

ALPHA, DIMENSION-FREE, AND MODEL-BASED INTERNAL CONSISTENCY RELIABILITY

PETER M. BENTLER

DEPARTMENTS OF PSYCHOLOGY AND STATISTICS, UCLA

As pointed out by Sijtsma ([in press](#)), coefficient alpha is inappropriate as a single summary of the internal consistency of a composite score. Better estimators of internal consistency are available. In addition to those mentioned by Sijtsma, an old dimension-free coefficient and structural equation model based coefficients are proposed to improve the routine reporting of psychometric internal consistency. The various ways to measure internal consistency are also shown to be appropriate to binary and polytomous items.

Key words: common, unique, true, error scores.

Sijtsma ([in press](#)) has ably critiqued the unfortunate ascendance of coefficient alpha as the almost universal and sole estimator of the somewhat vague population concept of internal consistency. Among several recommendations, he suggests to report the greatest lower bound, here denoted ρ_{glb} , as a measure of internal consistency reliability, and the explained common variance (ECV) based on minimum rank factor analysis, as a measure of unidimensionality. These are good recommendations, but they are incomplete. We provide additional technical detail on some methods, and also consider several alternative methods.

Using a slightly different notation than Sijtsma, we are interested in describing the reliability of a composite X that is a simple sum of p unit-weighted components such as $X = X_1 + X_2 + \dots + X_p$. An internal consistency reliability coefficient describes the quality of the composite or scale in terms of hypothesized constituents of the components X_i . These might represent true and error parts based on classical test theory ($X_i = T_i + E_i$), common and unique parts based on common factor analysis ($X_i = C_i + U_i$), or the loading of the component on its factor plus residual error ($X_i = \lambda_i F + E_i$). Since U_i is the sum of two uncorrelated parts called specificity and error, that is, $U_i = S_i + E_i$, where S_i is specificity, it is a more general decomposition than that of classical test theory and we focus on it. We assume that C_i and U_i are uncorrelated, and that the X_i are not linearly dependent. Then the covariance structure model $\Sigma = \Sigma_C + \Psi$ follows, where Σ_C is the positive semidefinite (*psd*) covariance matrix of the common variables and Ψ is the covariance matrix of the unique variables, typically taken as positive definite and diagonal. In general, the C_i are linearly dependent, so that Σ_C can be of low rank and can have such factor analytic (FA) decompositions as $\Sigma_C = \Lambda \Lambda'$ or $\Sigma_C = \Lambda \Phi \Lambda'$, or a structure based on a FA simultaneous equation system such as $\Sigma_C = \Lambda(I - B)^{-1} \Phi (I - B)^{-1'} \Lambda'$ (see Bentler, 2007). Here, Λ is a factor loading matrix, Φ is the covariance matrix of nondependent latent factors, and B is a matrix of coefficients relating latent factors.

Research supported in part by grants DA00017 and DA01070 from the National Institute on Drug Abuse. This paper is based in part on Bentler (2003).

Requests for reprints should be sent to Peter M. Bentler, Departments of Psychology and Statistics, UCLA, Franz Hall, Box 951563, Los Angeles, CA 90095-1563, USA. E-mail: bentler@ucla.edu

1. Reliability and Internal Consistency

As noted by Sijtsma, internal consistency reliability is a vague concept, and so we now give it a concrete definition. Given decomposition for the parts $X_i = T_i + E_i$ and further that $U_i = S_i + E_i$, we have that the total or composite score has a similar decomposition $X = C + U$, where $C = \sum_1^p C_i$, $U = \sum_1^p U_i = \sum_1^p (S_i + E_i) = S + E$, and S and E are specific and error total scores. Thus, $X = C + S + E$, i.e., the total score is made up of mutually uncorrelated total common, specific, and error variables. With $T = C + S$, the covariance matrices of the parts are $\Sigma = \Sigma_C + \Psi = \Sigma_C + \Psi_S + \Psi_E = \Sigma_T + \Psi_E$, where $\Sigma_T = \Sigma_C + \Psi_S$ and Ψ_E are the true and error covariance matrices and Ψ_S is the diagonal matrix of specific variances. The reliability of the composite ρ_{XX}^* is defined as the ratio of $\text{var}(T)/\text{var}(X)$, or

