Numerical differences between Guttman's reliability coefficients and the GLB

Oosterwijk, P.R.; van der Ark, L.A.; Sijtsma, K.

DOI
10.1007/978-3-319-38759-8_12

Publication date
2016

Document Version
Final published version

Published in
Quantitative Psychology Research

Numerical Differences Between Guttman's Reliability Coefficients and the GLB

Pieter R. Oosterwijk, L. Andries van der Ark, and Klaas Sijtsma

Abstract For samples smaller than 1000 observations and tests longer than ten items, the greatest lower bound (GLB) to the reliability is known to be biased and not recommended as a method to estimate test-score reliability. As a first step in finding alternative lower bounds under these conditions, we investigated the population values of seven reliability coefficients: Coefficients λ_1, λ_2, λ_3 (a.k.a Cronbach's alpha), λ_4, λ_5, λ_6 and the GLB under varying correlational structures, and varying levels of number of items and item variances. Coefficients λ_2, λ_4 and λ_6 had population values closest to the GLB and may be considered as alternatives for the GLB in small samples. A necessary second step, investigating the behavior of these coefficients in samples, is a topic for future research.

Keywords Classical test theory • Greatest lower bound • Guttman's lambda coefficients • Reliability

1 Introduction

The purpose of this study was to compare seven methods for computing a test score's reliability. The methods are reliability coefficients proposed by Guttman (1945) and the greatest lower bound to the reliability (GLB; Bentler & Woodward 1980; Ten Berge, Snijders, & Zegers 1981; Woodhouse & Jackson 1977). Guttman's coefficients are known as λ_1 through λ_6, of which λ_3 equals the well known coefficient α (e.g., Cronbach 1951). Results in this study were obtained for the population (i.e., parameters).

In the population, the GLB is known to be closest to the reliability (Sijtsma 2009) of all lower bounds used in classical test theory (CTT). In samples, due to chance capitalization, the GLB is known to overestimate test-score reliability when

P.R. Oosterwijk (✉) • K. Sijtsma
Tilburg University, Tilburg, The Netherlands
e-mail: p.r.oosterwijk@gmail.com; K.Sijtsma@tilburguniversity.edu

L.A. van der Ark
University of Amsterdam, Amsterdam, The Netherlands
e-mail: L.A.vanderArk@uva.nl

L.A. van der Ark et al. (eds.), *Quantitative Psychology Research*, Springer Proceedings in Mathematics & Statistics 167, DOI 10.1007/978-3-319-38759-8_12

the sample size is smaller than 1000 and the test length exceeds ten items (Ten Berge & Sočan 2004). The question at hand is whether the λ coefficients should be recommended as alternative lower bounds under these conditions. A first step, in answering this question is to investigate whether in the population, the values of the λ coefficients are close enough to the GLB to be viable candidates. This step is investigated in this paper. A second step is to investigate the bias and variance of the λ coefficients in samples. In particular, λ_4, λ_5, λ_6 are the result of maximization procedures and may be prone to chance capitalization just like the GLB. This second step is currently being investigated by the authors.

It is known that λ_1 is smaller than λ_3, that λ_3 is smaller than λ_2, and that all three coefficients are smaller than the GLB, but the relationships of the other three λs with λ_1, λ_2, and λ_3, with the GLB, and with one another are either unknown or only known for special situations. Also, for most λs it is unknown how much their values differ from each other, and how much they differ from the GLB. This study discusses the mutual relationships between the seven methods at the theoretical level, and uses a computational study to focus on the issue of numerical differences between the seven coefficients.

In addressing the numerical differences between the λs and the GLB, we assumed that differences varied across different test and item properties. We investigated the influence of the following factors on the values of seven reliability methods and their mutual differences: (1) the variation of the item variances, (2) the dimensionality due to the correlational structure, and (3) the strength of the inter-item correlations. To investigate the effects of these factors, computational studies were used.

We performed four computational studies addressing the effect on reliability methods of: (1) Size of equal item variances and equal inter-item correlations representing one-dimensional item structures. This setup was a benchmark for the next three studies; (2) Spread of item variances while keeping inter-item correlations equal representing one-dimensional item structures; (3) Varying correlations between items from two different dimensions while correlations between items within dimensions were fixed. Results were presented for both correlation and covariance matrices; and (4) Varying within-dimension inter-item correlations while between-dimension inter-item correlations were fixed.

This article is organized as follows. We briefly discuss CTT (Lord & Novick 1968). The λ coefficients have been studied by Jackson and Agunwamba (1977). We briefly reiterate their line of reasoning as it greatly helps to understand each of the reliability methods and their mutual relationships. Next, we discuss the research method for the computational studies followed by the results, and finish by discussing the results and their implications for follow-up research.

2 Classical Test Theory

According to CTT, observed test score X can be decomposed into a true score T and measurement error E, such that $X = T + E$. Suppose a test consists of J items. Let

X_j be the score on item j, hence $X = \Sigma_{j=1}^{J} X_j$. Let σ_Y^2 denote the variance of Y, then it follows from the assumptions of classical test theory that the test-score reliability is defined as

$$\rho = \frac{\sigma_T^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2}. \tag{1}$$

In this definition, both σ_T^2 and σ_E^2 are unobservable, and this was the reason why psychometricians proposed several methods to approximate ρ on the basis of the inter-item covariance matrix obtained in a single test administration. Because of its brevity, in this article we use the notation of Jackson and Agunwamba (1977) which we introduce first.

Let $\sigma_{jj} = \sigma_{X_j}^2$ denote the observable item-score variance, σ_{jk} the inter-item covariance between items j and k, $t_j = \sigma_{T_j}^2$ the item true-score variance, $\sigma_{T_jT_k}$ the inter-item true-score covariance, $\theta_j = \sigma_{E_j}^2$ the item measurement-error variance, and $\sigma_{E_jE_k}$ the inter-item measurement-error covariance. Notice that $\sigma_{T_jT_k} = \sigma_{jk}$ and $\sigma_{E_jE_k} = 0$, for all pairs $j \neq k$. Covariance matrices $\mathbf{\Sigma}_X$ and $\mathbf{\Sigma}_T$ are $J \times J$ symmetrical matrices, whereas $\mathbf{\Sigma}_E$ is a $J \times J$ diagonal matrix. Matrix $\mathbf{\Sigma}_X$ is positive definite (pd), meaning that for any vector $\mathbf{u}$ of size J, we have $\mathbf{u}'\mathbf{\Sigma}_X\mathbf{u} > 0$ (i.e., the determinant of $\mathbf{\Sigma}_X$ is positive), and $\mathbf{\Sigma}_T$ and $\mathbf{\Sigma}_E$ are positive semi-definite (psd), that is, $\mathbf{u}'\mathbf{\Sigma}_T\mathbf{u} \geq 0$ and $\mathbf{u}'\mathbf{\Sigma}_E\mathbf{u} \geq 0$ (i.e., the two matrices' determinants are non-negative). It may be noted that

$$\mathbf{\Sigma}_X = \mathbf{\Sigma}_T + \mathbf{\Sigma}_E. \tag{2}$$

For example, for $J = 4$ Eq. (2) equals

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{pmatrix} = \begin{pmatrix} t_1 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & t_2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & t_3 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & t_4 \end{pmatrix} + \begin{pmatrix} \theta_1 & 0 & 0 & 0 \\ 0 & \theta_2 & 0 & 0 \\ 0 & 0 & \theta_3 & 0 \\ 0 & 0 & 0 & \theta_4 \end{pmatrix}. \tag{3}$$

