaving the not-reached items as omitted, and begins with the expected distribution of omissions. This distribution needs to be modified to take into consideration that correct and incorrect responses at random replace omissions. For this purpose, a clearly defined expected probability of a correct response at random that is independent of the difficulty level of the item is necessary. We assumed that the data were collected with items showing a multiple-choice response format in order to have a basis for such probabilities. In this case, the expected probability of a correct response at random solely depends on the number of response options. The multiple-choice response format is the most popular response format (Johnson & Morgan, 2016). Correct responses at random obtained by

this response format are not only an issue with respect to rapid guessing, but also a general issue of assessment (Drasgow & Mattern, 2006).

Since the logistic function varies between zero and one and is assumed to provide values corresponding to the expected frequency of omissions divided by the upper limit for the frequency of omissions, it can be perceived as probability. Accordingly, in the following discussion we use probabilities for combining the description of the effect of a time limit with the description of the effect of rapid guessing. The expected probability $E[\Pr(\cdot)]$ for X_i ($i = 1, \dots, p$) to be an omission is described with respect to the set of omissions C_o . To keep this section connected to the previous discussion, we start from Equation 4:

$$\lambda_{\text{additional}_i} = \frac{e^{j-tp}}{1 + e^{j-tp}} = E[\Pr(X_i \in C_o)]. \quad (5)$$

Next, the influence of rapid guessing needs to be quantified. Rapid guessing means that X_i ($i = 1, \dots, p$) can be perceived as taken either from the set of correct responses C_c or the set of false (= incorrect) responses C_f . The expected probability depends on the number of response options. If we assume that this number is f , the expected probability of a correct response due to rapid guessing is given by

$$E[\Pr(X_i \in C_c)] = \frac{1}{f} \quad (6)$$

and the expected probability of an incorrect response by

$$E[\Pr(X_i \in C_f)] = \frac{f-1}{f}. \quad (7)$$

The majority of correct responses can be assumed to originate from the primary source of responding whereas omissions turned into incorrect responses are more likely than omissions turned into correct responses at random. This suggests that the focus has to be on the incorrect responses in quantifying the effect of rapid guessing on the detection of speededness. Accordingly, the expected probability of an incorrect response due to latent processing speed in combination with rapid guessing is given by

$$E\{\Pr[(X_i \in C_o) \cap (X_i \in C_f)]\} = \frac{e^{j-tp}}{1 + e^{j-tp}} \times \frac{f-1}{f}. \quad (8)$$

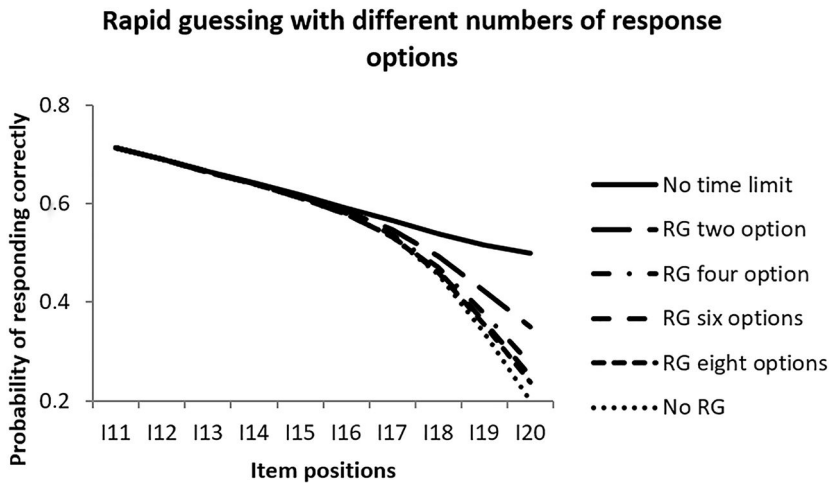
Since the expected probabilities of Equation 8 can be assumed to reflect the distribution of latent processing speed under the condition of rapid guessing, it is justified to employ it for achieving values for the fixation of factor loadings:

$$\lambda_{\text{additional_latent_source_}i} = \frac{e^{i-tp}}{1 + e^{i-tp}} \times \frac{f-1}{f}. \quad (9)$$

Figure 2 provides a graphical representation of how rapid guessing modifies the probability of a correct response.