$$\rho_{XX}^* = \frac{\sigma_T^2}{\sigma_X^2} = \frac{1' \Sigma_T 1}{1' \Sigma 1} = 1 - \frac{1' \Psi_E 1}{1' \Sigma 1}, \quad (1)$$

where 1 is a column vector of units. We also immediately have a clear-cut definition of internal consistency reliability ρ_{XX} as

$$\rho_{XX} = \frac{\sigma_C^2}{\sigma_X^2} = \frac{1' \Sigma_C 1}{1' \Sigma 1} = 1 - \frac{1' \Psi 1}{1' \Sigma 1} = 1 - \frac{\sigma_U^2}{\sigma_X^2} \leq \rho_{XX}^*, \quad (2)$$

which is a lower bound to the reliability of the composite. See Bentler (1968, (12)) and Heise and Bohrnstedt (1970, (32)) for the origin of these ideas. These equations make clear that an internal consistency coefficient is a downward biased estimator of reliability by the proportion of specific variance, that is,

$$\rho_{XX} + \frac{\sigma_S^2}{\sigma_X^2} = \rho_{XX}^*. \quad (3)$$

Equality holds only when there is no specific variance.

2. Alpha

Let σ_{ij} be an off-diagonal element of Σ and $\bar{\sigma}_{ij}$ be the average of all σ_{ij} . Then we have coefficient alpha as (see Sijtsma, (2))

$$\alpha = \frac{p^2 \bar{\sigma}_{ij}}{\sigma_X^2}. \quad (4)$$

We now assume that Ψ is diagonal, so that $\sigma_C^2 = p\bar{\sigma}_{Ci}^2 + p^2\bar{\sigma}_{ij} - p\bar{\sigma}_{ij} = p^2\bar{\sigma}_{ij} + p(\bar{\sigma}_{Ci}^2 - \bar{\sigma}_{CiCj})$, where $\bar{\sigma}_{Ci}^2$ and $\bar{\sigma}_{CiCj}$ are the average diagonal and off-diagonal elements of Σ_C . Since Σ_C is positive semidefinite, $(\bar{\sigma}_{Ci}^2 - \bar{\sigma}_{CiCj}) \geq 0$, and we have that

$$\alpha \leq \rho_{XX} \quad (5)$$

with equality if all the elements of Σ_C are equal. Although Sijtsma points out that α is downwardly biased, we now see by comparing (2), (3), and (4) that the extent of this bias is given by

$$\alpha + \frac{p(\bar{\sigma}_{Ci}^2 - \bar{\sigma}_{CiCj})}{\sigma_X^2} + \frac{\sigma_S^2}{\sigma_X^2} = \rho_{XX}^*. \quad (6)$$

A point not discussed by Sijtsma is that α can, in fact, overestimate ρ_{XX} . This can occur if Ψ is not diagonal, i.e., if there are correlated errors. Then (2) is still appropriate, but the numerator in (4) is inflated by these correlated errors. For some discussions of this problem, see, e.g., Raykov (2001) and Kano and Azuma (2003). Of course, if correlated errors are parameterized as factors so that Ψ is diagonal, then (5) again holds. It would seem that the question of whether to consider correlated errors as factors, and hence part of the common factor space, or as residual covariances, and hence as part of the unique space, should be left up to the goals of the investigator.

From (4), we also see that α reflects the number of parts in the composite and also a weighted average correlation among measures, which may be useful information. If the variances of the parts are all equal, it reflects the average correlation. More precisely, let the generalized classical test theory average item or part reliability (Bentler, 1964) be defined by

$$\rho_{X_i X_i} = \frac{\bar{\sigma}_{ij}}{\bar{\sigma}_i^2}, \tag{7}$$

where $\bar{\sigma}_i^2$ is the average item or part variance. If the item variances are equal, as in a parallel test situation, (7) is just the average correlation among items. In general, α is just the Spearman–Brown step-up of the part reliability (7), that is,

$$\alpha = \frac{p\rho_{X_i X_i}}{1 + (p - 1)\rho_{X_i X_i}}. \tag{8}$$

It can be argued that such an interpretation provides sufficient rationale for the continuing use of α as one of several measures of internal consistency.