Let $\mathbf{1}$ denote the unity vector of size J, then $\sigma_T^2 = \mathbf{1}'\mathbf{\Sigma}_T\mathbf{1}$, $\sigma_E^2 = \mathbf{1}'\mathbf{\Sigma}_E\mathbf{1}$, and the reliability definition in Eq. (1) may be written as

$$\rho = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\mathbf{1}'\mathbf{\Sigma}_T\mathbf{1}}{\sigma_X^2} = 1 - \frac{\mathbf{1}'\mathbf{\Sigma}_E\mathbf{1}}{\sigma_X^2} = 1 - \frac{\sum \theta_j}{\sigma_X^2}. \tag{4}$$

3 Guttman's Reliability Coefficients and the GLB

3.1 Guttman's Reliability Coefficients

The six λ coefficients (Guttman 1945) are lower bounds to the reliability. Each is derived from necessary (not: sufficient; hence, none provides the GLB) conditions for $\mathbf{\Sigma}_T$ to be psd. Jackson and Agunwamba (1977) provided derivations of these lower bounds, also illuminating some of the mutual relationships between the six λs and their relationship to the reliability. For each λ lower bound, we explain its formal basis and its definition, but we refer the reader to the original sources for more details about the steps leading from the formal basis to the specific λ definition. Logically, the definition of λ_3 precedes the discussion of λ_2.

Coefficient $\boldsymbol{\lambda}_1$. Because CTT also applies to item scores, we know that $\sigma_{jj} = t_j + \theta_j$, and because $t_j \geq 0$, it follows that $\theta_j \leq \sigma_{jj}$; thus, $\sum \theta_j \leq \sum \sigma_{jj}$. Hence, a most simple lower bound to the reliability is

$$\lambda_1 = 1 - \frac{\sum \sigma_{jj}}{\sigma_X^2}. \tag{5}$$

Coefficient λ_1 only exploits the information from $\mathbf{\Sigma}_T$ that $t_j \geq 0$ $(j = 1, \ldots, J)$. The other five λ coefficients extract more information from $\mathbf{\Sigma}_T$.

Coefficient $\boldsymbol{\lambda}_3$. Coefficient λ_3, also known as coefficient α (Cronbach 1951), uses information based on all 2×2 principal submatrices of $\mathbf{\Sigma}_T$, with diagonal elements t_j and t_k, and off-diagonal elements σ_{jk} and σ_{kj}; in particular, coefficient λ_3 uses the property

$$0 \leq \sigma^2_{X_j - X_k} = t_j + t_k - 2\sigma_{jk}, j \neq k. \tag{6}$$

Combining the sums for all $j \neq k$, λ_3 can be derived to be equal to

$$\lambda_3 = \lambda_1 + J^{-1}\lambda_1, \tag{7}$$

which is readily rewritten in its well known form as

$$\lambda_3 = \alpha = \frac{J}{J-1}\left(1 - \frac{\sum \sigma_{jj}}{\sigma_X^2}\right). \tag{8}$$

Coefficient λ_3 is a lower bound for reliability, ρ. Eq. (7) shows that as $J \rightarrow \infty$, we find that $\lambda_3 \rightarrow \lambda_1$. Obviously, for small J, λ_3 will clearly exceed λ_1 but the difference soon becomes smaller. However, for a 20-item test for which $\lambda_1 = 0.80$, we find that $\lambda_3 = 0.80 + 0.80/20 = 0.84$, a difference that still is worthwhile to report.

Coefficient $\boldsymbol{\lambda}_2$. Each of the 2×2 principal submatrices of $\mathbf{\Sigma}_T$ needs to be psd, but coefficient λ_3 does not use this property. The psd property implies that the submatrices' determinants must be non-negative; that is

$$t_j t_k \geq \sigma_{jk}^2, j \neq k, \tag{9}$$

and this result is used to derive

$$\lambda_2 = \lambda_1 + \frac{\left(J(J-1)^{-1} \sum \sum_{j \neq k} \sigma_{jk}^2\right)^{1/2}}{\sigma_X^2}. \tag{10}$$

Coefficient λ_2 is a lower bound for reliability, ρ, but usually it is not the greatest lower bound.

Coefficient $\boldsymbol{\lambda}_4$. Coefficient λ_4 exploits the information in $\boldsymbol{\Sigma}_T$ as follows. Let $\mathbf{u}$ be a J-vector with elements equal to either $+1$ or -1, and use the inequality $\mathbf{u}'\boldsymbol{\Sigma}_T\mathbf{u} \geq 0$. The effect of vector $\mathbf{u}$ in the quadratic form is to produce a sum of the J variances t_j and $J(J-1)$ covariances σ_{jk} in which each term has either positive or negative signs. For example, vector $\mathbf{u}' = (-1 \ +1 \ -1)'$ produces $\mathbf{u}'\boldsymbol{\Sigma}_T\mathbf{u} = t_1 + t_2 + t_3 - 2\sigma_{12} + 2\sigma_{13} - 2\sigma_{23} \geq 0$, and replacing $\mathbf{u}$ by $-\mathbf{u}$ yields the same result. Vector $\mathbf{u} = (+1 \ +1 \ +1)'$ yields $\mathbf{u}'\boldsymbol{\Sigma}_T\mathbf{u} = \sum_j t_j + \sum \sum_{j \neq k} \sigma_{jk} = \sigma_T^2 \geq 0$. Because each element of $\mathbf{u}$ has one of two possible values, in total 2^J different vectors $\mathbf{u}$ are possible, but because the effect of $\mathbf{u}$ and $-\mathbf{u}$ on the matrix product is the same, one may consider only 2^{J-1} vectors.