Figure 2

Illustration of the Probabilities of Responding Correctly if There is Rapid Guessing (RG) in Combination With Response Formats Including Two, Four, Six and Eight Response Options



This Figure includes curves depicting the probability of responding correctly if participants use rapid guessing in combination with response formats including two, four, six and eight options. The curves suggest that eight, six and even four response options only cause minor deviations from the curve for *no* rapid guessing.

Equations 8 and 9 enable a first evaluation of the consequences of rapid guessing for the detection of the effect of speeded testing. The comparison of Equation 9 and Equation 4 suggests that there is a decrease of the systematic variation due to latent processing speed. The decrease of systematic variation also means a reduced probability of detecting it. Further, the comparison of Equation 8 and Equation 5 reveals that there is a decrease in the probability of detecting the effect of speeded testing. The decrease depends on the number of response options. Therefore, we state the following hypothesis for the empirical investigation: the larger the number of response options, the larger the probability of the detection of rapid guessing.

Despite the indicated impairment of the detectability of the effect of a time limit in testing there is also positive news: there is still some chance to detect this effect.

Further, the larger the number of response options, the larger is the probability to detect it. Concerns about the effect of the number of response options leads [Johnson and Morgan \(2016\)](#) to recommend three or more response options in constructing multiple-choice items. Three response options are reported to be most common in applications ([Rodriguez, 2005](#)).

Analytic Strategy

Confirmatory factor analysis attempts to estimate a model that can reproduce the covariance matrix. This involves comparing the model-implied $p \times p$ covariance matrix Σ with the $p \times p$ empirical covariance matrix \mathbf{S} by means of a discrepancy function. Good model fit is considered as confirmation of the specified measurement model ([Graham, 2006](#)) that gives rise to Σ , and model misfit as its rejection. Since factor analysis is mostly conducted according to a measurement model that includes continuous and normally distributed variables, the appropriateness for investigating dichotomous data that are considered in the present work may be called into question ([Kubinger, 2003](#)).

Mathematics offers several solutions for relating different types of data to each other. There are link transformations as part of generalized linear models ([McCullagh & Nelder 1985](#); [Skrondal & Rabe-Hesketh, 2004](#)), variance stabilizing methods ([Guan, 2009](#); [Morgenthaler & Staudte, 2012](#)) and methods that are specific for factor analysis. Methods specific for factor analysis include the transformation of dichotomous data into frequencies or probabilities that are considered as continuous. This important step turns binary information into continuous information. Further transformations leads to tetrachoric correlations or probability-based covariances that serve as input to factor analysis ([Schweizer et al., 2015](#)). Whereas the computation of tetrachoric correlations includes the computation and use of thresholds that eliminates the effect of splitting data according to probability level p in dichotomization, probability-based covariances still include the effect of p that may be considered as a reason for observing spurious factors ([Kubinger, 2003](#)).

The model-implied $p \times p$ covariance matrix Σ is defined as

$$\Sigma = \Lambda\Phi\Lambda' + \Theta \quad (10)$$

where Λ (and its transpose Λ') is the $p \times q$ matrix of factor loadings, Φ the $p \times q$ matrix of variances and covariances of factors, and Θ the $p \times p$ diagonal matrix of error variances. In the case of one factor only and centered data it reflects the standard model of measurement of confirmatory factor analysis that is defined as

$$\mathbf{x} = \boldsymbol{\lambda} \times \xi + \boldsymbol{\delta} \quad (11)$$

(see also [Equation 1](#)). This model includes $p \times 1$ vector of manifest variables \mathbf{x} , $p \times 1$ vector of factor loadings $\boldsymbol{\lambda}$, latent variable ξ and $p \times 1$ vector of error variables $\boldsymbol{\delta}$. In the case of

a model that takes speed and rapid guessing into consideration additional latent variable $\xi_{\text{speed\&guessing}}$ needs to be considered besides latent variable $\xi_{\text{construct}}$. Furthermore, there are two $p \times 1$ vectors of factor loadings $\lambda_{\text{construct}}$ and $\lambda_{\text{speed\&guessing}}$. The factor loadings of $\xi_{\text{speed\&guessing}}$ corresponds to $\lambda_{\text{additional_latent_source}_i}$ ($i = 1, \dots, p$) of Equation 9. This enlargement of the model of measurement represented by Equation 11 gives

$$\mathbf{x} = \lambda_{\text{construct}} \times \xi_{\text{construct}} + \lambda_{\text{speed\&guessing}} \times \xi_{\text{speed\&guessing}} + \delta \quad (12)$$

(see also Equation 2).

