PSYCHOMETRIKA—VOL. 69, NO. 4, 613–625
DECEMBER 2004

THE GREATEST LOWER BOUND TO THE RELIABILITY OF A TEST AND THE HYPOTHESIS OF UNIDIMENSIONALITY

JOS M.F. TEN BERGE
GREGOR SOČAN

UNIVERSITY OF GRONINGEN

To assess the reliability of congeneric tests, specifically designed reliability measures have been proposed. This paper emphasizes that such measures rely on a unidimensionality hypothesis, which can neither be confirmed nor rejected when there are only three test parts, and will invariably be rejected when there are more than three test parts. Jackson and Agunwamba's (1977) greatest lower bound to reliability is proposed instead. Although this bound has a reputation for overestimating the population value when the sample size is small, this is no reason to prefer the unidimensionality-based reliability. Firstly, the sampling bias problem of the glb does not play a role when the number of test parts is small, as is often the case with congeneric measures. Secondly, glb and unidimensionality based reliability are often equal when there are three test parts, and when there are more test parts, their numerical values are still very similar. To the extent that the bias problem of the greatest lower bound does play a role, unidimensionality-based reliability is equally affected. Although unidimensionality and reliability are often thought of as unrelated, this paper shows that, from at least two perspectives, they act as antagonistic concepts. A measure, based on the same framework that led to the greatest lower bound, is discussed for assessing how close is a set of variables to unidimensionality. It is the percentage of common variance that can be explained by a single factor. An empirical example is given to demonstrate the main points of the paper.

Key words: Reliability, congeneric test, unidimensionality of a test.

It is a well-known fact that the reliability of a test, defined as the ratio of true score to observed score variance, cannot generally be determined from a single test administration, but requires the use of a parallel test. More often than not, parallel tests are not available. In such cases, two approaches are popular to obtain indirect information on the reliability of the test: Either lower bounds to reliability can be used, or one may resort to hypotheses about the nature of the test parts.

Evaluating lower bounds to the reliability, such as Guttman's λ_3 (Guttman, 1945), also known as coefficient alpha (Cronbach, 1951) has gained wide popularity. A lower bound that is nearly always better than alpha is Guttman's λ_4. It is the highest alpha that can be obtained by splitting up the items in two parts (not necessarily of equal numbers) and treating those two parts as novel "items." Jackson and Agunwamba (1977) proposed the greatest lower bound (glb) to reliability. It exceeds all other lower bounds by using the available information implied by the observed covariance matrix exhaustively.

When lower bounds are high enough, the reliability has been shown adequate by implication. However, when lower bounds are low, they are of no avail. Also, lower bounds to reliability involve some degree of sampling bias. This is well known for alpha, λ_4, and glb. Whereas alpha in small samples tends to slightly underestimate the population value (Feldt, Woodruff, & Salih, 1987; Van Zijl, Neudecker, & Nel, 2000; Yuan & Bentler, 2002), λ_4 and glb may grossly overestimate the population values when computed in small samples (Verhelst, 1998; Shapiro & Ten Berge, 2000).

Request for reprints should be sent to Jos Ten Berge, Department of Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, Netherlands. E-mail: tberge@ppsw.rug.nl.

The authors are obliged to Henk Kiers for commenting on a previous version. Gregor Sočan is now at the University of Ljubljana.

0033-3123/2004-4/2002-0997-A $00.75/0

To avoid these problems, it is tempting to follow a different approach by introducing hypotheses on the nature of the test parts, from which the reliability can be determined at once. For instance, when the test parts are essentially tau-equivalent, which means that their true scores are equal up to an additive constant, coefficient alpha is the reliability. A major limitation of essential tau-equivalence is that it requires equal covariances between test parts (a necessary, but not sufficient condition), which will never be encountered in practice (unless k, the number of test parts, is 2).

A weaker, and far more popular hypothesis, is that of a congeneric test, consisting of k test parts with perfectly correlated true scores. Equivalently, the test parts are assumed to fit a one-factor model. We shall refer to this assumption as the hypothesis of unidimensionality. For the case $k = 3$, Kristof (1974) has derived a closed-form expression for the reliability of the test, based on this hypothesis. For $k > 3$, coefficients based on common factor analysis have been proposed that can be seen as generalized versions of Kristof's reliability coefficient. We shall focus our attention on a coefficient proposed by McDonald (1970). Coefficients like Kristof's and McDonald's are often considered useful alternatives to lower bounds like glb, because they aim to estimate, rather than underestimate, reliability, and lack any reputation of sampling bias.

The main point of the present paper is that, when dealing with congeneric tests, the greatest lower bound is to be preferred to reliability coefficients based on the hypothesis of unidimensionality. The first reason is that the hypothesis is either beyond verification or it is false, which jeopardizes the very basis from which the corresponding reliability coefficients were derived. The second reason is that the sampling bias problems of the glb are of no impact when the number of test parts is small, and when the number is large, the reliability coefficients by Kristof and McDonald have exactly the same bias problem. In fact, when $k = 3$, Kristof's coefficient is often identical to the glb, as will be proven below, and for $k > 3$, glb and McDonald's coefficient typically behave very much the same (a fact to be demonstrated and explained below). Because perfect unidimensionality does not arise when $k > 3$, assessing how *close* is a given test to unidimensionality is far more interesting than testing *whether or not* the test is unidimensional. It is the second purpose of the present paper to describe a method for this, based on factor analysis, fully coherent with the Jackson and Agunwamba framework that was used to derive the greatest lower bound to reliability. It is the percentage of common variance that can be explained by a single factor.