Starting from Eq. (2), one may write for any vector $\mathbf{u}$,

$$\mathbf{u}'\boldsymbol{\Sigma}_X\mathbf{u} = \mathbf{u}'\boldsymbol{\Sigma}_T\mathbf{u} + \mathbf{u}'\boldsymbol{\Sigma}_E\mathbf{u}. \tag{11}$$

Because $\mathbf{u}'\boldsymbol{\Sigma}_T\mathbf{u} \geq 0$, it follows that

$$\mathbf{u}'\boldsymbol{\Sigma}_E\mathbf{u} = \sum \theta_j \leq \mathbf{u}'\boldsymbol{\Sigma}_X\mathbf{u}. \tag{12}$$

Substituting $\mathbf{u}'\boldsymbol{\Sigma}_X\mathbf{u}$ for $\sum \theta_j$ in Eq. (4), a lower bound for the reliability ρ is obtained by finding $\mathbf{u}$ such that

$$\lambda_4 = \max_{\mathbf{u}} \left(1 - \frac{\mathbf{u}'\boldsymbol{\Sigma}_X\mathbf{u}}{\sigma_X^2}\right). \tag{13}$$

Jackson and Agunwamba (1977) showed that coefficient λ_4 equals the maximum value of coefficient λ_3 (equal to α) for the split of the test into two subtests that have test scores Y_1 and Y_2 (i.e., two test parts, not necessarily of equal length, such that $X = Y_1 + Y_2, J = 2$), based on the items that correspond to the positive and the negative elements in $\mathbf{u}$, respectively.

Coefficient $\boldsymbol{\lambda}_5$. Coefficient λ_5 exploits the result that covariances which are relatively large in absolute value put the main constraint on $\sum t_j$. Arbitrarily, let us assume that column k of $\boldsymbol{\Sigma}_X$ contains the relatively large covariances. Then, based on $t_j t_k \geq \sigma_{jk}^2$ [Eq. (9)] one can deduce

$$\sum t_j \geq 2 \left(\sum_{j \neq k} \sigma_{jk}^2 \right)^{1/2}. \tag{14}$$

Because CTT holds for individual items, so that $X_j = T_j + E_j$, and hence for the item-variance decomposition [Eq. (2)], such that $\sigma_{jj} = t_j + \theta_j$, it is also true that

$$\sum \sigma_{jj} = \sum t_j + \sum \theta_j. \tag{15}$$

Substituting $\sum t_j$ in Eq. (14) by the right-hand side of Eq. (15), and rearranging the terms produces

$$\sum \theta_j \leq \sum \sigma_{jj} - 2 \left(\sum_{j \neq k} \sigma_{jk}^2 \right)^{1/2}. \tag{16}$$

Then, substituting $\sum \theta_j$ in the numerator of Eq. (4) by the right-hand side of Eq. (16) yields a lower bound for ρ. To find the greatest value for λ_5, one determines Eq. (16) for each of the J columns in $\mathbf{\Sigma}_T$, thus letting k play the role of index (i.e., $k = 1, \ldots, J$), and defines coefficient λ_5 as

$$\lambda_5 = \lambda_1 + \max_k \frac{2 \left(\sum_{j \neq k} \sigma_{jk}^2 \right)^{1/2}}{\sigma_X^2}. \tag{17}$$

Coefficient $\boldsymbol{\lambda}_6$. Coefficient λ_6 is based on the multiple regression of an item score X_j on the other $J-1$ item scores. Let matrix $\mathbf{\Sigma}_{jj}$ denote the $(J-1) \times (J-1)$ covariance matrix without row j and column j, and let $\boldsymbol{\sigma}'_j = (\sigma_{jk})'$, $k \neq j$, denote the $J-1$ vector containing covariances involving item j but not σ_{jj}. Then, it can be shown that the residual variance $\epsilon_j^2 = \sigma_{jj} - \boldsymbol{\sigma}'_j \mathbf{\Sigma}_{jj}^{-1} \boldsymbol{\sigma}_j$, and that this provides an upper bound for measurement error, θ_j; that is, $\theta_j \leq \epsilon_j^2$. Substituting θ_j in the numerator of Eq. (4) by ϵ_j^2 yields lower bound

$$\lambda_6 = 1 - \frac{\sum \epsilon_j^2}{\sigma_X^2}. \tag{18}$$

3.1.0.1 The Greatest Lower Bound The observed covariance matrix $\mathbf{\Sigma}_X$ can be produced by many different matrices $\mathbf{\Sigma}_T$ and $\mathbf{\Sigma}_E$. The GLB (Bentler & Woodward 1980) is computed using an algorithm that maximizes the estimate of the trace of measurement-error matrix $\mathbf{\Sigma}_E$. The resulting matrix $\widetilde{\mathbf{\Sigma}}_E$ and its complement $\widetilde{\mathbf{\Sigma}}_T$, must be positive semi-definite such that $\mathbf{\Sigma}_X = \widetilde{\mathbf{\Sigma}}_T + \widetilde{\mathbf{\Sigma}}_E$. The GLB estimates reliability formulated as $\rho = 1 - \frac{\sigma_E^2}{\sigma_X^2}$ [Eq. (1)]. Using the assumptions of CTT, reliability can be written as

$$\rho = 1 - \frac{tr(\mathbf{\Sigma}_E)}{\sigma_X^2}. \tag{19}$$

The GLB is obtained by replacing $tr(\mathbf{\Sigma}_E)$ with $tr(\widetilde{\mathbf{\Sigma}}_E)$, resulting in

$$GLB = 1 - \frac{tr(\widetilde{\mathbf{\Sigma}}_E)}{\sigma_X^2}. \tag{20}$$

If all items in the test are essentially tau-equivalent (Lord & Novick 1968p. 90), the GLB is equal to the reliability; that is, GLB $= \rho$. The GLB provides the worst-case scenario for the reliability given the covariance matrice $\mathbf{\Sigma}_X$ (Sijtsma 2009). There are multiple ways to estimate the GLB. For details we refer to Bentler and Woodward (1980) and Ten Berge et al. (1981). We used the function glb.algebraic from the *psych* r-package (Revelle 2015) to obtain the GLB.

3.2 Relations Between Methods

We reiterate the relationships between the six λ coefficients and the GLB, and between these methods and reliability ρ.