Using this model in combination with dichotomous data requires adaptation that is two-fold in the approach characterizing this work. First, there is adaptation of the scale level of data that occurs in computing probability-based covariances that changes from binary to continuous (see the paragraph preceding the previous paragraph). We symbolize this adaptation by transformation T of \mathbf{x} so that the manifest part of the model becomes $T_{\text{probability-based covariance}}(\mathbf{x})$. But there is still dependency on p used in dichotomization. Second, this dependency is removed by an additional transformation that is realized as weighting. The item-specific weight w_i ($i=1, \dots, p$) is defined as

$$w_i = \sqrt{\Pr(X_i = 1) \times [1 - \Pr(X_i = 1)]} \quad (13)$$

where $\Pr(X_i = 1)$ represents the probability of the correct response in completing item i (Zeller et al., 2017). For integrating the weights into Equation 12, they are assigned to the main diagonal of $p \times q$ diagonal matrix of weights \mathbf{W} . Finally, the model adapted to the characteristics of dichotomous data is given by

$$T_{\text{probability-based covariance}}(\mathbf{x}) = (\mathbf{W} \times \lambda_{\text{construct}}) \times \xi_{\text{construct}} + (\mathbf{W} \times \lambda_{\text{speed\&guessing}}) \times \xi_{\text{speed\&guessing}} + \delta. \quad (14)$$

Note, if the focus of the investigation is on the effect of latent speed, weights on $\lambda_{\text{construct}}$ are only necessary in combination with fixed factor loadings on this factor. Otherwise, they can be omitted since their omission does not influence model fit (Schweizer et al., 2018).

The corresponding model-implied $p \times p$ covariance matrix Σ is a matrix that is prepared for probability-based covariances as input:

$$\Sigma_{\text{probability-based covariances}} = (\mathbf{W}\Lambda)\Phi(\mathbf{W}\Lambda)' + \Theta. \quad (15)$$

Note, implicitly it is assumed that dichotomization influences the systematic part of the model of measurement but not the error part. Error is assumed to always follow a normal distribution.

The correctness of the model-implied $p \times p$ covariance matrix Σ specified according to Equation 15 for reproducing the $p \times p$ empirical covariance matrix S can be investigated by the maximum likelihood estimation method. This method maximizes the likelihood of the parameters of the model with respect to the data. For this purpose function F is minimized:

$$F = \log\|\Sigma\| + \text{tr}(\Sigma S^{-1}) - \log\|S\| - p + q \quad (16)$$

(Jöreskog & Sörbom, 2006). Function F shows the following characteristic: the more similar Σ and S , the smaller is the difference between $\log\|\Sigma\|$ and $\log\|S\|$ on one hand and the more similar is the product ΣS^{-1} to the identity matrix on the other hand. In the case of perfect correspondence trace $\text{tr}(\Sigma S^{-1})$ corresponds to p so that q remains. F is used for investigating the correctness of models because an incorrect model is likely to lead to a large value so that the comparison with the χ^2 distribution signifies model misfit (except of in cases where there are more factors included in the model than necessary).

The preconditions for making use of function F are continuous data, invertibility of Σ and positive definiteness whereas there is no precondition regarding distribution. Yet, in the application of F for comparing S and Σ there is restriction regarding the distribution of data that originates from Σ . The variables of this model are assumed to follow the normal distribution and are treated as such. Skewness is a deviation from normality that has been demonstrated to lead to model misfit (Lai, 2018). Correction methods and estimation methods have been developed that aim at overcoming such deviation of data from normality.

Our approach differs from the available data-focused approaches in that it seeks to modify the model in such a way that model and data correspond to each other according to major distributional properties. This means that it makes use of the characteristic of the maximum likelihood estimation function of no restriction regarding the distribution. The factor loadings are modified by multiplication with weights in such a way that the effect of splitting continuous data according to probability level p in dichotomization is compensated. This is an important precondition for the correct reproduction of the entries of S computed from dichotomous data. In the following we prefer to refer to our approach as ML-MA (model-adapted ML).

Objectives

The main objective of the empirical investigation was to examine if the effect of a time limit in testing was detectable in data despite participants' rapid guessing. The use of this guessing strategy was an important issue as its strict application would result in the complete disappearance of omissions. Complete disappearance of omissions meant that the effect of a time limit in testing was no longer apparent in descriptive statistics.