3. Dimension-Free Internal Consistency Reliability

Sijtsma described the coefficient ρ_{glb} , but did not review an earlier coefficient that is precisely equal to the glb under standard assumptions. Suppose now we take $\Sigma_C = \Lambda\Lambda'$ so that $\Sigma = \Lambda\Lambda' + \Psi$ and Ψ is diagonal, i.e., we have a traditional factor analysis model. Bentler (1972) developed a dimension-free reliability coefficient based on this decomposition without any assumption on the number of factors. He proposed that factor loading matrix Λ (of arbitrary dimension k) be chosen so that $\text{trace}(\Lambda\Lambda')$ is minimized while the model $\Sigma = \Lambda\Lambda' + \Psi$ holds precisely.¹ But $\min \text{trace}(\Lambda\Lambda') = \max \text{trace}(\Psi) = \max 1'\Psi 1$, so that Bentler’s dimension-free lower bound to reliability is

$$\rho_{\text{blb}} = 1 - \max \frac{1'\Psi 1}{1'\Sigma 1} = \min \left(1 - \frac{1'\Psi 1}{1'\Sigma 1} \right), \quad \text{with } (\Sigma - \Psi) \text{ psd.} \tag{9}$$

There is no smaller internal consistency reliability coefficient for which $(\Sigma - \Psi)$ is positive semidefinite, i.e., for which the factors are real, and not imaginary. If, at the solution, the unique variance matrix Ψ has nonnegative variances $\psi_{ii} \geq 0$, the dimension free lower bound is the greatest lower bound to reliability. Actually, we would argue that reliability theory requires every variable to have some nonzero error variance.

¹A more recent name for the optimization is minimum trace factor analysis or MTFA (e.g., della Riccia & Shapiro, 1982; Shapiro, 1982).

We had shown in (6) how α relates to reliability ρ_{XX}^* . In (6), however, no specific structure was assumed for Σ_C . Using the structure (9), for $\Sigma_C = \Lambda\Lambda'$ with minimum trace, we can say

$$\rho_{\text{blb}} = \alpha + \frac{p(\bar{\sigma}_{Ci}^2 - \bar{\sigma}_{CiCj})}{\sigma_X^2}. \quad (10)$$

That is, ρ_{blb} increases α by the smallest possible amount that would make Σ_C *psd* with minimum trace.

If we also constrain $\psi_{ii} \geq 0$ when minimizing (9), that is, we disallow Heywood cases, we obtain the greatest lower bound to internal consistency reliability (Woodhouse & Jackson, 1977; Bentler & Woodward, 1980). Specifically,

$$\rho_{\text{glb}} = 1 - \max \frac{1'\Psi 1}{1'\Sigma 1} = \min \left(1 - \frac{1'\Psi 1}{1'\Sigma 1} \right), \quad \text{with } (\Sigma - \Psi) \text{ and } \Psi \text{ psd.} \quad (11)$$

Since (11) is an optimization problem with an additional constraint as compared to (9), it immediately follows that

$$\rho_{\text{blb}} \leq \rho_{\text{glb}}, \quad (12)$$

with equality when there are no negative unique variances in ρ_{blb} .² Another way to say this is that ρ_{glb} improves on ρ_{blb} to the extent that variables in the population have negative error variances. This may be an illusory gain, since it implies that a factor model with nonnegative unique variances does not fit the population, and hence a reliability coefficient based on it may be questionable.

We now make several critical observations about these dimension-free coefficients. First, it was stated in association with (9) and (11) that the number of factors, k , is arbitrary. In covariance structure analysis, k is usually taken to be a small number so that a model would have positive degrees of freedom. Here, however, k is usually sufficiently large that a factor model would have negative degrees of freedom, that is, k exceeds the Ledermann bound. This might be an argument to consider model-based reliability coefficients, discussed in the next section. Closely related to this is the fact that ρ_{blb} and ρ_{glb} are based on the assumption that all covariation in the data should be modeled in the common factor space. Stated differently, Ψ is diagonal and no correlated errors are allowed. If the partitioning of variance that one has in mind implies that correlated errors should be considered part of error and not systematic common variance, these lower-bound coefficients overestimate internal consistency. Further, the computational procedures used to obtain solutions to (9) and (11) assume that there is no sampling variability in the sample covariances, that is, they treat sample covariances as if they are population covariances. Stated differently, while (9) and (11) are based on the population covariance matrix, the procedures are applied to a sample covariance matrix. This seems risky as any sample-based distortion of correlations will be modeled as real common variance. To take an extreme case, suppose the population covariance matrix $\Sigma = I$, that is, variables are uncorrelated. ρ_{blb} and ρ_{glb} would be zero. But a sample covariance matrix would have nonzero covariances, so that ρ_{blb} and ρ_{glb} would be positive. This illustrates the bias noted by Sijtsma. Shapiro and ten Berge (2000) offered an explicit expression for the asymptotic bias of MTFAs, and hence for ρ_{blb} . More effective non-asymptotic bias corrections also are available (Li & Bentler, 2004).