1. $\lambda_1 \leq \lambda_3 \leq \lambda_2$ (proof, see Jackson & Agunwamba 1977). For finite J, $\lambda_1 < \lambda_3$; for $J \to \infty$, $\lambda_1 = \lambda_3$, but this does not happen in practice. Furthermore, $\lambda_3 \leq \lambda_2 \leq GLB \leq \rho$ with equality if the items are essentially tau-equivalent (Lord & Novick 1968p. 50); that is, $T = T + a_{jk}$, where a_{jk} is a scalar. In this case, coefficient λ_1 falls short of ρ by a factor $(J-1)/J$, because $\lambda_1 = [(J-1)/J]\lambda_3$.
2. Lord and Novick (1968pp. 93–94) showed that coefficient λ_4 is a higher lower bound than coefficient λ_3; that is, $\lambda_3 = \alpha \leq \lambda_4$. Jackson and Agunwamba (1977) derived the conditions for which $\lambda_4 = GLB$. In particular, if $\boldsymbol{\nu} = (\nu_1, \ldots \nu_J)$ is the vector with elements $\nu = -1, +1$ that maximizes λ_4, then the authors prove that: If $\lambda_4 = GLB$, then it must hold that $\theta_j = \nu_j(\mathbf{\Sigma}_X\boldsymbol{\nu})_j$, all j, where $(\mathbf{\Sigma}_X\boldsymbol{\nu})_j$ denotes the jth element of the column vector $\mathbf{\Sigma}_X\boldsymbol{\nu}$, provided (1) $\mathbf{\Sigma}_T = \mathbf{\Sigma}_X - \sum \theta_j$ is psd, and (2) $\theta_j \geq 0$, all j. If these conditions are satisfied, $\lambda_4 = GLB$.

   For real-data problems one has to check whether the GLB solution provides error variances $\theta_j = \nu_j(\mathbf{\Sigma}_X\boldsymbol{\nu})_j$, all j; the latter quantities $\nu_j(\mathbf{\Sigma}_X\boldsymbol{\nu})_j$ can be derived from $\boldsymbol{\nu}$, the item weights vector that produced λ_4, and the observable covariance matrix $\mathbf{\Sigma}_X$. The authors noticed that in general coefficient λ_4 is a lower bound for ρ but that it does not equal the GLB.
3. Jackson and Agunwamba (1977) derived conditions for which $\lambda_2 < \lambda_5$ and $\lambda_4 < \lambda_5$, but these inequalities are not true in general.

We are unaware of other relationships that have been demonstrated between the six λ coefficients. In addition to algebraic relations, considering graphically displayed relations based on a computational study may be worthwhile, because (1) the mutual relations between the λs are unknown for several λ pairs or only known under particular conditions that often are unfulfilled, and (2) except for λ_1 and λ_3, it is unknown how far the different λ values are apart, how far they are apart from the GLB, and whether differences are large enough to be of practical interest. In the next section, we discuss and present results of such a computational study.

4 Method

In this section, we discuss the setup of four computational studies. The purpose of these studies was to explore how different factors influence the distance between the coefficients λ_1 through λ_6 in relation to the GLB, and to study the numerical differences among λ coefficients. In each computational study, λ_1 through λ_6 and the GLB were computed using correlation matrices and covariance matrices. The λ coefficients provided values at most equal to the GLB.

Study 1 was a benchmark for the other studies. In Study 1, the values of the seven methods were computed using constructed correlation matrices containing equal item variances and equal inter-item correlations, which may follow from essential tau-equivalence. In Study 2, the item-score variances were varied while keeping the inter-item correlations equal and fixed. In studies 3 and 4, the item-score variances were varied, and inter-item correlations were varied so as to construct a two-dimensional item structure. Thus, the effect of multi-dimensionality on the reliability methods could be studied. In Study 3, correlations between items from different dimensions were varied and correlations between items within dimensions were fixed. In Study 4, by contrast, between-dimension inter-item correlations were fixed and within-dimension inter-item correlations were varied.

4.1 Study 1: Equal Correlations

For 4, 6, and 8 standardized items, we investigated inter-item correlation matrices in which all inter-item correlations ρ_{jk} were equal. Different correlation matrices were constructed by letting ρ_{jk} run from 0 to 1 in steps equal to 0.025; that is, $\rho_{jk} = 0, 0.025, \ldots, 1$. This resulted in 41 correlation matrices $\mathbf{R}_n$ $(n = 1, \ldots, 41)$. For each of three test lengths, $J = 4, 6, 8$, λs were computed for each of the 41 correlation matrices. For example, for $J = 4$ the correlation matrices equaled

$$\mathbf{R}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{R}_2 = \begin{pmatrix} 1 & 0.025 & 0.025 & 0.025 \\ 0.025 & 1 & 0.025 & 0.025 \\ 0.025 & 0.025 & 1 & 0.025 \\ 0.025 & 0.025 & 0.025 & 1 \end{pmatrix},$$

$$\ldots, \quad \mathbf{R}_{41} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

All items are standardized (i.e., $\mu_j = 0$ and $\sigma_{jj} = 1$), so that $\mathbf{\Sigma}_{Xn} = \mathbf{R}_n$.

4.2 Study 2: Varying Item-Score Variances

To isolate the effect of variation of item-score variances, inter-item correlations were fixed at $\rho_{jk} = 0.3$, which is a value typical of empirical test research. For example, the NEO-PR big-five personality inventory (McCrae & Costa 1999, retrieved from *Psych* package, Revelle 2015) reports mean inter-item correlations within each of the five facets equal to approximately 0.3. Test length equaled $J = 4, 6, 8$. For example, for $J = 4$, the correlation matrix $\mathbf{R}$ equaled

$$\mathbf{R} = \begin{pmatrix} 1 & 0.3 & 0.3 & 0.3 \\ 0.3 & 1 & 0.3 & 0.3 \\ 0.3 & 0.3 & 1 & 0.3 \\ 0.3 & 0.3 & 0.3 & 1 \end{pmatrix}. \tag{21}$$

Covariance matrices were constructed as follows. For $j = 1, 2, 3$, we chose $\sigma_{jj} = 1.5$, which is representative of 5-point Likert scales regularly found in psychological research. Across different covariance matrices, σ_{44} varied from 0.25 to 4.00 in steps equal to 0.05; that is, $\sigma_{44} = 0.25, 0.30, \ldots, 4.00$, resulting in 76 covariance matrices. Covariance matrices equaled

$$\mathbf{\Sigma}_{X_1} \approx \begin{pmatrix} 1.5 & 0.45 & 0.45 & 0.18 \\ 0.45 & 1.5 & 0.45 & 0.18 \\ 0.45 & 0.45 & 1.5 & 0.18 \\ 0.18 & 0.18 & 0.18 & 0.25 \end{pmatrix}, \quad \mathbf{\Sigma}_{X_2} \approx \begin{pmatrix} 1.5 & 0.45 & 0.45 & 0.20 \\ 0.45 & 1.5 & 0.45 & 0.20 \\ 0.45 & 0.45 & 1.5 & 0.20 \\ 0.20 & 0.20 & 0.20 & 0.30 \end{pmatrix},$$

$$\ldots, \quad \mathbf{\Sigma}_{X_{76}} \approx \begin{pmatrix} 1.5 & 0.45 & 0.45 & 0.73 \\ 0.45 & 1.5 & 0.45 & 0.73 \\ 0.45 & 0.45 & 1.5 & 0.73 \\ 0.73 & 0.73 & 0.73 & 4 \end{pmatrix}.$$

For 6 and 8 items, inter-item correlations equaled those used in [Eq. (21)]; that is, $\rho_{jk} = 0.3$. For $J = 6$, for the first four items, item-score variance $\sigma_{jj} = 1.5$ ($j = 1, \ldots, 4$), and for the last two items 5 and 6, σ_{55} and σ_{66} varied by increasing steps equal to 0.05, so that $\sigma_{55} = \sigma_{66} = c$, with $c = 0.25, 0.30, \ldots, 4.00$. For $J = 8$, the same numerical choices were made, keeping σ_{11} through σ_{55} equal to 1.5, and varying σ_{66}, σ_{77}, and σ_{88} by increasing steps equal to 0.05, starting at 0.25 and ending with 4.00, so that $\sigma_{66} = \sigma_{77} = \sigma_{88} = c$, with $c = 0.25, 0.30, \ldots, 4.00$.