The organization of this paper is as follows. First, we discuss λ_3 (alpha) and λ_4, and explain the rationale behind the glb, relying on the concept of the set of all possible solutions. Next, reliability coefficients based on the unidimensionality hypothesis are discussed for $k = 3$ (Kristof) and for $k > 3$ (McDonald). It is proven that Kristof's coefficient is often the glb, and that it always exceeds λ_4. It will be explained that unidimensionality usually can neither be proved nor disproved when $k = 3$. From the literature on maximal rank reduction of a covariance matrix, it appears that the unidimensionality hypothesis is untenable when $k > 3$. The question of how to evaluate "closeness to unidimensionality" will be addressed, in terms of the same theoretical framework that led up to the greatest lower bound to reliability. This results in the percentage of common variance that can be explained by a single factor as a natural measure of unidimensionality. In passing, unidimensionality and reliability appear to be antagonistic concepts. Finally, an empirical data set, expected to be close to unidimensional, will be analyzed. In particular, it will be shown and explained that McDonald's reliability coefficient, when applied to such data, will behave very much like the glb in every respect.

Guttman's λ_3 (Coefficient alpha), λ_4, and the glb

Lower bounds to reliability start from the assumption that each test part X_i consists of a true score T_i and error E_i, that is,

$$X_i = T_i + E_i, \tag{1}$$

$i = 1, \ldots, k$, the error term E_i only correlating with itself and with $X_i, i = 1, \ldots, k$. Throughout this paper, we maintain this assumption of "linearly independent errors." In cases where items of a test are experimentally dependent, the assumption does not make sense, and collapsing such items into one test part would be indicated (Cronbach, 1988). The reliability of test X is defined as

$$\frac{\text{Var}\,(T)}{\text{Var}\,(X)} = 1 - \frac{\text{Var}\,(E)}{\text{Var}\,(X)}, \tag{2}$$

where $X = X_1 + \cdots + X_k$, $T = T_1 + \cdots + T_k$, and $E = E_1 + \cdots + E_k$. Under these classical assumptions, one may express the relation between alpha and the reliability by the formula

$$\alpha + \frac{1}{(k-1)\text{Var}\,(X)} \sum_{i<j} \text{Var}\,(T_i - T_j) = \text{reliability}, \tag{3}$$

see Appendix A. Formula (3) gives a quick summary of two well-known results. First, it reveals that alpha is a lower bound to the reliability, as we know from Guttman (1945). In addition, it shows that alpha is the reliability if and only if $\text{Var}\,(T_i - T_j) = 0$ for all $i \neq j$ (the condition of essential tau-equivalence) as has been proven by Novick and Lewis (1967). Because equal covariances between all test parts are necessary for this, equality in (3) will never hold in practice, whence alpha is best thought of as a lower bound to reliability.

The definitions of true scores and errors remain valid when the test parts are added together in two groups. Treating these two groups as two "new" test parts Y_1 and Y_2, the alpha of $X = Y_1 + Y_2$ can be found to be $\alpha_Y = 4\sigma(Y_1, Y_2)/\sigma^2(X)$, another lower bound to reliability. Of course, α_Y will vary according to the two-split adopted. The maximum value of α_Y over all two-splits is λ_4 (Guttman, 1945). Usually, λ_4 is a better lower bound than alpha, but exceptions may be encountered when k is odd.

Jackson and Agunwamba noted that all lower bounds to reliability use *part of the information* implicit in the classical assumptions. For instance, alpha uses nonnegativity of all variances of true score differences $(T_i - T_j)$, as is clear from (3); λ_4 uses nonnegativity of the true score difference of Y_1 and Y_2, which is equivalent to demanding that $\text{Var}\,(w_1T_1 + \cdots + w_kT_k)$ be nonnegative, where w_j is either $+1$ or -1, and $+1$ and -1 occur at least once. Jackson and Agunwamba went on to ask the fundamental question of how to use *all information* implied by the classical assumptions. Obviously, nonnegativity of $\text{Var}\,(w_1T_1 + \cdots + w_kT_k)$ should hold *for every set of weights*, which means that $\mathbf{C}_T$, the covariance matrix of the true scores, should be positive semi-definite (psd), that is, it cannot have a negative eigenvalue. In addition, the zero covariance of error terms with true scores and with errors of other test parts implies that the observed covariance matrix $\mathbf{C}_X$ can be decomposed as $\mathbf{C}_X = \mathbf{C}_T + \mathbf{C}_E$, where the error covariance matrix $\mathbf{C}_E$ is diagonal, and also psd. To set the stage for the glb, Jackson and Agunwamba thus arrived at the set of all possible states of nature: It is the set of all nonnegative diagonal matrices $\mathbf{C}_E$ for which $\mathbf{C}_X - \mathbf{C}_E$ has no negative eigenvalue. The set is convex, and part of it can be pictured as in Figure 1.

The area below the curve represents the set of all cases where $\mathbf{C}_T - \mathbf{C}_E$ is psd. The marked area is the intersection of that set with the set of nonnegative diagonal matrices $\mathbf{C}_E$. This intersection is the set of possible states of nature. It contains all diagonal matrices $\mathbf{C}_E$ that are compatible with the classical definitions.