The simulated data for this investigation had to show 1) the characteristics of data originating from a time limit situation in testing leading to omissions, 2) the use of a multiple-choice response formats, and 3) rapid guessing. The selected time limit was assumed to allow all participants to complete approximately two-thirds of the 20-item set before they would gradually stop responding properly. Furthermore, the data had to show the consequence of the participants' rapid guessing. For this purpose, the simulated omissions due to the testing time limit were replaced by simulated random responses.

Method

Data

Data Generation

Data matrices composed of 500 rows and 20 columns were generated by means of three 20×20 relational patterns (Jöreskog & Sörbom, 2001). These patterns showed off-diagonal entries that could be reproduced by factor loadings of 0.325, 0.375 and 0.425 of a one-factor model; we referred to them as low, medium, and high levels. As they reflected the influence of the primary latent source of responding to be used for controlling effects, we refer to them as *levels of source influence*. The diagonal elements of the relational patterns were set equal to one. Each one of the three relational patterns served the generation of 500 matrices of continuous and normally distributed random data $[N(0,1)]$. In the following, we refer to rows of the matrices as simulated participants and to the columns as simulated items.

Establishment of Data Characteristics

The continuous data were dichotomized so that the first simulated item showed a simulated probability of a correct response of .95 and the last simulated item of .50. The simulated probabilities of the simulated items in-between linearly decreased. Furthermore, omissions were integrated into the data matrices using the logistic function. That is for each simulated item (= column) the percentage of simulated participants (=rows) who were expected to be unable to respond within the available time span was determined by the logistic function. After the selection of a simulated participant the entries to this and all following simulated items were turned into omissions. The turning point that marks the switch from the increase in steepness to the decrease of steepness of the logistic function was set to item 18 (Schweizer, Wang et al., 2020).

Simulation of Response Format

The omissions were replaced by random data (correct responses or incorrect responses at random), as could be expected because of rapid guessing. Because of the crucial influence of the number of response options different multiple-choice response formats

were considered. Eight, six, four and two response options were selected for this study. The corresponding probabilities of a correct response at random were 1/8, 1/6, 1/4 and 1/2 respectively. They served the investigation of the hypothesis regarding the number of response option (see end of the theoretical section). Furthermore, no replacement of omissions, that is, no rapid guessing, was also considered in order to have a comparison level. Altogether, there were 500×3 (source influence levels) $\times 5$ (response option levels) matrices.

Models

The confirmatory factor models included either one or two latent variables (=factors). One of them was designed to capture systematic variation due to the primary source of responding and the other one to capture systematic variation due to the additional source that was assumed to be latent processing speed. The latent variables were not allowed to correlate with each other. Furthermore, there were 20 manifest variables. The constraints for the factor loadings on the latent variable representing the effect of the time limit in testing were obtained according [Equation 4](#) or [Equation 9](#). The variance parameters of the latent variable with fixed loadings were set free. The factor loadings on the substantive latent variable were freely estimated while the corresponding variance parameter was fixed to one. The input to confirmatory factor analysis was achieved by computing probability-based covariances.

Design

The study included the type of model (one-factor vs two-factor models) as main independent variable, the response formats as a minor latent variable and the levels of source influence as control variable. The dependent variable was model fit measured by CFI.

Statistical Investigation

The statistical investigation was conducted using ML-MA version of maximum likelihood estimation (Jöreskog & Sörbom, 2006). Only the Comparative-fit-index (CFI; see DiStefano, 2016; Peterson et al., 2020) was considered for model evaluation in this study since this index was found to be especially sensitive to the effect of a time limit in testing (Schweizer, Troche et al., 2019). Models were compared by means of the CFI and χ^2 differences; where a difference of .01 could be considered as substantial regarding the CFI difference (Cheung & Rensvold, 2002) and of 3.84 regarding the χ^2 difference ($df = 1$). Since each condition required analyzing 500 matrices, there were 500 estimates of model fit for each condition. The mean fit results were evaluated and compared.