4.3 Study 3: Two Dimensions, Varying Correlations Between Dimensions

In this example, the correlation matrices had a two-dimensional structure. Inter-item correlations were manipulated such that for different matrices the two dimensions either were negatively related, unrelated, positively related, or indistinguishable, thus representing one dimension. The range of correlations between items from different dimensions was based on the range of correlations between items from different facets of the NEO-PI-R available in the *Psych* package (Revelle 2015). In the first condition item variances were equal to 1, so the correlation and covariance matrix were equal. In the second condition the item variance varied. In the second example covariance matrices were based on the correlation matrices in the first example. This was done so as to create matrices resembling those found in empirical research.

For 4, 6, and 8 items, the inter-item correlations and the item variances were manipulated, resulting in 37 correlation and covariance matrices. Consider $J = 4$: A two-dimensional structure was constructed by dividing correlation matrix $\mathbf{R}$ into block $\mathbf{A}$ and block $\mathbf{B}$ [Eq. (22)], such that

$$\mathbf{R} = \left(\begin{array}{cc|cc} \rho_{11} & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{21} & \rho_{22} & \rho_{23} & \rho_{24} \\ \hline \rho_{31} & \rho_{32} & \rho_{33} & \rho_{34} \\ \rho_{41} & \rho_{42} & \rho_{43} & \rho_{44} \end{array}\right) = \left(\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{B} & \mathbf{A} \end{array}\right). \tag{22}$$

For all matrices, block $\mathbf{A}$ was a $J/2 \times J/2$ matrix with diagonal elements equal to 1 and off-diagonal elements equal to 0.6. Block $\mathbf{B}$ differed for all matrices. Matrix $\mathbf{B}_n$ $(n = 1, \ldots, 37)$ is a $J/2 \times J/2$ matrix with all off-diagonal elements equal to $\rho_{jk(n)} = -0.3225 + 0.0225 \times n$. For example, for $J = 4$ the correlation matrices $\mathbf{R}_n$ equaled

$$\mathbf{R}_1 = \begin{pmatrix} 1 & 0.6 & -0.3 & -0.3 \\ 0.6 & 1 & -0.3 & -0.3 \\ -0.3 & -0.3 & 1 & 0.6 \\ -0.3 & -0.3 & 0.6 & 1 \end{pmatrix}, \quad \mathbf{R}_2 \approx \begin{pmatrix} 1 & 0.6 & -0.28 & -0.28 \\ 0.6 & 1 & -0.28 & -0.28 \\ -0.28 & -0.28 & 1 & 0.6 \\ -0.28 & -0.28 & 0.6 & 1 \end{pmatrix},$$

$$\ldots, \quad \mathbf{R}_{37} = \begin{pmatrix} 1 & 0.6 & 0.6 & 0.6 \\ 0.6 & 1 & 0.6 & 0.6 \\ 0.6 & 0.6 & 1 & 0.6 \\ 0.6 & 0.6 & 0.6 & 1 \end{pmatrix}.$$

Correlation matrices for 6 and 8 items were constructed by extending blocks $\mathbf{A}$ and $\mathbf{B}$ by 1 or 2 rows and columns, respectively.

Covariance matrices were constructed as follows. For the three test lengths, the variances of the even numbered items equaled 1 and the variances of the odd numbered items equaled 2. For $J = 4$, we constructed 37 covariance matrices that equaled

$$\boldsymbol{\Sigma}_{X_1} \approx \begin{pmatrix} 2 & 0.85 & -0.6 & -0.6 \\ 0.85 & 1 & -0.6 & -0.6 \\ -0.6 & -0.6 & 2 & 0.85 \\ -0.6 & -0.6 & 0.85 & 1 \end{pmatrix}, \quad \boldsymbol{\Sigma}_{X_2} \approx \begin{pmatrix} 2 & 0.85 & -0.56 & -0.56 \\ 0.85 & 1 & -0.56 & -0.56 \\ -0.56 & 0.56 & 2 & 0.85 \\ -0.56 & 0.56 & 0.85 & 1 \end{pmatrix},$$

$$\ldots, \quad \boldsymbol{\Sigma}_{X_{37}} \approx \begin{pmatrix} 2 & 0.85 & 1.2 & 1.2 \\ 0.85 & 1 & 1.2 & 1.2 \\ 1.2 & 1.2 & 2 & 0.85 \\ 1.2 & 1.2 & 0.85 & 1 \end{pmatrix}.$$

4.4 Study 4: Two Dimensions, Varying Correlations Within Dimensions

As in Study 3, the coefficients were calculated using covariance matrices belonging to one of two conditions. In the first condition item variances were all equal to 1, so the correlation matrix and covariance matrix were equal. In the second condition the item variance varied. Again, the correlation matrices were divided into blocks A and B [Eq. (22)] but the inter-item correlations between the dimensions (block B in Study 3) were fixed and the inter-item correlations within the dimensions (block A in Study 3) were varied. This resulted in a $J/2 \times J/2$ matrix $\mathbf{A}_n (n = 1, \ldots, 41)$, with diagonal elements 1 and off-diagonal elements $\rho_n = -0.025 + 0.025 \times n$. All the elements of the $J/2 \times J/2$ matrix $\mathbf{B}$ are 0.1. This resulted in correlation matrices

$$\mathbf{R}_n = \left(\begin{array}{c|c} \mathbf{A}_n & \mathbf{B} \\ \hline \mathbf{B} & \mathbf{A}_n. \end{array}\right), \quad \text{for} \quad n = 1, \ldots, 41. \tag{23}$$

For $J = 4$, the correlation matrices $\mathbf{R}_n$ equaled

$$\mathbf{R}_1 = \begin{pmatrix} 1 & 0 & 0.1 & 0.1 \\ 0 & 1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 1 & 0 \\ 0.1 & 0.1 & 0 & 1 \end{pmatrix}, \quad \mathbf{R}_2 = \begin{pmatrix} 1 & 0.025 & 0.1 & 0.1 \\ 0.025 & 1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 1 & 0.025 \\ 0.1 & 0.1 & 0.025 & 1 \end{pmatrix},$$

$$\ldots, \quad \mathbf{R}_{41} = \begin{pmatrix} 1 & 1 & 0.1 & 0.1 \\ 1 & 1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 1 & 1 \\ 0.1 & 0.1 & 1 & 1 \end{pmatrix}.$$

Correlation matrices for $J = 6, 8$ were obtained by adding one or two rows and columns to blocks **A** and **B**, respectively.