By mapping the set of possible $\mathbf{C}_E$ into the interval [0, 1] by the reliability function $r_{XX} = 1 - \text{tr}\,(\mathbf{C}_E)/\text{Var}\,(X)$, that interval is split in an area of possible versus impossible values for the reliability. Clearly, because $\mathbf{C}_E = \mathbf{O}$ (the origin) belongs to the set (note that $\mathbf{C}_X - \mathbf{O} = \mathbf{C}_X$ is

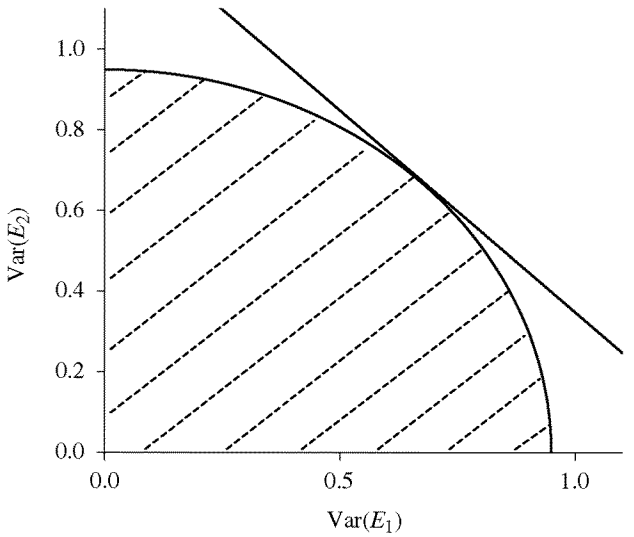

FIGURE 1.
An example of the admissible set ($k = 2$).

psd), a reliability as high as 1 can never be ruled out on the basis of a single test administration. On the other hand, the worst possible situation for reliability is when the sum of error variances is a maximum over the set of possible $\mathbf{C}_E$. This point is where the tangent plane orthogonal to the vector of ones, that is, the plane defined by a constant sum of error variances, touches the possible set (when $k = 2$, the tangent plane becomes the tangent line with angle $-45°$). It defines the worst-case scenario for reliability. It has the largest possible sum of error variances under full consideration of all information. Accordingly, the glb is defined as the reliability value associated with that point. Hence, in the [0, 1] interval, the possible reliability values are in the range [glb, 1], and the impossible reliability values are in the range [0, glb). All lower bounds (except when they happen to coincide with the glb) are in the interval [0, glb) of impossible reliability values (or they are impossible because they are negative).

It may be counterintuitive that respectable lower bounds such as alpha are outside the possible set. An example may clarify this. Suppose we have three test parts, with covariance matrix

$$\mathbf{C}_X = \begin{bmatrix} 3 & 1 & 2 \\ 1 & 4 & 3 \\ 2 & 3 & 6 \end{bmatrix}. \tag{4}$$

As will be explained below, see (7), the glb solution for this case implies true score variances 1, 2, and 5, whence

$$\mathbf{C}_T = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 5 \end{bmatrix} \quad \text{and} \quad \mathbf{C}_E = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{5}$$

and the glb is

$$1 - \frac{\text{tr}\,(\mathbf{C}_E)}{\text{Var}\,(X)} = 1 - \frac{5}{25} = .8.$$

Alpha, on the other hand, is

$$\frac{3}{2}\left(1-\frac{13}{25}\right)=1-\frac{7}{25}=.72.$$

Because 25 is the observed test variance, alpha implies a sum of error variances 7, whence the implied sum of the three true score variances is $13-7=6$. To see that this is incompatible with the classical definitions, consider the implied true score covariance matrix $\mathbf{C}_T=\mathbf{C}_X-\mathbf{C}_E$, of the form

$$\mathbf{C}_T=\begin{bmatrix}? & 1 & 2\\ 1 & ? & 3\\ 2 & 3 & ?\end{bmatrix}. \tag{6}$$

The diagonal elements are unknown, but their sum is 6. This would mean that the variance of $(T_1+T_2-T_3)=6+2\times(1-2-3)=-2$, which is impossible. This shows that alpha implies error variances that are too large to be true. They correspond to a point outside the possible set. Failure to use all available information causes alpha to be smaller than the glb.

How to Compute the glb

In general, an iterative procedure is needed to find the glb. Bentler (1972) has proposed an algorithm for a lower bound to reliability that is, in fact, the glb whenever the constraint of nonnegative error variances is inactive. A generalized iterative algorithm by Bentler and Woodward (1980), also see Ten Berge, Snijders and Zegers (1981), will find the glb in all cases. Although local minima may occur, they can be detected by a necessary and sufficient condition for the global optimum (Ten Berge et al., 1981), and the program will restart whenever that condition is not met. This approach has yet to fail. A program (TiaPlus) can be obtained from `TON.HEUVELMANS@CITOGROEP.NL`. The glb algorithm together with Bentler's (1972) algorithm and Shapiro's weighted glb (Shapiro, 1982b) has also been included in EQS 6.0. Finally, the factor analysis program mrfa2 by Henk Kiers also gives the glb. The program and documentation can be downloaded from `ppswmm.ppsw.rug.nl/~kiers/main.html`.

Closed-form solutions for the glb only exist for $k=2$ or 3. When $k=2$, the glb will often be alpha. For $k=3$, a closed-form solution exists, which is valid only when the constraint of nonnegative error variances is inactive; see Woodhouse and Jackson (1977, Table 4.1). When cases of negative covariances between test parts are ignored, and test parts are arranged such that $\sigma(1,2)\le\sigma(1,3)\le\sigma(2,3)$, that solution is as follows:

When

$$\sigma_{12}\ge\frac{\sigma_{13}\sigma_{23}}{\sigma_{13}+\sigma_{23}},$$

the glb is

$$\frac{(\sigma_{12}\sigma_{13}+\sigma_{12}\sigma_{23}+\sigma_{13}\sigma_{23})^2}{\sigma_{12}\sigma_{13}\sigma_{23}\mathrm{Var}\,(X)}, \tag{7a}$$

based on true score variances

$$\sigma^2(T_1)=\sigma(1,2)\sigma(1,3)/\sigma(2,3),$$
$$\sigma^2(T_2)=\sigma(1,2)\sigma(2,3)/\sigma(1,3),$$
$$\sigma^2(T_3)=\sigma(1,3)\sigma(2,3)/\sigma(1,2).$$