Results

The mean CFI results observed for the one-factor and two-factor CFA models are presented as bars in Figure 3. The first set of three bars provides the results for data showing no effect of a time limit in testing, that is, the data were complete and without rapid guessing. The mean CFIs depicted by these bars were larger or close to .95 that indicated good or marginally good model fit, by standard criteria for fit. The following five sets of three bars picture the results obtained by the one-factor confirmatory factor model for different probabilities of a correct response at random. There was a monotonic CFI increase that was associated with an increase in the probability of a correct response at random; that is, a reduction of the number of response options. Whereas model misfit was indicated for no random responses, good model fit was signified for the probability of .5 (that means two response options) in two levels of source influence. The sets of bars on the right side of the Figure depict the results observed by the two-factor confirmatory factor model. As suggested by the sizes of the bars, the model fit was always good for two levels of source influence and the results for the third level were close to .95. In each set of three neighboring bars, the CFIs showed dependency on the levels of source influence. Always the highest level led to the largest CFI and the smallest level to the lowest CFI.

Figure 3

Mean CFI Results as Bars for the Different Levels of Source Influence and Probabilities of a Correct Response at Random Achieved by Investigating Data Constructed to Include Rapid Guessing Using One-Factor and Two-Factor Confirmatory Factor Models

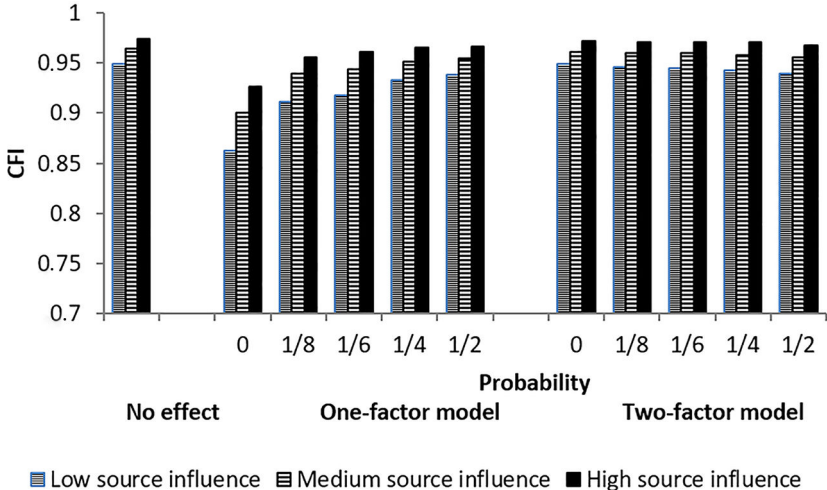


Table 1 comprises the CFI differences between the CFI results of the corresponding one-factor and two-factor confirmatory factor models.

Table 1

Differences Between the CFIs of the One-Factor and Two-Factor Models Observed for Rapid Guessing

Source influence level	CFI differences for the following probabilities of a correct response				
	0 ^a	1/8 ^b	1/6 ^b	1/4 ^b	1/2 ^b
.325	0.086*	0.035*	0.027*	0.010*	0.001
.375	0.061*	0.021*	0.016*	0.007	0.001
.425	0.045*	0.015*	0.010*	0.006	0.001

^aComparison level. There was no replacement of omissions. ^bThe probability of a correct response due to chance (instead of an omission).

* $p < .05$ (according to Cheung & Rensvold, 2002).

The columns of the table refer to the probability levels (numbers of response options) and the rows to the source influence levels. All differences of the first to third columns were larger than or equal to .01. These result signified a substantial improvement in model fit from the one-factor to the two-factor models for the probabilities of zero, 1/8 and 1/6; that is, for no replacement of omissions and response formats with eight and six response options. In the fourth column there was only one other substantial difference for the lowest source influence level. Not one of the differences reported in the last column reached the level of statistical significance.

The χ^2 differences for the corresponding one-factor and two-factor confirmatory factor models are included in Table 2

Table 2

Differences Between the χ^2 s of the One-Factor and Two-Factor Models Observed for Rapid Guessing

Source influence level	χ^2 differences for the following probabilities of a correct response				
	0 ^a	1/8 ^b	1/6 ^b	1/4 ^b	1/2 ^b
.325	32.5 ^{c*}	16.7*	13.6*	8.0*	2.6
.375	33.4 ^{c*}	18.7*	15.1*	10.0*	2.8
.425	33.5 ^{c*}	20.9*	17.1*	11.3*	3.2

^aComparison level. There was no replacement of omissions due to rapid guessing. ^bThe probability of a correct response due to chance (instead of an omission). ^cSince the two-factor model with free factor loadings on the first factor led in a large number of datasets to estimation problems in this condition, the factor loadings on this factor were fixed to one.