Covariance matrices were constructed using the correlation matrices. Similar to Study 3, for the even numbered items the item-score variances equaled 1 and for the odd numbered items the item-score variances equaled 2. Using the 41 correlation matrices and the item-score variances, 41 covariance matrices $\boldsymbol{\Sigma}_{X_n}$ were constructed. For example, for $J = 4$, the matrices equaled

$$\boldsymbol{\Sigma}_{X_1} \approx \begin{pmatrix} 2 & 0 & 0.2 & 0.14 \\ 0 & 1 & 0.14 & 0.1 \\ 0.2 & 0.14 & 2 & 0 \\ 0.14 & 0.1 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\Sigma}_{X_2} \approx \begin{pmatrix} 2 & 0.04 & 0.2 & 0.14 \\ 0.04 & 1 & 0.14 & 0.1 \\ 0.2 & 0.14 & 2 & 0.04 \\ 0.14 & 0.1 & 0.04 & 1 \end{pmatrix},$$

$$\ldots, \quad \boldsymbol{\Sigma}_{X_{41}} \approx \begin{pmatrix} 2 & 1.41 & 0.2 & 0.14 \\ 1.41 & 1 & 0.14 & 0.1 \\ 0.2 & 0.14 & 2 & 1.41 \\ 0.14 & 0.1 & 1.41 & 1 \end{pmatrix}.$$

5 Results

5.1 Study 1: Equal Correlations

Figure 1 (left panel) shows that for fixed test length, reliability increases as inter-item correlations, ρ_{jk}, increase. This increase is faster for longer tests. By definition, λ_1 produced the lowest values of the λs, and the GLB produced the highest value. Because in each matrix **R** all inter-item correlations were equal, a necessary condition for essential tau-equivalence was satisfied; hence, λ_2, λ_3, λ_4 and GLB provided the same values. Equal inter-item correlations do not imply essential tau-equivalence; hence, λ_2, λ_3, λ_4 and GLB do not necessarily provide the reliability, ρ. At best λ_5 and λ_6 produced values that were lower than GLB by 0.04 and 0.01 units, respectively.

The difference between λ_6 on the one hand and λ_2, λ_3, λ_4 and the GLB on the other hand was smallest for the lowest and highest values of ρ_{jk}. As inter-item correlation ρ_{jk} increased, the difference between λ_1 and λ_5 on the one hand, and the GLB on the other hand increased. Method λ_5 was only closer to the GLB than λ_6 (at most by 0.01 units) for lower values of ρ_{jk} and the difference was greater as fewer items were used. When $\rho_{jk} = 1$, matrix **R** had determinant equal to 0; hence, λ_6 which uses the multiple regression model could not be computed.

For this study and the next three studies, method λ_1 not only was furthest from the GLB, but the distance was so large that λ_1 was useless compared to the other λs. Therefore, there is no discussion of the results for λ_1 in the remainder of this section. Results for λ_1 can be found in all figures.

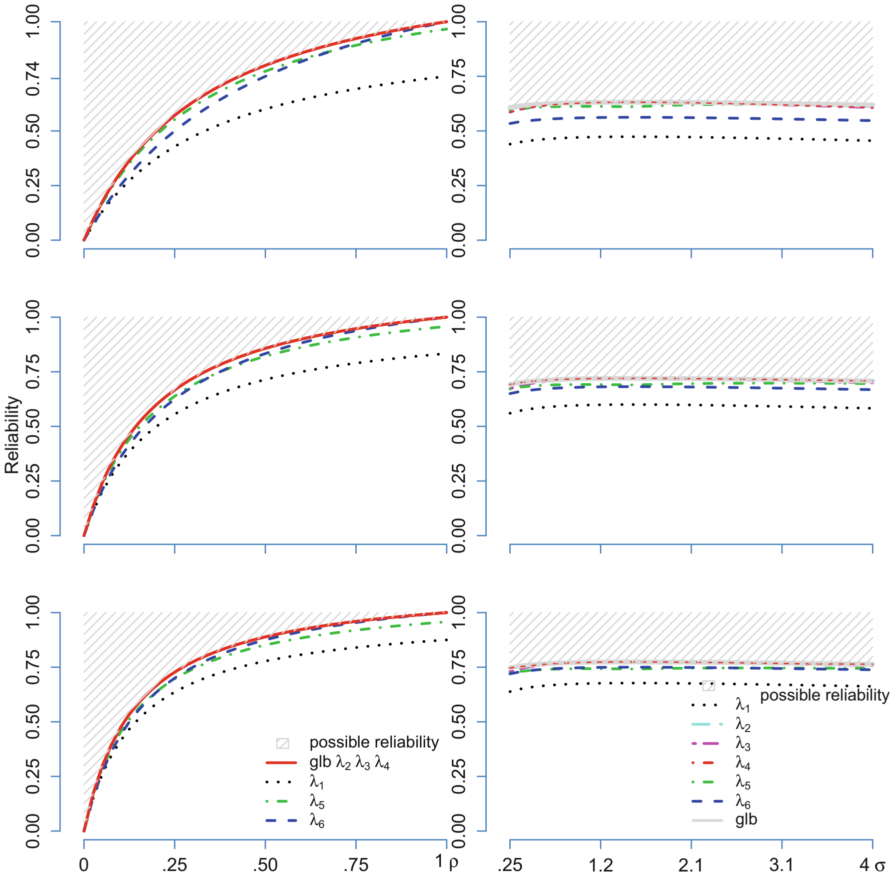

Fig. 1 Reliability coefficients as function of inter-item correlation ρ, or item variance σ, for $J = 4$ (*top*), $J = 6$ (*middle*), and $J = 8$ (*bottom*), with equal correlations (*left*) and varying item variances (*right*)

5.2 Study 2: Varying Item-Score Variances

Figure 1 (right panel) shows that the effect of manipulating the item variances on the differences between the λs and the GLB was small. The differences were approximately equal to the differences found in Study 1 for $\rho_{jk} = 0.3$. λ_2, λ_3, and λ_4 almost always yielded higher values than λ_5 and λ_6, except for a few conditions discussed in the next paragraph. λ_2, λ_3, and λ_4 differed equally from the GLB, but the difference was negligible, and was always smaller than 0.02. For $J = 4$, when the item variances differed the most, λ_2 produced slightly higher values than the other methods.