When

$$\sigma_{12} < \frac{\sigma_{13}\sigma_{23}}{\sigma_{13} + \sigma_{23}}$$

the glb is

$$\frac{4(\sigma_{13} + \sigma_{23})}{\text{Var}\,(X)}, \tag{7b}$$

based on true score variances

$$\sigma^2(T_1) = \sigma(1,3) - \sigma(1,2),$$
$$\sigma^2(T_2) = \sigma(2,3) - \sigma(1,2),$$
$$\sigma^2(T_3) = \sigma(1,3) + \sigma(2,3).$$

In retrospect, it can now be verified that (5) was an application of (7b). Woodhouse and Jackson (1977, p. 584) noted that, when $\sigma(1,2) \leq \sigma(1,3) \leq \sigma(2,3)$, (7b) is λ_4. It implies a rank-2 solution for $\mathbf{C}_T$. An example where (7a) is the glb is when

$$\mathbf{C}_{X*} = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 4 & 3 \\ 2 & 3 & 6 \end{bmatrix}. \tag{8}$$

The true score variances 4/3, 3, and 3 yield a glb of .79. On the other hand, λ_4 is only .74 (smaller yet than $\alpha = .78$). It is not a possible reliability value in this case, because the implied $\mathbf{C}_T$ would have a negative eigenvalue.

The true score variances of (7a) always generate a solution with $\mathbf{C}_T$ of rank one. However, even when the condition for (7a) to be the glb is met, this need not correspond to a possible solution: When at least one of the implied error variances is negative, we are outside the possible set, and (7a) does not apply. For instance, suppose we change the second diagonal element in $\mathbf{C}_{X*}$ of (8) to have

$$\mathbf{C}_{X+} = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 2.8 & 3 \\ 2 & 3 & 5 \end{bmatrix}. \tag{9}$$

The solution (7a) still implies a true score variance of 3 for the second test part, which exceeds the observed variance of that test part. In such cases, (7a) does not apply, and the associated glb has to be evaluated numerically. It is .8422 (based on error variances 1.4286, 0, and 2.6429), whereas (7a) now gives pseudo error variances of 5/3, −.20, and 3, implying an impossibly low reliability value of .8269.

Apart from cases of negative error variances, the rank one solution also fails when one of the three observed covariances is negative. In that case, $\mathbf{C}_T$ will have a negative diagonal element, which is incompatible with having all eigenvalues nonnegative. We shall ignore such cases, because they can only meaningfully arise by sampling error.

Reliability Coefficients under the Hypothesis of Unidimensionality

Kristof (1974) has shown that, when a test is composed of three congeneric test parts, the reliability is

$$r(X,T)^2 = \frac{1}{\text{Var}\,(X)}\left(\frac{\sigma_{12}\sigma_{13}}{\sigma_{23}} + \frac{\sigma_{12}\sigma_{23}}{\sigma_{13}} + \frac{\sigma_{13}\sigma_{23}}{\sigma_{12}} + 2\sigma_{12} + 2\sigma_{13} + 2\sigma_{23}\right). \tag{10}$$

Upon rewriting this expression as one with denominator $\sigma_{12}\sigma_{13}\sigma_{23}\text{Var}(X)$, it is trivial that (10) is identical to (7a), a fact that seems to have gone unnoticed. Another fact about Kristof's coefficient, equally unnoticed, is that Kristof's coefficient is always at least as large as (7b), the glb when it happens to coincide with λ_4. Specifically, from $(\sigma_{12}\sigma_{13}+\sigma_{12}\sigma_{23}-\sigma_{13}\sigma_{23})^2 \geq 0$, we have $(\sigma_{12}\sigma_{13}+\sigma_{12}\sigma_{23}+\sigma_{13}\sigma_{23})^2 \geq 4\sigma_{12}\sigma_{13}\sigma_{23}(\sigma_{13}+\sigma_{23})$. Using that $\lambda_4 = 4(\sigma_{13}+\sigma_{23})/\text{Var}(X)$ when $\sigma(1,2) \leq \sigma(1,3) \leq \sigma(2,3)$, it is clear that λ_4 cannot exceed Kristof's coefficient. The implications are twofold. When the glb is λ_4, Kristof's reliability gives just a possible value of the reliability, in between λ_4 and 1. When the glb is Kristof's reliability, λ_4 gives an impossible value of the reliability, because a lower bound less than the glb implies error variances too large to be true.

When $k > 3$, and the hypothesis of unidimensionality holds, it is possible to find a decomposition $\mathbf{C}_X = \mathbf{C}_T + \mathbf{C}_E$ with $\mathbf{C}_T$ of rank one. Then $\mathbf{C}_T$ can be factored as $\mathbf{aa}'$, for some vector $\mathbf{a}$, and the reliability is the sum of elements of $\mathbf{C}_T$, divided by $\text{Var}(X)$. This gives, as a direct generalization of Kristof's coefficient, the reliability

$$(\mathbf{1}'\mathbf{C}_T\mathbf{1})/\text{Var}(X) = (\mathbf{1}'\mathbf{aa}'\mathbf{1})/\text{Var}(X) = (\mathbf{1}'\mathbf{a})^2/\text{Var}(X), \tag{11}$$

see, for example, McDonald (1970).

