Psychological Test and Assessment Modeling, 2025.67:3-22
DOI: https://doi.org/10.2440/001-0016

Discrimination between types of common systematic variation in data contaminated by method effects using CFA models

Karl Schweizer[1]

[1] Goethe University Frankfurt, Germany

Abstract:
In data contaminated by method effects, common systematic variation is inhomogeneous requiring that attribute-related common systematic variation is in structural investigations discriminated from other variation. In the reported study, CFA measurement models dealing differently with such inhomogeneity were compared with respect to their performance in investigating data contaminated by either speededness or high subset homogeneity. For this purpose, structured random data with five different levels of speededness respectively subset-homogeneity were generated and investigated. The investigations were conducted by the one-factor congeneric and tau-equivalent CFA models, as well as the bifactor CFA model designed as mixture of tau-equivalent and fixed-links models. In data with speededness the congeneric model indicated good model fit while the tau-equivalent model showed sensitivity for the effect. In data with subset-homogeneity both models showed sensitivity. Only the bifactor model accounted for the common systematic variation and discriminated well between the attribute and method effects.

Keywords: comfirmatory factor analysis, discrimination, speededness, subset homogeneity, method effect

Correspondence:

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

Introduction

This essay reports research addressing the question whether confirmatory factor analysis (CFA) is suited for discriminating different types of common systematic variation in considering different CFA measurement models. Since CFA is designed according to the model-fit approach (Gumedze & Dunne, 2011; Jöreskog, 1969, 1970) that highlights the correspondence of the model and data, it does not comprise discrimination as one of its primary properties. CFA in the first place means checking whether the latent variable(s) of the measurement model account(s) for the common systematic variation of data that is due to individual differences in the attribute measured by the scale employed in data collection. This implicitly means that common systematic variation is assumed to be homogeneous so that discrimination is not necessary. Most CFA measurement models perform on the basis of this assumption (Alwin & Jackson, 1980; Graham, 2006).

But common systematic variation of data may be inhomogeneous instead of homogeneous. While the process of item selection in test construction (Johnson & Morgan, 2016) may exclude inhomogeneity regarding the attribute, there is the possibility that method effects create additional common systematic variation. Method effects are described as systematic variation in measurements that is not due to the attribute intended to be measured but to measurement (Maul, 2013; Schweizer, 2020). The strongest evidence that method effects contribute to common systematic variation is provided by multitrait-multimethod (MTMM) research using multitrait-multimethod measurement models (Byrne, 2016). The results of MTMM research suggest that differences between observers, instrumentation or the circumstances of measurement can lead to extra common systematic variation that may falsely be captured by the attribute latent variable of the CFA model. In MTMM research, the MTMM design is basic for the discrimination between common systematic variation associated with the attribute and common systematic variation due to the method of measurement.

Further, there are method effects that do not fit with an MTMM design because they cannot be captured by grouping the manifest variables of a CFA model as, for example, the item-position effect (Knowles, 1980; Kubinger, 2008) or the speededness of tests (Oshima, 1994). The item-position effect and speededness are effects that unfold toward the end of a sequence of items. The item-position effect steadily increases and is accepted as a regular nuisance that needs to be controlled in experimental research. The control of speededness is more demanding since this effect follows a non-linear course (Ren et al., 2017). Besides inhomogeneity of common systematic variation involving all manifest variables of a CFA model, there are the cases of inhomogeneity restricted to subsets of manifest variables as, for example, the wording effect (DiStefano & Motl, 2006). The most extreme case of inhomogeneity of common systematic variation is the phenomenon referred to as correlated residual (Landis et al., 2009) including two items of a scale that correlate to a much higher degree among each other than the remaining items.

Discrimination between different types of common systematic variation using CFA models is well-established in MTMM research (Byrne, 2016). There is also some research demonstrating that the bifactor CFA model enables the discrimination of common systematic variation due to the attribute from common systematic variation due to the item-position effect (e.g., Ren et al., 2012; Thomas et al., 2015). Even there is the proposal to transfer the experimental approach of dealing with the position effect to the field of differential research (Goodhew & Edwards, 2019).

Less well established is discrimination when the additional common systematic variation is due to speededness or high subset homogeneity although the efficiency of the bifactor model with respect to data displaying these effects has already been studied. Regarding speededness, it has been demonstrated that, when using a latent variable reflecting the distribution of processing speed as part of a CFA measurement model, it is possible to account for the common systematic variation due to speededness (e.g., Ren et. al., 2017). Regarding subset homogeneity, there is already some evidence of the feasibility of representing subsets of two highly correlated manifest variables by a latent variable in order to overcome the inhomogeneity due to the different types of common systematic variation (Schweizer et al., 2023).

Another issue that has to be taken into consideration is that CFA measurement models (Alwin & Jackson, 1980; Graham, 2006) can include two different types of factor loadings: free factor loadings and fixed factor loadings that may be differently suited for the discrimination of types of common systematic variation. In the case of free factor loadings, the latent variable is expected to account for the complete common systematic variation (Jöreskog, 1971) whereas in the other case, the fixation of factor loadings restricts the latent variable in accounting for the complete common systematic variation.

The research that is described in the following sections investigated the efficiency of CFA models with free and fixed factor loadings in discriminating between common systematic variation associated with the attribute and common systematic variation due to speededness on one hand and due to high subset homogeneity on the other hand.