* $p < .05$.

This table shows the same structure as [Table 1](#). The χ^2 results were in line with the CFI results with two exceptions. The first exception was the result for the combination of the source level of .375 and the probability of a correct response of 1/4 and the second exception the result for the combination of the source level of .425 and the same probability of a correct response. In both cases the χ^2 difference signified a substantial difference whereas the CFI difference did not.

In sum, an effect of a time limit in testing was detected if there were no less than either six response options (CFI difference) or four response option (χ^2 difference), or expressed in a different way, if the probability of a correct response at random was not larger than .167 (or .25) with one exception.

Discussion

Accurate data are the precondition for the achievement of new insights in science; the control of sources that potentially impair data is an important part of scientific research. A long known issue regarding measurement validity is a time limit in testing ([Lu & Sireci, 2007](#); [Wise, 2017](#)). A time limit is likely to prevent some or even all participants from completing all items so that omissions are inevitable. This means that the validity of measurement not only depends on the structural validity of the scale but also on the time limit in testing. Data collected with a time limit in testing are likely to lead to model misfit in an investigation of the structural validity of the scale by standard confirmatory factor analysis. Model misfit is normally interpreted as indication of the lack of unidimensionality or some specified structure, although in this case the true reason of the failure is the time limit in testing.

A time limit in testing creates a special precondition for the statistical investigation of internal structure that requires adaption of the factor model. The special precondition is that two sources of responding need to be considered instead of only one ([Partchev et al., 2013](#)). There are further preconditions that are of importance but need not to be considered as long as conventions hold. An important convention is that the participants behave as expected. They are expected to thoroughly process the items of the scale successively. Scale instruction and item text play a key role in invoking this convention ([Johnson & Morgan, 2016](#)). Statistical investigations by a two-factor measurement model can be expected to yield accurate results despite a time limit in testing as long as the participants' behavior is guided by this convention.

Rapid guessing means a violation of this convention. Various reasons can lead to its violation including test preparation courses that advise participants to respond to all items even if there is not enough time for completing them appropriately. Following this advice leads to complete data that may be regarded as desirable because the missing data problem is avoided ([Little & Rubin, 2019](#)) in the statistical investigation. However, this solution to the missing data problem does not take into consideration that responses at

random, as data at random do not harmonize with data originating from the latent source of responding that is measured by the scale. The lack of harmonization means that rapid guessing cannot prevent the impairment of measurement validity due to a time limit in testing (Lu & Sireci, 2007).

As is demonstrated in our results, there remains the possibility to capture systematic variation that is due to latent processing speed despite rapid guessing. It is not even necessary to modify the model for the investigation of speeded data because of rapid guessing and also not necessary to measure processing times (Wise & Kong, 2005). The expected values obtained by the logistic function with and without the consideration of rapid guessing only differ by a constant multiplier. The results suggest that despite the use of rapid guessing in completing the items, it is possible to detect latent processing speed as one source of responding.

Although the use of rapid guessing does not prevent the detection of latent speed as one source of responding, it is not without a negative consequence for structural investigations. The integration of the probability of a correct response at random into the formal representation of the expected effect of the time limit in testing suggests an impairment of the probability of detecting this effect in structural investigations. This impairment is demonstrated to depend on the number of response options. Our results support the hypothesis suggesting such impairment.

A limitation of the present study is the assumption that all participants perform rapid guessing and that omissions completely disappear. Another limitation is the assumption of independence of ability and rapid guessing. Further limitations are the considerations of a single test length, the arrangement of items, the absence of omissions due to other sources, constancy of sample size, independence of factors. Moreover, free factor loadings (Estrada et al., 2017) are not considered besides fixed factor loadings. The assumptions may not hold in real test taking situations. The behavior observed in real test taking situations appears to show a large degree of variability (McNulty et al., 2007) instead of the homogeneity assumed for the reported study and to depend on the participants cognitive ability (Lindner et al., 2019). There may also be participants who behave as expected besides others who behave inconsistently. But, given that we investigated a kind of worst-case scenario, our results are promising regarding the detection of latent processing speed as source of responding.

Funding: The authors have no funding to report.

Acknowledgments: The authors have no additional (i.e., non-financial) support to report.

Competing Interests: The authors have declared that no competing interests exist.