The use of coefficients (10) or (11) is warranted to the extent that the hypothesis of unidimensionality is valid. We shall say that unidimensionality holds "patently" when the observed data admit a rank-one factor solution; that is, when a nonnegative diagonal matrix $\mathbf{C}_E$ exists such that $\mathbf{C}_T = \mathbf{C}_X - \mathbf{C}_E$ has rank one. We shall say that it holds "latently" when there is only one underlying factor. For instance, an IQ test with test parts verbal IQ, numerical IQ, and spatial IQ will be latently three-dimensional, yet patently unidimensional whenever the true score variances of (7a) imply a possible solution. Clearly, patent unidimensionality is necessary but not sufficient for latent unidimensionality.

The conditions under which patent unidimensionality is possible have been known for a long time. Spearman (1927) already noted that unidimensionality is impossible when $k > 3$. This is easily verified. When a 4×4 covariance matrix $\mathbf{C}_X - \mathbf{C}_E$ is of rank 1, every 2×2 submatrix must also be of rank 1. This means, among other things, that $\sigma_{13}\sigma_{24} - \sigma_{23}\sigma_{14}$ must vanish, an event of probability zero. Accordingly, when $k > 3$, factor analysis with only one common factor will never give perfect fit. More generally, Wilson and Worcester (1939), Guttman (1958), and Bekker and De Leeuw (1987) have argued that rank-reduction of a covariance matrix by reducing diagonal elements does not carry a long way. Shapiro (1982a) has proven that the minimal reduced rank will be at or above the Ledermann bound (Ledermann, 1937) almost surely. It means that the minimal reduced rank is almost surely at or above 1 when $k = 3$, at or above 2 when $k = 4$, at or above 3 when $k = 5$ or 6, and so on. The notion of "almost surely" reflects the fact that although covariance matrices that do have low reduced rank are easily constructed, they will never be encountered in practice.

The first implication of the above results is that patent unidimensionality is impossible when $k > 4$. It will be rejected by any statistical test provided with enough power. This means that latent unidimensionality is also untenable. Therefore, the rationale underlying (11) is in jeopardy when $k > 3$.

For $k = 3$, the situation is different. When the three test parts are positively correlated, and the one-factor solution does not entail negative unique variances, the data do satisfy the hypothesis of unidimensionality. However, two or three latent factors are equally compatible with the data, and there is no way of telling which is true. Because the hypothesis of latent unidimensionality may or may not be true when $k = 3$, and certainly will not be true when $k > 3$, it seems safer not to rely on the hypothesis at all. This is why lower bounds to reliability should *in theory* be preferred to (10) and (11). The lower bounds do rely on the assumption of

uncorrelated errors of the test parts, but do not involve any additional assumptions concerning the latent dimensionality. Still, there may be objections to using bounds like the glb *in practice*, because of its reputation for sampling bias. As we shall see in the sequel, these objections are not compelling.

A second implication of Shapiro's result is that we get farther away from unidimensionality as k increases. This has not been fully appreciated. It is well known that unidimensionality and reliability are different things. Time and again, the positive impact of test length on reliability has been invoked to demonstrate this; see, for instance, Cortina (1993) and Schmitt (1996). Apparently, unidimensionality is tacitly conceived of as being insensitive to test length. Appealing as this may be *latently*, Shapiro's result implies that, *patently*, dimensionality does go up with the number of test parts, whence we get farther away from unidimensionality as test length is stepped up. In this sense, unidimensionality and reliability should be conceived of as antagonistic concepts.

Assessing Closeness to Unidimensionality

Although the one-factor hypothesis is false when $k > 3$, and farther out as the number of test parts increases, it still may be close. Accordingly, when perfect unidimensionality is impossible, it may be desirable knowing how close is a test to being unidimensional. In principle, one may fit the one-factor model and use any of a number of fit measures to express the degree of unidimensionality. Here, an alternative is proposed by adhering to the Jackson–Agunwamba framework, which led to the greatest lower bound as the least reliable point in the set of all possible solutions with uncorrelated errors.

Consider, again, the set of possible matrices $\mathbf{C}_E$ of Figure 1. If we reinterpret $\mathbf{C}_E$ as the nonnegative diagonal matrix of unique variances, and $\mathbf{C}_X - \mathbf{C}_E$ as the "reduced" covariance matrix (also psd) with communalities instead of observed variances on the diagonal, the set of all possible matrices $\mathbf{C}_E$ represents the set of all possible factor solutions, where the unique factors are assumed to correlate zero with every variable of which they are not part. Although this set (when $k > 3$) does not contain a $\mathbf{C}_E$ such that $\mathbf{C}_X - \mathbf{C}_E$ has rank one, some points are much closer to this ideal than others, and we may look for the point closest to unidimensionality. This requires a criterion of closeness to unidimensionality.

It is well known that the best least squares rank-one approximation to a matrix is determined by the first component of its singular value decomposition, the squared distance being the sum of squares of all singular values *except the first*. Hence, the matrix $\mathbf{C}_E$ such that $\mathbf{C}_X - \mathbf{C}_E$ has the smallest *sum of squares* of the last $k - 1$ eigenvalues would qualify for "the solution closest to unidimensionality." However, any other monotonic function of the last $k - 1$ eigenvalues of $\mathbf{C}_X - \mathbf{C}_E$ also qualifies. We shall take the sum of the last $k - 1$ *unsquared* eigenvalues as criterion, for two reasons. The first one is very practical. The only method currently available for minimizing a relevant function of the last $k - 1$ eigenvalues involves the plain sum of these eigenvalues: Although other functions (sum of squares, or maximum likelihood) would also be relevant, methods to optimize such functions over the possible set are not yet available.[1] The second reason is the existence of a natural measure of unidimensionality associated with minimizing the sum of (unsquared) last $k - 1$ eigenvalues. This measure will be discussed shortly.