The Outset

The CFA measurement model with free factor loadings is the congeneric CFA measurement model that is a one-factor model (Brown, 2015; Jöreskog, 1971). The version for centered data is given by

$$\mathbf{x} = \boldsymbol{\lambda}_{\text{attribute}} \xi_{\text{attribute}} + \boldsymbol{\delta} \quad (1)$$

where $\mathbf{x}$ is the $p \times 1$ vector of centered manifest variables representing the observations, $\boldsymbol{\lambda}_{\text{attribute}}$ is the $p \times 1$ vector including the factor loadings that quantify the relationships of the attribute latent variable and the manifest variables, $\xi_{\text{attribute}}$ is the latent variable representing the attribute, and $\boldsymbol{\delta}$ is the $p \times 1$ vector of residuals that quantifies the influences that are unique for each manifest variable. Free factor loadings are parameters included in $\boldsymbol{\lambda}_{\text{attribute}}$ that are estimated in addition to the parameters included in $\boldsymbol{\delta}$. These characteristics are shared with the common-factor model of exploratory factor analysis (Lawley & Maxwell, 1971).

A basic characteristic of fixed factor loadings is that they are restricted. Mostly the sizes of fixed factor loadings are restricted to correspond while the overall size is estimated. For example, the tau-equivalent CFA measurement model (Graham, 2006) requires the estimation of the general factor loading that serves as estimate for all factor loadings. We represent it by the symbol τ. Therefore, our formalization of the concept of the tau-equivalent measurement model includes $\tau_{\text{attribute}}\mathbf{1}$ instead of $\boldsymbol{\lambda}_{\text{attribute}}$:

$$\mathbf{x} = \tau_{\text{attribute}} \mathbf{1} \xi_{\text{attribute}} + \boldsymbol{\delta} \quad (2)$$

where $\mathbf{x}$ is the $p \times 1$ vector of centered manifest variables representing the observations, $\tau_{\text{attribute}}$ is the general factor loading, $\mathbf{1}$ is the $p \times 1$ unit vector, $\xi_{\text{attribute}}$ is the latent variable representing the attribute and $\boldsymbol{\delta}$ is the $p \times 1$ vector of residuals. The tau-equivalent CFA measurement model implicitly assumes that the latent variable equally relates to all manifest variables and suggests the interpretation that all items (→ manifest variables) used in data collection show the same degree of discrimination.

The fixed-links CFA measurement model (Schweizer, 2006) in a way extends the tau-equivalent CFA measurement model to be applicable for investigating experimental effects. The extension consists in replacing the unity factor, $\mathbf{1}$, by a $p \times 1$ vector, $\boldsymbol{\upsilon}_{\text{experimental\_effect}}$, that includes numbers representing partial hypotheses on the relationships of manifest variables and the latent variable reflecting the expected experimental effect. Replacing them by numbers characterizing a method effect turns $\boldsymbol{\upsilon}_{\text{experimental\_effect}}$ into $\boldsymbol{\upsilon}_{\text{method\_effect}}$ for investigating the presence of common systematic variation due to a method effect:

$$\mathbf{x} = \omega_{\text{method\_effect}} \boldsymbol{\upsilon} \xi_{\text{method\_effect}} + \boldsymbol{\delta} \quad (3)$$

where $\mathbf{x}$ is the $p \times 1$ vector of centered manifest variables representing the observations, $\omega_{\text{method\_effect}}$ is the parameter serving as the general factor loading, $\boldsymbol{\upsilon}_{\text{method\_effect}}$ is the $p \times 1$ vector specifying the assumed relationships of manifest variables and the latent variable, $\xi_{\text{method\_effect}}$ is the latent variable representing the method effect and $\boldsymbol{\delta}$ is the $p \times 1$ vector of residuals.

Since the tau-equivalent model and the fixed-links model include restrictions, they are not suited to account alone for the complete common systematic variation of data in cases of inhomogeneity. In such a case the replacement of a one-factor CFA model by the bifactor model (Reise, 2012) is required. One component of this bifactor model may be specified according to the tau-equivalent model and another one according to the fixed-links model so that

$$\mathbf{x} = \tau_{\text{attribute}} \mathbf{1} \xi_{\text{attribute}} + \omega_{\text{method\_effect}} \boldsymbol{\upsilon}_{\text{method\_effect}} \xi_{\text{method\_effect}} + \boldsymbol{\delta} \; . \qquad (4)$$

In Equation 4 **x** reflects available observations, $\tau_{\text{attribute}}$ is the parameters servings as general factor loading regarding the attribute, **1** is the $p \times 1$ unity vector, $\xi_{\text{attribute}}$ is the latent variable representing the attribute, $\omega_{\text{method\_effect}}$ is the parameter servings as the weight regarding the influence of the method effect, $\boldsymbol{\upsilon}_{\text{method\_effect}}$ is the $p \times 1$ vector including numbers representing the assumed effect, $\xi_{\text{method\_effect}}$ is the latent variable representing the method effect, and $\boldsymbol{\delta}$ is the $p \times 1$ vector of residuals.

Some Considerations Regarding the Data Structure

Investigating the appropriateness of measurement models for discriminating types of common systematic variation requires data that allow for at least two types of common systematic variation. One type of common systematic variation has to be the attribute common systematic variation while the other type can be the common systematic variation of a method effect. In a simulation study, this requires the mimicking of the selected method effect in data generation.

The first method effect (speededness) is taken from the set of end-of-scale effects that comprises effects with an increasing probability to influence the outcome of assessment along the sequence of items of a scale. These effects are either obvious in missing data or alternatively in random responses. For example, there is the effect due to a time limit in testing that leads to missing data if the participant’s speed of cognitive processing is insufficient (Oshima, 1994; Partchev et al., 2013). But some participants who are about to run out of time in completing the items may not simply go on as long as possible. Instead they may resort to the response strategy titled rapid guessing (Wise, 2017) leading to random responses. Decline of the motivation and fatigue toward the end of a scale may also lead to this type of response.