For $J = 4$, the four covariance matrices having the most extreme item-score variance (i.e., $\sigma_{44} = 0.25, 0.30, 3.95, 4.00$) produced the smallest difference between λ_5 and the GLB. The difference between λ_5 and the GLB was largest when item variances were equal. This results from λ_5 utilizing differences between columns of the covariance matrix to find the best possible estimate for item true-score variance (Verhelst 2000p. 7). Because the inter-item correlations in this study were equal, the differences between columns were smallest when item variances were identical.

Because the differences between methods λ_2 through λ_5 and the GLB were small, the effect of increasing test length was not clear-cut. For method λ_6, compared to manipulating item variance, increasing test length had a stronger effect. This can be understood from the regression model containing more predictors as tests grow longer, hence producing smaller residual item variances.

5.3 Study 3: Two Dimensions, Varying Correlations Between Dimensions

Figure 2 shows that for all λs the distance to the GLB was smaller as the inter-item correlations were more similar, thus causing the two-dimensional structure of the matrices to disappear. In most conditions, λ_4 was closest to the GLB (difference always < 0.08). Only when $J = 6$, all item variances equaled $\sigma_{jj} = 1$, and the between-dimension inter-item correlations were approximately $\rho_{jk} = 0$, the difference between λ_6 and the GLB was smaller than the difference between λ_4 and the GLB (at most 0.01).

In most conditions, λ_3 differed the most from the GLB. When all inter-item correlations were equal (i.e., $\rho_{jk} = 0.6$), it holds that $\lambda_2 = \lambda_3 = \lambda_4 = GLB$. When ρ_{jk} approached 0.6 from below, λ_3 eventually was closer to the GLB than λ_5 and λ_6 (at most 0.03 and 0.04, respectively). Figure 2 shows that as test length increased, the λ_3 curve intersected with the λ_5 and λ_6 curves at lower ρ_{jk} values.

Coefficients λ_2, λ_5, and λ_6 all had similar distances to the GLB, with distances between λ coefficients being more extreme as test length grew (Fig. 2). λ_6 was almost always closest to the GLB, except when $J = 4$ and approximately $\rho_{jk} = 0.6$. For all conditions, we found $\lambda_2 > \lambda_5$. Creating covariance matrices from the correlation matrices by increasing the variance of even numbered items by 1 was not sufficient to create a column in the covariance matrix with a sum of squared covariances larger than $\frac{J^2}{4}$ times the mean item variance (Verhelst 2000p. 8).

Differences between results from correlation matrices and results from covariance matrices were small. The two most noticeable differences were found for $J = 6$. The difference between both λ_4 and λ_5 and the GLB were notably smaller (0.04 and 0.03, respectively). Increasing item variances by 1 for uneven items did not produce differences between the columns of the covariance matrices that were large enough to result in favorable results for λ_5.

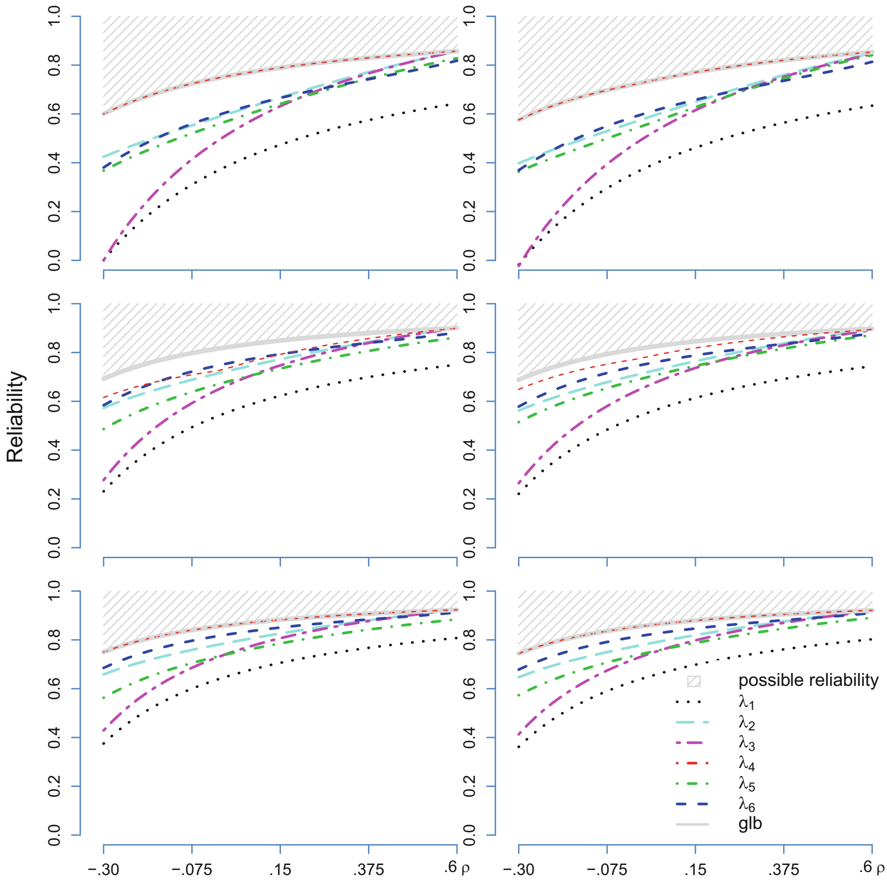

Fig. 2 Reliability coefficients for two-dimensional structure as a function of inter-item correlations (ρ) between dimensions, for $J = 4$ (*top*), $J = 6$ (*middle*), and $J = 8$ (*bottom*), with standardized items (*left*) and unstandardized items (*right*)

5.4 Study 4: Two Dimensions, Varying Correlations Within Dimensions

Figure 3 shows the results for the two-dimensional item structure when dimensions were weakly related. Similar to the previous studies, for most conditions λ_4 was closest to the GLB. Except when $J = 6$, for the top half of the within-dimension inter-item correlations (for inter-item correlations approximately larger 0.48), λ_6 outperformed λ_4. Compared to λ_4, λ_6 was closer to the GLB, and the difference between the λs and the GLB was greater as the correlation between dimensions increased (being 0.04 at its maximum). Also similar to Study 3, except for λ_5 differences between results for correlation matrices and covariance matrices were