Having adopted the sum of the last $k - 1$ eigenvalues of $\mathbf{C}_X - \mathbf{C}_E$ as criterion of nonunidimensionality, it is obvious that the "least unidimensional" point corresponds to the origin ($\mathbf{C}_E = \mathbf{O}$), because subtracting any positive number from any of the diagonal elements of $\mathbf{C}_X$ will reduce all eigenvalues, including the last $k - 1$. Subtracting zero therefore maximizes their sum. Thus, PCA, considered as a possible solution for factor analysis, appears as the "least-

[1] By the time this paper went to press, this statement has become obsolete as far as the sum of squares is concerned.

unidimensional" solution. To find the "most-unidimensional" counterpart, a special iterative procedure is needed. Ten Berge and Kiers (1991) (also see Ten Berge (1998) and Shapiro & Ten Berge (2002)), have proposed Minimum Rank Factor Analysis (MRFA, downloadable as mrfa2.exe from `ppswmm.ppsw.rug.nl/~kiers/main.html`) as the method that picks, from the possible set, that matrix $\mathbf{C}_E$ for which the sum of the $k-r$ smallest eigenvalues of $\mathbf{C}_X - \mathbf{C}_E$ is a minimum. The number of common factors r has to be specified in advance. Clearly, applying this method with $r = 1$ yields the solution from the possible set which is most unidimensional, because it has the smallest possible sum of the last $k - 1$ eigenvalues of $\mathbf{C}_X - \mathbf{C}_E$. It does not correspond to the glb, although it may be close. It always is another boundary point of the marked area in Figure 1.

After determining this $\mathbf{C}_E$, all eigenvalues of $\mathbf{C}_X - \mathbf{C}_E$ can be evaluated. The ratio of the first eigenvalue to the sum of all eigenvalues is a natural coefficient of "closeness to unidimensionality." It will be one if and only if patent unidimensionality holds, and it can be converted to a percentage of explained *common* variance (ECV), parallel to the percentage of explained *observed* variance in PCA. An example will be given below. The unidimensionality measure proposed here is similar to one proposed by Bentler (1972, p. 355). Whereas Bentler used the ratio of the first to the sum of all eigenvalues implied by his lower bound theta (which often coincides with the glb, the least reliable amid the possible solutions), we use instead the eigenvalues that belong with the most unidimensional solution.

It should be clear that reliability and patent unidimensionality, as measured by ECV, are distinct concepts. The test parts may have large error variances, implying low reliability, and yet ECV, which entirely depends on the eigenvalues of $\mathbf{C}_X - \mathbf{C}_E$, may be high. Conversely, when the test parts have little error variance, they may still be far away from unidimensionality. It is revealing that ECV can be increased, at the cost of losing reliability, by simply reducing the number of test parts, a point already noted above. Similarly, in terms of the possible set of solutions for $\mathbf{C}_E$, the choice $\mathbf{C}_E = \mathbf{O}$ reflects the best possible case for reliability, and the worst case scenario (the least unidimensional solution) for unidimensionality. This is another instance where reliability and unidimensionality behave as opposite concepts, rather than as just distinct.

Sampling Bias of the glb

Above, a case has been made for the glb as preferable to reliability coefficients based on assumptions of unidimensionality, because it relieves us of the need to uphold an untenable hypothesis of unidimensionality. However, the glb does have a problem of its own. It has been known from the start (Woodhouse & Jackson, 1977; Cronbach, 1988) that the glb may have a fierce positive sampling bias. Its sample estimates tend to overestimate the population value. This problem already shows up when the number of variables is 10, say. Even in samples of size 1000, cases exist where the population glb is overestimated by more than .10. But the sampling bias problem cannot be used to dismiss the glb altogether.

Firstly, Shapiro and Ten Berge (2000) have shown that the asymptotic (large sample) bias of the glb is *negative* when the implied $\mathbf{C}_T$ has rank $k - 1$. This happens quite often when $k = 3$, 4, 5, or 6. Simulations indicate that sample bias is in fact negligible in such cases. It may be noted that (7b), when the glb equals λ_4, is a case in point where the implied rank of $\mathbf{C}_X - \mathbf{C}_E$ is indeed $k - 1$. However, when the number of test parts is large, (positive) sampling bias does require special measures. Such measures are available. Verhelst (1998) has offered a method for bias correction based on resampling. The program to carry out the correction, called TiaPlus, is available from `TON.HEUVELMANS@CITOGROEP.NL`.

Secondly, it seems that the factor analysis based reliability, see (11), has never been under suspicion of sampling bias. However, the present authors have found that (11) often behaves much like the glb. A prototypical example and an explanation will be reported below. Accord-

ingly, whenever the glb is under suspicion for possible sampling bias, it seems wise to also scrutinize its competitor (11) on the same account.

An Example: Six Political Items

Although the glb is known to have a sampling bias problem, its competitors may have the same problem. This will now be demonstrated in an example, the six political survey items described by De Leeuw (1983). These items are meant to measure the same trait, and therefore should be close to unidimensional. The correlation matrix is

$$\mathbf{C}_X = \begin{bmatrix} 1.000 & & & & & \\ .446 & 1.000 & & & & \\ .462 & .380 & 1.000 & & & \\ .398 & .241 & .589 & 1.000 & & \\ .583 & .536 & .569 & .459 & 1.000 & \\ .516 & .483 & .417 & .403 & .514 & 1.000 \end{bmatrix}. \tag{12}$$

Following De Leeuw (1983), we treat the correlation matrix (based on $n = 119$ members of parliament) as if it were based on the population, and construct 500 samples, of sizes 100, 250, 500, and 1000, respectively, under the assumption of multivariate normality. For each sample, we evaluate alpha, λ_4, glb, its asymptotic bias approximation (Shapiro & Ten Berge, 2000), and the factor analysis based reliability coefficient (11), based on the rank-one MRFA solution. Also, we include the "unidimensionality" measure ECV (percentage of Explained Common Variance with one factor) associated with the latter solution. Table 1 gives average results (always over 500 replications) for each sample size, and for the population. For each sample size, we report means and standard deviations. Empirical bias (emp. bias) was computed for each statistic by subtracting the population parameters from the sample means.