We concentrate on the end-of-scale effect that appears to have attracted most attention in science so far that is the speed-related missing data effect observable if participants perform as they are instructed to do but are not fast enough. We refer to it as the *speededness condition*. Systematic variation due to speed-related missing data is not at random (Little & Rubin, 2019) and may not be captured by the latent variable of a model representing the attribute measured by the scale. Research demonstrates that this effect can be represented by a latent variable reflecting the distribution of missing data (Borter et al., 2023; Ren et al., 2017). This knowledge can be used to simulate data including a speed-related missing data effect. An increase following the curve characterizing the cumulative normal distribution (function $g_{\text{trajectory}}$) must characterize not only the simulated data but also the fixations serving as factor loadings on the latent variable representing this effect (ith factor loading λ_i, $i = 1,\ldots, p$):

$$\lambda_i = g_{\text{trajectory}}(i)\ .$$

The other method effect selected for this study is frequently reported and commonly referred to as correlated residuals (Landis et al., 2009). Since a basic characteristic of the model-fit approach excludes correlations of residuals, we prefer to address this method effect as high subset homogeneity. The set of manifest variables displaying this effect can be perceived as a set composed of at least two non-overlapping subsets ($\{\{X_1, \ldots, X_{p-2}\}, \{X_K, X_L\}\}$) that differ according to the sizes of their intercorrelations:

$$r_{KL} > r_{ij}\ .$$

Correctly conceptualized, such a method effect can be captured by an additional latent variable of a bifactor model (Schweizer et al., 2023), that is, each subset can be assumed to give rise to its own common systematic variation. To mimic this method effect means generating data matrices including one or a few pairs of columns with especially large intercorrelation(s) while the correlations among the remaining columns are smaller. We refer to data with the described characteristic as *subset-homogeneity condition*.

The Empirical Section

The empirical section reports a simulation study that served the aim to investigate whether the described CFA models were appropriate for discriminating between two types of common systematic variation when data with inhomogeneous common systematic variation were to be investigated. The data for this study were generated to display either the speededness or subset-homogeneity conditions in considering different levels reaching from no effect to a large effect.

Method

The two data conditions outlined in the previous section gave rise to two parts of the study report. Each condition included five levels, and the number of data matrices created and investigated was 500 at each level. The generation of data according to the speededness condition was accomplished in the following way: in the first step, 500×13 matrices including normally distributed random data [$N(0,1)$] were generated on the basis of a relational pattern by means of PRELIS (Jöreskog & Sörbom, 2001). The off-diagonal entries of the relational pattern corresponded and were selected to give rise to expected factor loadings of 0.35 while the diagonal entries were one. In the next step, the effect of speededness, that is, information on omissions due to lack of enough processing speed, was integrated into the matrices. The following columns were selected for the onsets of the speed effect conditions: the seventh (80 percent), eights (60 percent), ninth (40 percent) and tenth (20 percent) columns.

Figure 1 illustrates the expected numbers of omissions in a sample of 500 observations for the various items position and levels, as they were realized in the study. Next, zeros were inserted in the matrices at randomly selected rows for representing omissions. Zeros representing omission in zero-standardized data could be expected to reduce the common systematic variation associated with the attribute systematically. In addition, the distribution of zeros across the item positions was expected to create additional common systematic variation that indirectly represented processing speed and provided the basis for what was captured by the speededness latent variable (for more information see Borter et al., 2023; Ren et al., 2017). Finally, the data matrices were transformed into covariance matrices.

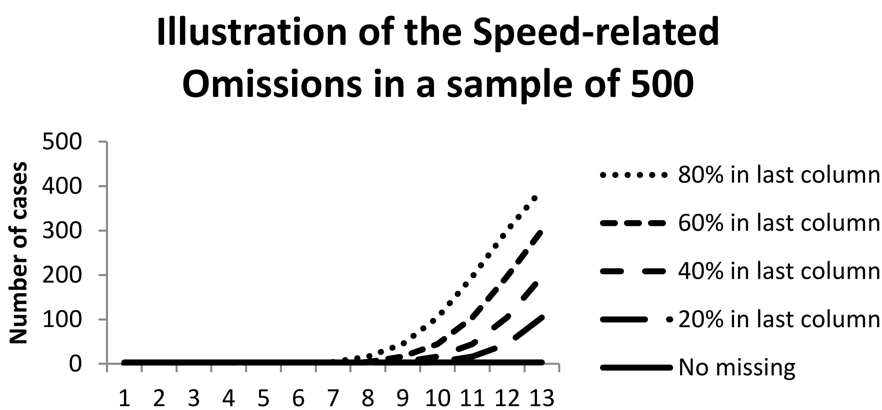

Figure 1. Illustration of the distribution of omissions under the levels of the speededness condition.

The generation of data according to the subset-homogeneity condition occurred in a similar way. A set of five-hundred 500 × 13 data matrices was generated by means of a relational pattern that differed from the one used for the speededness condition in three spots. The numbers at the intersections of the *2*nd row and *5*th column, the *7*th row and *8*th column as well as *9*th row and *12*th column of the relational pattern were replaced. The replacements were enlargements that were expected to create the subset-homogeneity condition in the three spots instead of one in order to have a strong overall effect. The following increments were selected to establish five levels: 0 (first level), 0.20 (second level), 0.25 (third level), 0.30 (fourth level) and 0.35 (fifth level).