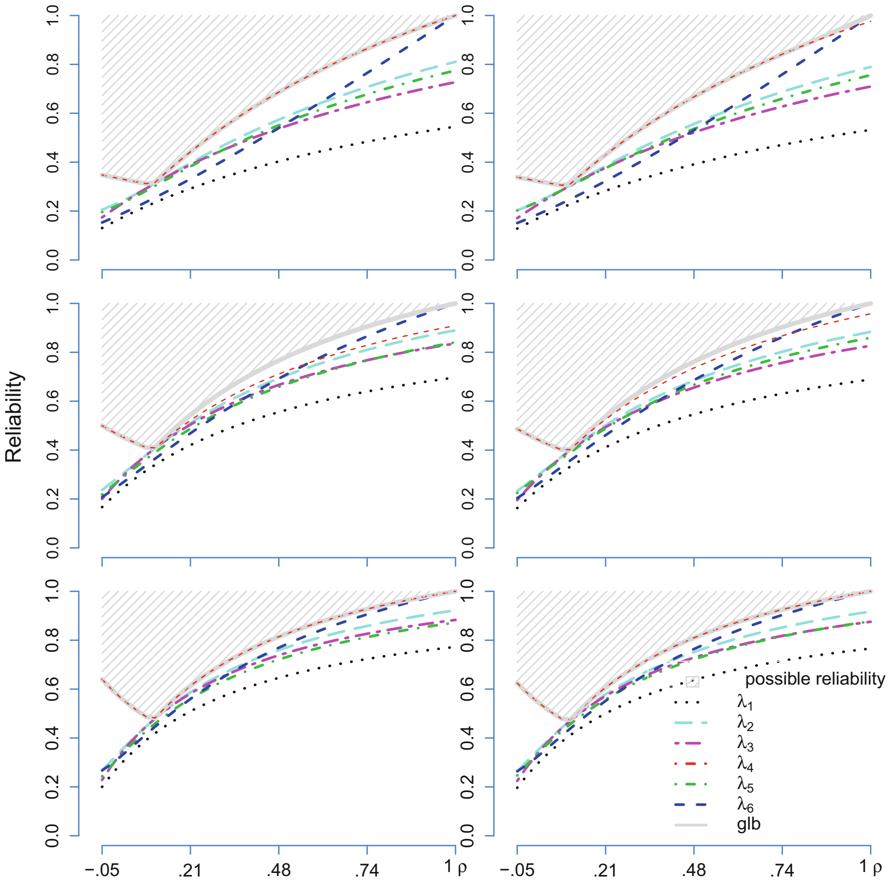

Fig. 3 Reliability coefficients for two-dimensional structure as a function of inter-item correlations (ρ) within dimensions, for $J = 4$ (*top*), $J = 6$ (*middle*), and $J = 8$ (*bottom*), with standardized items (*left*) and unstandardized items (*right*)

small. For $J = 6$ and $J = 8$, λ_5 produced higher values for the covariance matrices than for the correlation matrices but these higher values were not closer to the GLB than for example λ_4 and λ_6.

Of the remaining λs, λ_6 benefited most from higher within-dimension inter-item correlations. This result was found especially for the top half of the within-dimension inter-item correlations (again for inter-item correlations approximately larger than 0.48). Across all conditions, λ_2 was closer to the GLB than λ_3 and λ_5.

6 Discussion

None of the λs was closest to the GLB for all conditions discussed. However, compared to the other λs, in general method λ_4 was closest to the GLB. This result may have been facilitated by the structure of the correlation matrices that made selection of similar test halves easy. For 4 and 8 items and equal item variances this structure was perfect. Methods λ_1 and λ_3 are not serious competitors for the GLB. Method λ_1 not only is the smallest lower bound of the six λs but the difference with the other λs and the GLB is too large to be useful. Although generally much higher than λ_1, method λ_3 also appears rather useless, a result that has been discussed in different contexts (e.g., Cortina 1993; Cronbach 2004; Schmitt 1996; Sijtsma 2009; Zinbarg, Revelle, Yovel, & Li 2005).

Intuitively, method λ_5 might have been considered a good alternative to the GLB because of its capacity to cope with variation within the covariance matrix. However, even though the computational examples in this study may be considered rather representative of data structures typically encountered in psychological research, λ_5's performance was worse than that of the other methods (except λ_1). For all λs, in general differences between results for covariance matrices and correlation matrices caused by varying item variance were modest to small.

For small to moderate samples not containing more than 1000 cases, the GLB suffers from strong positive sampling bias (Ten Berge & Sočan 2004) and alternative methods may be considered. Candidates replacing the GLB for small to moderate samples are λ_2, λ_4 and λ_6. Only when differences in item variance are large and inter-item correlations are very similar is λ_5 a viable candidate. For λ_4 results are available showing bias is likely to be small for values greater than 0.85, test length smaller than 25 items and sample size greater than 3000 (Benton 2015). Research addressing the sampling variance of these methods is needed and we are currently studying this issue.

References

Bentler, P. M., & Woodward, J. A. (1980). Inequalities among lower bounds to reliability: With applications to test construction and factor analysis. *Psychometrika*, *45*, 249–267.

Benton, T. (2015). An empirical assessment of Guttman's lambda 4 reliability coefficient. In R. E. Millsap, D. M. Bolt, L. A. van der Ark, & W. -C. Wang (Eds.), *Quantitative psychology research: The 78th annual meeting of the Psychometric Society* (pp. 301–310). New York, NY: Springer.

Cortina, J. M. (1993). What is coefficient alpha? an examination of theory and applications. *Journal of Applied Psychology*, *78*, 98–104.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Cronbach, L. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, *64*, 391–418.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*, 255–282.

Jackson, P. H., & Agunwamba, C. C. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: I: Algebraic lower bounds. *Psychometrika*, *42*, 567–578.

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

McCrae, R. R., & Costa, P. T. (1999). A five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 139–153). New York: Guilford Press.

Revelle, W. (2015). *Psych: Procedures for personality and psychological research Version 1.5.8 [computer software].* Evanston, IL. Retrieved from http://CRAN.R-project.org/package=psych.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*, 350–353.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*, 107–120.

Ten Berge, J. M. F., Snijders, T. A. B., & Zegers, F. E. (1981). Computational aspects of the greatest lower bound to the reliability and constrained minimum trace factor analysis. *Psychometrika*, *46*, 201–213.

Ten Berge, J. M. F., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, *69*, 613–625.

Verhelst, N. (2000). *Estimating the reliability of test from single test administration.* Unpublished report. Arnhem, The Netherlands: Cito. Retrieved from http://www.cito.com/research_and_development/psychometrics/~/media/cito_com/research_and_development/publications/cito_report98_2.ashx.

Woodhouse, B., & Jackson, P. H. (1977). Lower bounds for the reliability of a test composed of nonhomogeneous items II: A search procedure to locate the greatest lower bound. *Psychometrika*, *67*, 251–259.

Zinbarg, R., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω_w: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*, 122–133.