In general, the sampling bias of glb is quite small for the correlation matrix (12), yet it is slightly larger (in absolute size) than for alpha. On the other hand, glb had a smaller sampling variance than alpha, the difference being larger in smaller samples. As expected, glb was notice-

TABLE 1.
Six variables of De Leeuw (1983).

	alpha	λ_4	Glb	As. bias	(11)	ECV
Population	0.840	0.883	0.885		0.885	80.10
M 100	0.836	0.889	0.893		0.892	77.48
SD 100	0.026	0.020	0.019		0.019	4.44
emp. bias	−0.004	0.006	0.008	0.004	0.007	−2.62
M 250	0.839	0.885	0.888		0.888	79.40
SD 250	0.017	0.014	0.013		0.013	3.41
emp. bias	−0.001	0.002	0.003	0.001	0.003	−0.70
M 500	0.840	0.884	0.887		0.886	79.76
SD 500	0.011	0.010	0.009		0.009	2.37
emp. bias	0.000	0.001	0.002	0.001	0.001	−0.34
M 1000	0.839	0.883	0.886		0.885	79.92
SD 1000	0.008	0.007	0.007		0.007	1.77
emp. bias	0.000	0.000	0.001	0.000	0.000	−0.18

Note. M = average of the sample coefficients at the respective sample size; SD = standard deviation of the sample coefficients at the respective sample size; ECV = percentage of common variance explained by the first factor.

ably higher than alpha both in population and in samples. But λ_4 and (11) in particular behave very similarly to the glb in every respect. The similarity of λ_4 and glb is not surprising. Jackson and Agunwamba (1977) and Woodhouse and Jackson (1977) have already pointed to conditions for these two to coincide. Even if the conditions are not met, the similar behavior of the two might have been expected. However, the near identity of glb and (11) is remarkable. It shows that (11) shares the bias problems with the glb. An explanation will now be given.

For the correlation matrix (12), the true score variances implied by the glb happen to be equal (within four decimal places) to the communalities of the single factor MRFA solution. This means that $\mathbf{C}_T$ will be nearly the same in both cases. Upon factoring $\mathbf{C}_T$ as $\mathbf{C}_T = \mathbf{a}\mathbf{a}' + \mathbf{B}\mathbf{B}'$, where $\mathbf{a}$ is the vector of loadings on the first factor, and $\mathbf{B}$ the matrix of loadings on the remaining common factors, the glb can be evaluated as $\mathbf{1}'\mathbf{C}_T\mathbf{1}/\mathbf{1}'\mathbf{C}_X\mathbf{1} = \mathbf{1}'(\mathbf{a}\mathbf{a}' + \mathbf{B}\mathbf{B}')\mathbf{1}/\mathbf{1}'\mathbf{C}_X\mathbf{1}$, whereas (11) yields $\mathbf{1}'\mathbf{a}\mathbf{a}'\mathbf{1}/\mathbf{1}'\mathbf{C}_X\mathbf{1}$. Because all elements of $\mathbf{C}_T$ are positive, all loadings in $\mathbf{a}$ have the same sign, whence the column sums of $\mathbf{B}$ are close to zero, hence $\mathbf{1}'\mathbf{B}\mathbf{B}'\mathbf{1}$ is near zero. This means that, although the associated factors do explain some variance, they contribute nothing to the reliability, because positive and negative contributions to true scores cancel. This explains why the glb must be close to (11) in data sets like (12). However, whereas (11) is based on an assumption of unidimensionality, glb is not. Therefore, the glb is to preferred.

It can further be noted from Table 1 that, in the present case, the asymptotic (large sample) estimate of bias seems to underestimate the bias observed in small samples, a tendency also found by Shapiro and Ten Berge (2000). Furthermore, it is instructive to consider the question of how far away are the six variables of (12) from a one-factor solution. The one-factor MRFA solution yields 80.1% of explained common variance (last column of Table 1), a quantity underestimated in small samples. On the other hand, the largest eigenvalue of (12) is 3.346, implying that the first component in PCA yields 55.8% of explained variance, representing the worst possible case for unidimensionality. It can be concluded that, within the possible set, "closeness to unidimensionality" varies from 55.8 to 80.1%. The best possible case for unidimensionality is 80.1%, which does not seem bad for $k = 6$. This percentage would certainly have been deemed bad had there been only three test parts.

Discussion

Although the glb is theoretically superior to all other lower bounds to reliability, it has gained little popularity, mainly because of the sampling bias problem. Until bias correction methods have been proven adequate, the glb will continue to be ignored. Also, whenever alpha is high, there is no reason to resort to superior lower bounds, except when a correction for attenuation is desired. Therefore, it would be premature to demand that the glb be considered whenever alpha is, except when samples sizes run in the thousands. Nevertheless, we have argued that the glb should not be ignored in the context of reliability of congeneric tests. The most obvious way to see this is that, when $k = 3$, the glb is typically identical to either Kristof's coefficient or Guttman's λ_4, two coefficients that have been proposed for the reliability of congeneric tests (e.g., Osburn, 2000).

It seems that bias problems, notoriously invalidating the glb when the number of test parts is large, do not pose much of a problem when the number of test parts is small. In addition, competing coefficients are liable to the same bias problems. This goes in particular for factor analysis based reliability coefficients. They behave much like the glb, have the same bias problem as the glb, but, in addition, rely on an unwarranted single factor (unidimensionality) hypothesis. This is why the glb is to be preferred.