The data were investigated using three CFA models. These models were specified as congeneric model (Equation 1), tau-equivalent model (Equation 2) and the combination of tau-equivalent model and fixed-links model (Equation 3) realized as bifactor model (Equation 4). The congeneric model included one latent variable with free factor loadings while the corresponding variance parameter was set to one so that *p* factor loadings had to be estimated. The other models included fixed numbers instead of parameters while either one or two variance parameters had to be estimated. The fixations for the attribute (tau-equivalent) latent variable were equal-sized numbers. Note. Any positive number can be used since the size of this number only influences the size of the variance parameter estimate (Schweizer et al., 2019) but not the degree of model fit. The fixations for the method effect (fixed-links) latent variable that represented speededness were selected according to the cumulative normal distribution function, that is, they corresponded to the numbers for generating the curves of Figure 1. Equal-sized numbers surmounting the numbers used for the fixation of the attribute (tau-equivalent) latent variable were used as factor loadings on the three factors representing subset homogeneity. The latent variables were not allowed to correlate with each other to avoid adverse effects on factor variances (Schweizer et al., 2024)

LISREL software (Jöreskog & Sörbom, 2006) specified to conduct maximum likelihood estimation was selected for the investigation of the generated covariance matrices. The evaluation of model fit occurred using recommended fit indices (established fit criteria are provided in parentheses): RMSEA (≤ 0.06), SRMR (≤ 0.08), NNFI (≥ 0.95), and CFI (≥ 0.95) (see DiStefano, 2016; Hu & Bentler, 1999). We also report χ^2s, dfs, and AICs. Means and standard deviations were computed and included in tables. Comparisons between models and across levels were accomplished using the CFI difference (0.01) and the RMSEA difference (0.015) according to Cheung and Rensvold (2002).

Results

Since the focus was on discrimination between the two types of common systematic variation, one-factor models were expected to yield model misfit for larger method effects (= highest levels). Otherwise the outcome would imply that there was only one type of common systematic variation that was incorrect. Since the data included two types of common systematic variation, only the bifactor model that included two latent variables, one for each type of common systematic variation, could be expected to correctly signify good model fit.

Results of investigating data with speededness. The fit results of this investigation are included in Table 1 that comprises of three sections: the first sections with the results obtained by the one-factor congeneric CFA model, the second section with the results by the one-factor tau-equivalent CFA model, and the third section with the results by the bifactor CFA model specified as two-factor model with fixed factor loadings.

All estimates for the congeneric model regarding the fit indices with a cutoff (RMSEA, SRMR, NNFI, CFI) reported in the first section indicated good model fit. Comparing the results across the levels revealed only two substantial differences: the CFI of the fifth level was substantially smaller than the CFIs of the first and second levels. In contrast, only the RMSEA and SRMR results for the tau-equivalent model included in the second section indicated good model fit across all levels whereas good NNFI and CFI results were restricted to the level one to three. Further, the RMSEAs of levels one to three differed from that of levels four and five while the CFIs of all levels differed except of the CFIs of levels one and two. Regarding the last section concerning the bifactor model, there were only results for the fourth and fifth levels because in the other levels either a second latent variable was not expected (first level) or there was no valid/substantial estimate for the speededness latent variable in a large number of cases. All estimates for the remaining levels (fourth and fifth levels) indicated good model fit. Further, the RMSEAs and CFIs of this model did not differ from the RMSEAs and CFIs of the levels one to three observed for the congeneric and tau-equivalent models.

In sum, the results for the tau-equivalent model displayed strong sensitivity for inhomogeneous common systematic variation whereas the congeneric model only weak sensitivity. The bifactor model discriminated well between the two types of common systematic variation in the levels where it was applicable.

Results of investigating data with high subset homogeneity. The report of the fit results is included in Table 2 that is structured in the same way as Table 1: the first sections includes the results obtained by the one-factor congeneric CFA model, the second section the results by the one-factor tau-equivalent CFA model, and the third section the results by the bifactor CFA model specified as two-factor model.

All RMSEAs and SRMRs of the first section indicated good model fit for the congeneric model whereas only the first to third levels of NNFIs and CFIs also signified good model fit. In comparing RMSEAs and CFIs across the five levels, several RMSEA differences and CFI differences reached the significance level suggesting a decrease of model fit across the treatment levels. Similar results were observed for the tau-equivalent model. All RMSEAs and SRMRs indicated good model fit while regarding NNFI and CFI only the estimates for the levels one to three were good. Additionally, several of the RMSEA differences and CFIs differences across the five levels were substantial and suggested a decrease of model fit. Furthermore, the RMSEA and CFI estimates of the congeneric and tau-equivalent models did neither differ according to any RMSEA difference nor according to any CFI difference. The last section including results for the bifactor model is incomplete since a second latent variable was not expected for the first level. All estimates of all fit indices for the remaining levels indicated good model fit. Further, the RMSEAs did not differ across the levels while regarding the CFIs there were three (marginal) cases. The CFI differences of the last level and the remaining levels were exactly 0.01.

In sum, the results for the congeneric and tau-equivalent models displayed sensitivity for inhomogeneous common systematic variation and did not differ substantially from each other. In contrast, the bifactor model discriminated well between the two types of common systematic variation.