References

- Borner, N., Völke, A. E., & Troche, S. J. (2020). The development of inductive reasoning under consideration of the effect due to test speededness. *Psychological Test and Assessment Modeling*, *62*, 344-358.
- Byrne, B. M. (2016). Using multitrait-multimethod analysis in testing for evidence of construct validity. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction* (pp. 288-307). Hogrefe Publishing.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255.
https://doi.org/10.1207/S15328007SEM0902_5
- DiStefano, C. (2016). Examining fit with structural equation models. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction* (pp. 26-51). Hogrefe Publishing.
- Drasgow, F., & Mattern, K. (2006). New tests and new items: Opportunities and issues. In D. Bartram & R. Hambleton (Eds.), *Computer-based testing and the Internet: Issues and advances* (pp. 59-75). John Wiley & Sons.
- Estrada, E., Román, F., Abad, F., & Colom, R. (2017). Separating power and speed components of standardized intelligence measures. *Intelligence*, *61*, 159-168.
<https://doi.org/10.1016/j.intell.2017.02.002>
- Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability. *Educational and Psychological Measurement*, *66*(6), 930-944.
<https://doi.org/10.1177/0013164406288165>
- Guan, Y. (2009). Variance stabilizing transformations of Poisson, binomial and negative binomial distributions. *Statistics and Probability Letters*, *79*(14), 1621-1629.
<https://doi.org/10.1016/j.spl.2009.04.010>
- Johnson, R. L., & Morgan, G. (2016). Item types, response formats, and consequences for statistical investigations. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction* (pp. 83-103). Hogrefe Publishing.
- Jöreskog, K. G., & Sörbom, D. (2001). *Interactive LISREL: User's guide*. Scientific Software International.
- Jöreskog, K. G., & Sörbom, D. (2006). *LISREL 8.80: User's reference guide*. Scientific Software International.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). The Guilford Press.
- Kubinger, K. D. (2003). On artificial results due to using factor analysis for dichotomous variables. *Psychology Science*, *45*, 106-110.
- Lai, K. (2018). Estimating standardized SEM parameters given nonnormal data and incorrect model: Methods and comparisons. *Structural Equation Modeling*, *25*(4), 600-620.
<https://doi.org/10.1080/10705511.2017.1392248>

- Lindner, M. A., Lüdtke, O., & Nagy, G. (2019). The onset of rapid-guessing behavior over the course of testing time: A matter of motivation and cognitive resources. *Frontiers in Psychology, 10*, Article 1533. <https://doi.org/10.3389/fpsyg.2019.01533>.
- Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data* (3rd ed.). John Wiley and Sons.
- Lu, Y., & Sireci, S. G. (2007). Validity issues in test speededness. *Educational Measurement, 26*(4), 29-37. <https://doi.org/10.1111/j.1745-3992.2007.00106.x>
- McCullagh, P., & Nelder, J. A. (1985). *Generalized linear models*. Chapman and Hall.
- McNulty, J. A., Sonntag, B., & Sinacore, J. M. (2007). Test-taking behaviors on a multiple-choice exam are associated with performance on the exam and with learning style. *International Association of Medical Science Educators, 17*, 52-57.
- Morgenthaler, S., & Staudte, R. G. (2012). Advantages of variance stabilization. *Scandinavian Journal of Statistics, 39*(4), 714-728. <https://doi.org/10.1111/j.1467-9469.2011.00768.x>
- Partchev, I., De Boeck, P., & Steyer, R. (2013). How much power and speed is measured in this test? *Assessment, 20*(2), 242-252. <https://doi.org/10.1177/1073191111411658>
- Peterson, R. A., Kim, Y., & Choi, E. (2020). A meta-analysis of construct reliability indices and measurement of model fit metrics. *Methodology, 16*(3), 208-233. <https://doi.org/10.5964/meth.2797>
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667-696. <https://doi.org/10.1080/00273171.2012.715555>
- Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice, 24*, 3-13.
- Schweizer, K., DiStefano, C., & Reiß, S. (2018). Structural validity of the OSA figures scale for the online self-assessment of fluid reasoning. *European Journal of Psychological Assessment, 34*(5), 321-327. <https://doi.org/10.1027/1015-5759/a000345>
- Schweizer, K., Gold, A., & Krampen, D. (2020). A semi-hierarchical confirmatory factor model for speeded data. *Structural Equation Modeling, 27*(5), 773-780. <https://doi.org/10.1080/10705511.2019.1707083>
- Schweizer, K., Reiß, S., Ren, X., Wang, T., & Troche, S. (2019). Speed effect analysis using the CFA framework. *Frontiers in Psychology (Section Quantitative Psychology and Measurement), 10*, Article 239. <https://doi.org/10.3389/fpsyg.2019.00239>.
- 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). Springer.
- Schweizer, K., Troche, S., & Reiß, S. (2019). Does the effect of a time limit for testing impair structural investigations by means of confirmatory factor models? *Educational and Psychological Measurement, 79*(1), 40-64. <https://doi.org/10.1177/0013164418770824>