Reliability has often been mistaken for unidimensionality. By considering the issue of unidimensionality in the framework proposed by Jackson and Agunwamba for the context of reliability, unidimensionality has appeared inversely related to reliability. The most favorable point

for reliability corresponds to the least favorable point for unidimensionality. In the same vein, increasing the number of test parts will boost reliability, but it will reduce "closeness to unidimensionality." The antagonism of reliability and unidimensionality is a surprising consequence of abiding by the Jackson and Agunwamba framework. When that antagonism is deemed unacceptable, one will have to part with the classical conception of reliability, and resort to a different framework; see, for instance, Nicewander (1990).

Appendix A: Proof of the Equality

$$\alpha + \frac{1}{(k-1)\text{Var}(X)} \sum_{i<j} \text{Var}(T_i - T_j) = \text{reliability}. \tag{A1}$$

It is readily verified that

$$\sum_{i<j} \text{Var}(T_i - T_j) = (k-1) \sum_{i=1}^{k} \text{Var}(T_i) - \sum_{i \neq j} \sigma_{ij}$$

so

$$\frac{1}{k-1} \sum_{i<j} \text{Var}(T_i - T_j) = \sum_{i=1}^{k} \text{Var}(T_i) - \frac{\sum_{i \neq j} \sigma_{ij}}{k-1}. \tag{A2}$$

Hence, writing

$$\alpha = \frac{k}{k-1} \sum_{i \neq j} \frac{\sigma_{ij}}{\text{Var}(X)},$$

the left-hand side of (A1) becomes

$$\frac{k \sum_{i \neq j} \sigma_{ij}}{(k-1)\text{Var}(X)} + \frac{\sum_{i=1}^{k} \text{Var}(T_i)}{\text{Var}(X)} - \frac{\sum_{i \neq j} \sigma_{ij}}{(k-1)\text{Var}(X)} = \frac{\sum_{i \neq j} \sigma_{ij}}{\text{Var}(X)} + \frac{\sum_{i=1}^{k} \text{Var}(T_i)}{\text{Var}(X)} = \frac{\text{Var}(T)}{\text{Var}(X)}. \tag{A3}$$

References

Bekker, P.A., & De Leeuw, J. (1987). The rank of reduced dispersion matrices. *Psychometrika, 52*, 125–135.

Bentler, P.M. (1972). A lower-bound method for the dimension-free measurement of reliability. *Social Science Research, 1*, 343–357.

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.

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. (1988). Internal consistency of tests. *Psychometrika, 53*, 63–70.

De Leeuw, J. (1983). Models and methods for the analysis of correlation coefficients. *Journal of Econometrics, 22*, 113–137.

Feldt, L.S., Woodruff, D.J., & Salih, F.A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement, 11*, 93–103.

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

Guttman, L. (1958). To what extent can communalities reduce rank. *Psychometrika, 23*, 297–308.

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

Kristof, W. (1974). Estimation of reliability and true score variance from a split of the test into three arbitrary parts. *Psychometrika, 39*, 245–249.

Ledermann, W. (1937). On the rank of reduced correlation matrices in multiple factor analysis. *Psychometrika, 2*, 85–93.

McDonald, R.P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology, 23*, 1–21.

Nicewander, W.A. (1990). A latent-trait based reliability estimate and upper bound. *Psychometrika, 55*, 65–74.

Novick, M.R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika, 32*, 1–13.

Osburn, H.G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods, 5*, 343–355.

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

Shapiro, A. (1982a). Rank reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis. *Psychometrika, 47*, 187–199.

Shapiro, A. (1982b). Weighted minimum trace factor analysis. *Psychometrika, 47*, 243–264.

Shapiro, A., & Ten Berge, J.M.F. (2000). The asymptotic bias of Minimum Trace Factor Analysis, with applications to the greatest lower bound to reliability. *Psychometrika, 65*, 413–425.

Shapiro, A., & Ten Berge, J.M.F. (2002). Statistical inference of minimum rank factor analysis. *Psychometrika, 67*, 79–94.

Spearman, C.E. (1927). *The abilities of man*. London: McMillan.

Ten Berge, J.M.F. (1998). Some recent developments in factor analysis and the search for proper communalities. In A. Rizzi, M. Vichi, and H.-H.Bock (Eds.): *Advances in data science and classification* (pp. 325–334). Berlin: Springer.

Ten Berge, J.M.F., & Kiers, H.A.L. (1991). A numerical approach to the exact and the approximate minimum rank of a covariance matrix. *Psychometrika, 56*, 309–315.

Ten Berge, J.M.F., Snijders, T.A.B., & Zegers, F.E. (1981). Computational aspects of the greatest lower bound to reliability and constrained minimum trace factor analysis. *Psychometrika, 46*, 357–366.

Van Zijl, J.M., Neudecker, H., & Nel, D.G. (2000). On the distribution of the maximum likelihood estimator of Cronbach's alpha. *Psychometrika, 65*, 271–280.

Verhelst, N.D. (1998). *Estimating the reliability of a test from a single test administration* (Measurement and Research Department Report No. 98-2). Arnhem: CITO.

Wilson, E.B., & Worcester, J. (1939). The resolution of six tests into three general factors. *Proceedings of the National Academy of Sciences, 25*, 73–79.

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, 42*, 579–591.

Yuan, K.-H., & Bentler, P.M. (2002). On robustness of the normal-theory based asymptotic distributions of three reliability coefficient estimates. *Psychometrika, 67*, 251–259.

Manuscript received 13 JUN 2002
Final version received 8 SEP 2003