Table 1
Results of Investigating Model Fit in Five Sets of Five-hundred 500 × 13 Matrices of Structured Random Data under the Speededness Condition

Model type	Level	Analysis type	χ^2	df	RMSEA	SRMR	NNFI	CFI	AIC
Congeneric	1	Mean	65.51	65	0.008	0.033	0.998	0.991	117.51
		SD	*11.65*		*0.010*	*0.003*	*0.025*	*0.013*	*11.65*
	2	Mean	65.60	65	0.008	0.033	0.998	0.991	117.60
		SD	*11.44*		*0.009*	*0.003*	*0.025*	*0.013*	*11.44*
	3	Mean	65.90	65	0.008	0.033	0.997	0.989	117.90
		SD	*12.05*		*0.010*	*0.003*	*0.028*	*0.015*	*12.05*
	4	Mean	67.63	65	0.010	0.034	0.992	0.986	119.63
		SD	*12.47*		*0.010*	*0.003*	*0.032*	*0.019*	*12.47*
	5	Mean	71.39	65	0.012	0.035	0.982	0.979	123.39
		SD	*13.23*		*0.011*	*0.003*	*0.037*	*0.024*	*13.23*
Tau-equivalent	1	Mean	77.10	77	0.007	0.041	0.999	0.991	105.10
		SD	*12.26*		*0.009*	*0.004*	*0.022*	*0.014*	*12.26*
	2	Mean	79.15	77	0.008	0.042	0.995	0.989	107.15
		SD	*12.52*		*0.009*	*0.004*	*0.023*	*0.016*	*12.52*
	3	Mean	87.05	77	0.014	0.046	0.980	0.977	115.05
		SD	*14.39*		*0.010*	*0.005*	*0.028*	*0.023*	*14.39*
	4	Mean	111.63	77	0.029	0.055	0.934	0.934	139.63
		SD	*19.02*		*0.009*	*0.006*	*0.037*	*0.036*	*19.02*
	5	Mean	168.46	77	0.048	0.069	0.825	0.827	196.48
		SD	*30.37*		*0.008*	*0.007*	*0.053*	*0.053*	*30.37*

Table 1 continued

Bifactor	1[1]	Mean	-	-	-	-	-	-	-
		SD	-	-	-	-	-	-	-
	2[2]	Mean	-	-	-	-	-	-	-
		SD	-	-	-	-	-	-	-
	3[2]	Mean	-	-	-	-	-	-	-
		SD	-	-	-	-	-	-	-
	4	Mean	76.31	76	0.008	0.040	0.996	0.987	108.31
		SD	*12.95*		*0.009*	*0.001*	*0.029*	*0.019*	*12.95*
	5	Mean	78.41	76	0.010	0.041	0.991	0.983	110.41
		SD	*13.10*		*0.010*	*0.004*	*0.032*	*0.022*	*13.10*

[1] Since a second source of common systematic variation was not expected. the bifactor model was not applied under this condition.
[2] No results are reported since in the majority of cases the factor variance of the method-effect latent variable was insignificant or negative or error estimates of parameters were not provided.

Discussion

Discrimination between types of common systematic variation is an issue of relevance for the establishment of the construct validity of a scale (Messick, 1981) that is closely linked to the multitrait-multimethod approach of investigating the validity of psychological scales (Campbell & Fiske, 1959). Using multitrait-multimethod CFA models for investigating data means discrimination between variation that is due to attributes and variation that is due to a characteristic of measurement on the basis of the multitrait-multimethod design. Discrimination by the bifactor model serves the exactly same aim but not on the basis of a multitrait-multimethod design. It seeks to discriminate common systematic variation into two types, one type that reflects the attribute and another type that reflects a characteristic of assessment other than the attribute.

In contrast, discrimination between different types of common systematic variation is not a property of the standard version of confirmatory factor analysis including the congeneric measurement model (Brown, 2015; Graham, 2006). The standard version is a one-factor model with a latent variable that is designed to account for as much common systematic variation as possible. This version can be expected to perform well if there is no other common systematic variation than common systematic variation due to the attribute that is measured. But this condition is unlikely to hold outside of simulation studies since, for example, repeated measurements are likely to display an effect that is referred to as sequence effect and is routinely controlled in experimental psychology. There is also an abundance of evidence of this effect in differential psychology where it is referred to as item-position effect (e.g., Kubinger, 2008; Ren et al., 2012; Scharfen, 2018, Zeller et al., 2019).

Table 2
Results of Investigating Model Fit in Five Sets of Five-hundred 500 × 13 Matrices of Structured Random Data under the Subset-Homogeneity Condition

Model type	Level	Analysis type	χ^2	df	RMSEA	SRMR	NNFI	CFI	AIC
Congeneric	1	Mean	65.42	65	0.008	0.033	0.996	0.989	117.42
		SD	*11.65*		*0.009*	*0.003*	*0.025*	*0.013*	*11.65*
	2	Mean	74.48	65	0.015	0.035	0.981	0.982	126.48
		SD	*13.26*		*0.011*	*0.003*	*0.025*	*0.018*	*13.26*
	3	Mean	91.51	65	0.027	0.039	0.951	0.959	143.51
		SD	*15.21*		*0.010*	*0.003*	*0.028*	*0.023*	*15.21*
	4	Mean	117.93	65	0.040	0.044	0.909	0.924	169.93
		SD	*18.49*		*0.007*	*0.004*	*0.033*	*0.028*	*18.49*
	5	Mean	154.08	65	0.052	0.049	0.857	0.880	206.08
		SD	*22.75*		*0.007*	*0.004*	*0.040*	*0.034*	*22.75*
Tau-equivalent	1	Mean	77.10	77	0.007	0.041	0.999	0.991	105.10
		SD	*12.26*		*0.009*	*0.004*	*0.022*	*0.014*	*12.26*
	2	Mean	86.62	77	0.013	0.043	0.983	0.981	114.62
		SD	*13.67*		*0.010*	*0.004*	*0.022*	*0.019*	*13.67*
	3	Mean	104.01	77	0.025	0.046	0.957	0.958	132.01
		SD	*15.64*		*0.009*	*0.004*	*0.025*	*0.024*	*15.64*
	4	Mean	131.35	77	0.037	0.051	0.920	0.920	159.35
		SD	*18.74*		*0.007*	*0.004*	*0.029*	*0.028*	*18.74*
	5	Mean	168.90	77	0.048	0.057	0.873	0.875	196.90
		SD	*18.74*		*0.007*	*0.004*	*0.029*	*0.028*	*18.74*