- Schweizer, K., Wang, T., & Ren, X. (2020). On the detection of speededness in data despite selective responding using factor analysis. *Journal of Experimental Education*. Advanced online publication. <https://doi.org/10.1080/00220973.2020.1808942>.
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modelling: Multilevel, longitudinal and structural equation models*. Chapman & Hall/CRC.
- Stenlund, T., Lyrén, P.-E., & Eklöf, H. (2018). The successful test taker: Exploring test-taking behavior profiles through cluster analysis. *European Journal of Psychology of Education*, 33, 403-417.
- Vogt, W. P., & Johnson, R. B. (2015). *The SAGE dictionary of statistics & methodology: A nontechnical guide for the social sciences* (5th ed.). SAGE.
- Wise, S. L. (2017). Rapid guessing behavior: Its identification, interpretation, and implications. *Educational Measurement: Issues and Practice*, 36(4), 52-61. <https://doi.org/10.1111/emip.12165>
- Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, 18(2), 163-183. https://doi.org/10.1207/s15324818ame1802_2
- Zeller, F., Reiss, S., & Schweizer, K. (2017). Is the item-position effect in achievement measures induced by increasing item Difficulty? *Structural Equation Modeling*, 24(5), 745-754. <https://doi.org/10.1080/10705511.2017.1306706>

Appendices

Appendix A

Steps of a CFA Investigation for the Detection of Rapid Guessing in Data With an Hidden Effect of a Time Limit in Testing

Data:

1. Compute probability-based covariances according to the following equation:

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

where X_i and X_j ($i, j = 1, \dots, p$) are dichotomous variables.

Model:

2. Select the bifactor model of measurement for the investigation
3. Select number 1 as fixation for factor loadings on the first factor or set them free
4. Select an item position as preliminary turning point (tp) for the effect of a time limit
5. Compute fixations for factor loadings on the second factor according to [Equation 9](#).
6. Compute weights according to [Equation 13](#)
7. Insert the information on the number of factors, the weights and the factor loadings in the statistics software
8. Assure that variance parameters of factors with fixed factor loadings are set free for estimation

9. Assure that the error variances are set free for estimation
10. Select the maximum likelihood estimation method
11. Select the matrix including the probability-based covariances of step 1 as input
12. Start the program and save the fit results
13. Repeat the steps 4 to 13 with varying item positions as turning point to identify the turning point yielding the best degree of model fit (if this point is not known)
14. Compare the fit result for this turning point with the result for a one-factor model

Appendix B

List of MPLUS Commands (Provided by Brian French)

```

TITLE: Karl's Example
DATA:
FILE=S:\COEPrivate\frenchb\Papers\Karl_speed_2021\example_cov.txt;
nobservations = 500;
type = covariance;
VARIABLE: NAMES ARE  I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12
                    I13 I14 I15 I16 I17 I18 I19 I20;
USEVARIABLES ARE
    I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12
    I13 I14 I15 I16 I17 I18 I19 I20;
Analysis:
    ESTIMATOR = ML;
MODEL:
    GEN BY  I1* I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12
           I13 I14 I15 I16 I17 I18 I19 I20;
    Speed by I1@ I2@ I3@ I4@ I5@ I6@ I7@ I8@ I9@ I10@ I11@ I12@
           I13@ I14@ I15@ I16@ I17@ I18@ I19@ I20@;
           I11@.00041 I12@.00114 I13@.00316 I14@.00862 I15@.02309
           I16@.05875 I17@.13413 I18@.24850 I19@.34373 I20@.34113;
    Gen@1;
    Gen with Speed@0;
OUTPUT: STDYX;

```



Methodology is the official journal
of the European Association of
Methodology (EAM).



leibniz-psychology.org

PsychOpen GOLD is a publishing
service by Leibniz Institute for
Psychology (ZPID), Germany.