Table 2 continued

Bifactor	1[1]	Mean	-	-	-	-	-	-	-
		SD	-	-	-	-	-	-	-
	2	Mean	73.87	76	0.007	0.040	0.999	0.993	107.87
		SD	*11.99*		*0.009*	*0.004*	*0.018*	*0.011*	*11.99*
	3	Mean	73.87	76	0.007	0.040	0.999	0.993	107.87
		SD	*11.99*		*0.009*	*0.004*	*0.018*	*0.011*	*11.99*
	4	Mean	73.51	76	0.007	0.040	1.000	0.993	107.51
		SD	*11.58*		*0.008*	*0.004*	*0.019*	*0.011*	*11.58*
	5	Mean	82.01	76	0.012	0.044	0.987	0.983	116.01
		SD	*20.42*		*0.013*	*0.010*	*0.030*	*0.024*	*20.42*

[1] Since a second source of common systematic variation is not expected. the bifactor model is not applied under this condition.

Latent variables can be prevented from accounting for the complete common systematic variation by constraining factor loadings. An example of a CFA measurement model including such a latent variable is the tau-equivalent CFA measurement model (Lord & Novick, 1968, p. 58; Jöreskog, 1970). Factor loadings constrained to equal sizes characterize this measurement model. Accounting for method effects typically requires sequences of numbers varying in size to reflect the trajectory of the effect for replacing factor loadings, as is enabled by the fixed-links model (Schweizer, 2006) proposed for representing experimental effects. Discrimination between different types of common systematic variation also requires that each one of the targeted types of common systematic variation is represented by an own latent variable as part of a bifactor model (Reise, 2012).

Data designed to include additional common systematic variation according to a method effect were expected to exclude the observation of good model fit when the contribution of the method effect was large and the CFA measurement model included one latent variable only. It turned out that neither RMSEA nor SRMR estimates were in line with this expectation; only NNFI and CFI estimates displayed some degree of sensitivity for method effects. In speededness data the results observed for the tau-equivalent model corresponded to the expectation. In the highest speededness level the NNFI and CFI results for this model indicated bad model fit while the fit results for the congeneric model decreased but were still good. When subset homogeneity served as method effect, both one-factor models were sensitive for the increasing effect size, and indicated bad model fit in the highest level. Overall, the tau-equivalent model seems to be under more conditions sensitive for method effects according to NNFI and CFI than the congeneric model.

The bifactor model led to good model fit under both conditions when it was applicable. Each latent variable accounted well for the corresponding part of the common systematic variation. We only applied this model when there was reason for assuming that the data included two types of common systematic variation. Otherwise it might indicate good model fit but the estimate of the variance parameter for the effect latent variable would be insignificant calling the validity of the model into question. That actually happened for the applications to data with speededness when the effect size was small. Despite this restriction to the applicability, the bifactor model proved to be well suited for the discrimination of two types of common systematic variation. These results are in line with previous investigations using similar conditions (Borter et al., 2023; Schweizer et al., 2023).

A limitation of this study is that the two considered method effects turn out as different types of effects. It would be preferable to have two method effects of the same type included in this study. Further studies are necessary to find out whether these types are general types. Another limitation is that the generated data are continuous and normally distributed data that avoids complications with binary data (Schweizer et al., 2021) but differ from the data investigated in applied studies and previous simulation studies.

As a further limitation with respect to applications, it may be argued that there can be other influences leading to what we have referred to as end-of-scale effects that may stay undetected in the described way of investigating data. So the additional latent variable of a model thought to capture speededness may actually not only account for speededness but also for other method effects. Regarding this argument we like to highlight that there is the possibility to include several method latent variables into the bifactor CFA model in order to also check the presence of further effects. For example, the simultaneous presence of the item-position effect and the difficulty effect in addition to the main effect were confirmed by a bifactor model with two additional latent variables (Schweizer, Troche et al., 2021). There is also the possibility to compare several models with each other were each one includes another method latent variable besides the main latent variable. For example, the CFA model including the speed-effect latent variable was compared with CFA models including other additional latent variables representing the difficulty effect, the item-position effect and the homogeneity effect (Schweizer, Reiß et al., 2019). Further, it is possible to investigate how response styles influences speededness (Schweizer et al., 2020).

Overall this research reveals that the utility of the one-factor measurement models for the demonstration of validity depends on the absence of method effects. While fixed factor loadings appear to fail under the influence of any method effect, free factor loadings seem to tolerate specific kinds of method effects despite an impairment of construct validity. In contrast, the bifactor model proves to be suitable for discriminating between different types of common systematic variation under various conditions, given that a suitable representation of the method effect is available.

Literature

Alwin, D. F., & Jackson, D. J. (1980). Measurement models for response errors in surveys: Issues and applications. In K. F. Schuessler (Ed.), *Sociological methodology* 1980 (pp. 69–119). Jossey-Bass Publishers.

Borter, N., Schlegel, K., & Troche, S. (2023). How speededness of a reasoning test and the complexity of mental speed tasks influence the relation between mental speed and reasoning ability. *Journal of Intelligence, 11*: 89. https://doi.org/10.3390/jintelligence11050089

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.

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.

Campbell, D. T., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105. https://doi.org/10.1037/h0046016

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. 166–193). Hogrefe Publishing.

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling: A Multidisciplinary Journal*, *13(3)*, 440–464. https://doi.org/10.1207/S15328007SEM1303_6

Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual-differences research: Contributions, challenges, and practical recommendations. *Consciousness and Cognition*, *69,* 14–25. https://doi.org/10.1016/j.concog.2019.01.008

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

Gumedze, F. N., & Dunne, T. T. (2011). Parameter estimation and inference in the linear mixed model. *Linear Algebra and Its Application, 435(8),* 1920–1944. https://doi.org/10.1016/j.Iaa.2011.04.015

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. https://doi.org/10.1080/10705519909540118

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*, 183–202. https://doi.org/10.1007/BF02289343

Jöreskog, K. G. (1970). A general method for analysis of covariance structure. *Biometrika, 57(2*), 239–251. https://doi.org/10.1093/biomet/57.2.239

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

Jöreskog, K. G., & Sörbom, D. (2001). *Interactive LISREL: User's guide*. Scientific Software International Inc.

Jöreskog, K. G., & Sörbom, D. (2006). *LISREL 8.80*. Scientific Software International Inc.

Johnson, R. L., & Morgan, G. B. (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.

Knowles, E. S. (1988). Item context effects on personality scales: Measuring changes the measure. *Journal of Personality and Social Psychology*, *55(2)*, 312–320. doi:10.1037/0022-3514.55.2.312

Kubinger, K. D. (2008). On the revival of the Rasch model-based LLTM: From constructing tests using item generating rules to measuring item administration effects. *Psychology Science Quarterly, 50*, 311–327.

Landis, R. S., Edwards, B. D., & Cortina, J. M. (2009). On the practice of allowing correlated residuals among indicators in structural equation models. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: verity and fable in the organizational and social sciences* (pp. 193–214). Routledge/ Taylor & Francis Group.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. American Elsevier.

Little, R. J. A., & Rubin, D. B. (2019). *Statistical Analysis with Missing Data* (3nd ed.). Wiley.

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

Maul, A. (2013). Method effects and the meaning of measurement. *Frontiers in Psychology*. http://doi.org/10.3389/fpsyg.2013.00169

Messick, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. *Psychological Bulletin*, *89*, 575–588. https://doi.org/10.1037/0033-2909.89.3.575

Oshima, T. C. (1994). The effect of speededness on parameter estimation in item response theory. *Journal of Educational Measurement, 31(3)*, 200–219. https://doi.org/10.1111/j.1745-3984.1994.tb00443.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

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement, 21(2)*, 173–184. https://doi.org/10.1177/01462216970212006

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

Ren, X., Goldhammer, F., Moosbrugger, H., & Schweizer, K. (2012). How relates attention to the ability-specific and position-specific components of reasoning measured by APM? *Learning and Individual Differences, 22,* 1–7. https://doi.org/10.1016/j.lindif.2011.09.009

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(1),* 64–71. https://doi.org/10.1016/j.intell.2017.11004

Scharfen, J., Peters, J. M., & Holling, H. (2018). Retest effects in cognitive ability tests: A meta-analysis. *Intelligence, 67(1),* 44–66. https://doi.org/10.1016/j.intell.2018.01.003

Schweizer, K. (2006). The fixed-links model in combination with the polynomial function as tool for investigating choice reaction time data. *Structural Equation Modeling, 13(3)*, 403–419. https://doi.org/10.1207/s15328007sem1303_4

Schweizer, K. (2020). Method Effects in Psychological Assessment. *Psychological Test and Assessment Modeling, 64*, 337–343.

Schweizer, K., Gold, A., & Krampen, D. (2021). Effect of dichotomization on the latent structure of data. *Frontiers in Applied Mathematics and Statistics, 7:*723258. http://doi.org/10.3389/fams.2021. 723258

Schweizer, K., Gold, A., & Krampen, D. (2023). High subset homogeneity impairs structural validity of the Barratt Impulsiveness Scale. *Journal of Individual Differences*. http://doi.org/10.1027/1614-0001/a000407

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*, ArtID 239. doi: 10.3389/fpsyg.2019.00239

Schweizer, K., Troche, S., Rammsayer, T., & Zeller, F. (2021). Inductive Reasoning and Its Underlying Structure: Support for Difficulty and Item Position Effects. *Advances in Cognitive Psychology, 17(4),* 250-260. http://doi.org/10.5709/acp-0334-7

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, 90,* 486-504. http://doi.org/10.1080/00220973.2020.1808942

Schweizer, K., Ren, X., Wang, T. (2024). Negativity of factor correlations biases the sizes of factor variances in bifactor CFA models. *Frontiers in Applied Mathematics and Statistics, 10*:1423726. http://doi.org/10.3389/fams.2024.1423726

Thomas, P., Rammsayer, T. H., Schweizer, K. & Troche, S. J. (2015). Elucidating the functional relationship between working memory capacity and psychometric intelligence: A fixed-links modeling approach for experimental repeated-measures designs. *Advances in Cognitive Psychology, 11,* 3–13.

Zeller, F., Reiß, S., & Schweizer, K. (2017). Is the item-position effect in achievement measures induced by increasing item difficulty? *Structural Equation Modeling: A Multidisciplinary Journal, 24(5),* 745–754. https://doi.org/10.1080/10705511.2017.1306706

Appendix

1. LISREL Code for the Speededness Study adapted to Covariances Computed from one Set of Structured Random Data

```
Speededness Study
DA NI=13 NO=500 MA=CM

LA
I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12 I13

CM

 1.3664
 0.3522  1.3775
 0.3591  0.3489  1.3583
 0.3773  0.3334  0.4198  1.3750
 0.4094  0.3048  0.3699  0.3417  1.4363
 0.3751  0.4170  0.4590  0.4056  0.3618  1.4023
 0.3509  0.3252  0.2612  0.4066  0.4610  0.4048  1.7687
 0.4064  0.2939  0.2474  0.3711  0.3283  0.3652  0.6956  1.6759
 0.4004  0.2414  0.2951  0.5311  0.3912  0.4230  0.7460  0.6360  1.6396
 0.4399  0.3308  0.3440  0.4537  0.3577  0.5206  0.7436  0.6662  0.6477  1.6941
 0.5118  0.3171  0.4149  0.4767  0.4741  0.4960  0.7795  0.6961  0.7212  0.6698  1.8099
 0.3220  0.3449  0.4481  0.4329  0.4021  0.5352  0.8225  0.7508  0.6743  0.8054  0.7532  1.8717
 0.4422  0.3134  0.3443  0.4826  0.3993  0.3548  0.7752  0.7784  0.7786  0.7464  0.8071  0.8742  1.8409


MO NX=13 NK=2 TD=FU,FI PH=FU,FI

LK
Attribute Speed

VA      0.2774   LX   1    1
VA      0.2774   LX   2    1
VA      0.2774   LX   3    1
VA      0.2774   LX   4    1
VA      0.2774   LX   5    1
VA      0.2774   LX   6    1
VA      0.2774   LX   7    1
VA      0.2774   LX   8    1
VA      0.2774   LX   9    1
VA      0.2774   LX   10   1
VA      0.2774   LX   11   1
VA      0.2774   LX   12   1
VA      0.2774   LX   13   1

VA      0        LX   1    2
VA      0        LX   2    2
VA      0        LX   3    2
VA      0        LX   4    2
VA      0        LX   5    2
VA      0        LX   6    2
VA      0.00736  LX   7    2
VA      0.02945  LX   8    2
```

```
VA      0.08099   LX      9       2
VA      0.19144   LX      10      2
VA      0.36080   LX      11      2
VA      0.55224   LX      12      2
VA      0.72160   LX      13      2

FR PH 1 1
FR PH 2 2
! FR PH 1 2

FR TD 1 1
FR TD 2 2
FR TD 3 3
FR TD 4 4
FR TD 5 5
FR TD 6 6
FR TD 7 7
FR TD 8 8
FR TD 9 9
FR TD 10 10
FR TD 11 11
FR TD 12 12
FR TD 13 13

OU ML SC IT=1000 AD=OFF ND=3
```

2. PRELIS Code for the Generation of Structured Random Data Serving as Input to LISREL

```
! Generating Multivar. Normal variables
DA NO=500
NE V1=NRAND
NE V2=NRAND
NE V3=NRAND
NE V4=NRAND
NE V5=NRAND
NE V6=NRAND
NE V7=NRAND
NE V8=NRAND
NE V9=NRAND
NE V10=NRAND
NE V11=NRAND
NE V12=NRAND
NE V13=NRAND
NE V14=NRAND
NE V15=NRAND

NE X1=1.0000*V1
NE X2=0.0000*V1+1.0000*V2
NE X3=0.0000*V1+0.0000*V2+1.0000*V3
NE X4=0.0000*V1+0.0000*V2+0.0000*V3+1.0000*V4
```

NE X5=0.0000*V1+0.0000*V2+0.0000*V3+0.0000*V4+1.0000*V5
NE X6=0.0000*V1+0.0000*V2+0.0000*V3+0.0000*V4+0.0000*V5+1.0000*V6
NE X7=0.0000*V1+0.0000*V2+0.0000*V3+0.0000*V4+0.0000*V5+0.0000*V6+1.0000*V7
NE
X8=0.0000*V1+0.0000*V2+0.0000*V3+0.0000*V4+0.0000*V5+0.0000*V6+0.0000*V7+1.0000*V8
NE
X9=0.0000*V1+0.0000*V2+0.0000*V3+0.0000*V4+0.0000*V5+0.0000*V6+0.0000*V7+0.0000*V8+
1.0000*V9
NE
X10=0.0000*V1+0.0000*V2+0.0000*V3+0.0000*V4+0.0000*V5+0.0000*V6+0.0000*V7+0.0000*V8
+0.0000*V9+1.0000*V10
NE
X11=0.0000*V1+0.0000*V2+0.0000*V3+0.0000*V4+0.0000*V5+0.0000*V6+0.0000*V7+0.0000*V8
+0.0000*V9+0.0000*V10+1.0000*V11
NE
X12=0.0000*V1+0.0000*V2+0.0000*V3+0.0000*V4+0.0000*V5+0.0000*V6+0.0000*V7+0.0000*V8
+0.0000*V9+0.0000*V10+0.0000*V11+1.0000*V12
NE
X13=0.0000*V1+0.0000*V2+0.0000*V3+0.0000*V4+0.0000*V5+0.0000*V6+0.0000*V7+0.0000*V8
+0.0000*V9+0.0000*V10+0.0000*V11+0.0000*V12+1.0000*V13
NE
X14=0.0000*V1+0.0000*V2+0.0000*V3+0.0000*V4+0.0000*V5+0.0000*V6+0.0000*V7+0.0000*V8
+0.0000*V9+0.0000*V10+0.0000*V11+0.0000*V12+0.0000*V13+1.0000*V14
NE
X15=0.0000*V1+0.0000*V2+0.0000*V3+0.0000*V4+0.0000*V5+0.0000*V6+0.0000*V7+0.0000*V8
+0.0000*V9+0.0000*V10+0.0000*V11+0.0000*V12+0.0000*V13+0.0000*V14+1.0000*V15

CO all
SD V1-V15
OU RA=RU.txt WI=7 ND=3 XM IX=402