The very nature of k -dimensional coefficients also can be questioned. Perhaps one would like to assure that any measure of reliability is based on a single latent dimension, yet ρ_{blb} and ρ_{glb}

²This optimization problem is also called constrained minimum trace factor analysis (e.g., ten Berge, Snijders, & Zegers, 1981).

allow k dimensions. Actually, this is an illusory problem. Bentler (2007) shows that there exists a rotation in k -dimensional factor space that gives a unidimensional subspace of maximum internal consistency, and that the resulting coefficients are actually given by ρ_{blb} and ρ_{glb} .

Sijtsma recommends assessing unidimensionality by the ratio of the first eigenvalue of Σ_C to their sum, expressed as a percent. Called explained common variance, or ECV, this is a measure of how close Σ_C is to being unidimensional. This is an informative measure. Sijtsma follows ten Berge and Sočan (2004) in recommending that Σ_C and ECV be computed by a procedure called minimum rank factor analysis. But there is no reason to limit the definition in this way. Since $\text{tr}(\Sigma_C)$ equals the sum of the common factor eigenvalues, and these are minimized by the dimension-free methods described above, as suggested by Bentler (1972) it would be at least as meaningful to apply the ECV measure to the Σ_C obtained from dimension-free methods than from the minimum rank method.

4. Model-Based Internal Consistency Coefficients

The internal consistency coefficients discussed by Sijtsma, and also those reviewed above, are specified at a population level and are mathematical lower bounds to population internal consistency reliability. In practice, sample covariance and correlation matrices must be used in the computation instead of their population counterparts, which are essentially never available. It is well known that a model-based covariance matrix $\hat{\Sigma}$ can be a more efficient estimator of the population covariance matrix Σ than the product moment sample covariance matrix S_n based on a sample of size n , and hence it may be useful to consider reliability coefficients based on a model.

A particular model-based coefficient that has been proposed in the literature is based on the 1-factor model $\Sigma_C = \lambda\lambda'$, where λ ($p \times 1$) is the factor loading vector. Then $\Sigma = \lambda\lambda' + \Psi$, and we can follow Jöreskog (1971, p. 112) to define

$$\begin{aligned} \rho_{11} &= \frac{1'\lambda\lambda'1}{1'\Sigma 1} = \frac{(1'\lambda)^2}{1'\Sigma 1} = \frac{(\sum_1^p \lambda_i)^2}{1'\Sigma 1} \\ &= 1 - \frac{1'\Psi 1}{1'\Sigma 1}. \end{aligned} \tag{13}$$

Zinbarg, Revelle, Yovel, and Li (2005) equate ρ_{11} with McDonald's (1985) ω_H . Typically, Ψ is taken to be a diagonal matrix but it need not be so (see Bollen, 1980). To make ρ_{11} operational, the one-factor model is estimated based on S_n , and $\hat{\rho}_{11}$ is computed from (13) using the parameter estimates $\hat{\lambda}$ and $\hat{\Psi}$ and the resulting $\hat{\Sigma}$ in place of their population counterparts.

However, a 1-factor model hardly ever describes real data with a reasonable large p , that is, the null hypothesis $\Sigma = \lambda\lambda' + \Psi$ will be invariably rejected (see e.g., ten Berge & Sočan, 2004, p. 613). McDonald (1999, p. 89) stated that if the 1-factor fit is poor "...we should not be using the coefficient anyway," a point emphasized by Bentler (2003, 2007) in his proposal to use any statistically acceptable structural model that contains additive random errors to estimate internal consistency reliability. A good example is a variant of the FA simultaneous equation model, which can be written as

$$\Sigma = \Lambda(I - B)^{-1}\Phi(I - B)^{-1'}\Lambda' + \Psi. \tag{14}$$

This fits in our framework with

$$\Sigma_C = \Lambda(I - B)^{-1}\Phi(I - B)^{-1'}\Lambda'.$$

Hence, after obtaining parameter estimates \hat{A} , \hat{B} , $\hat{\Phi}$, $\hat{\Psi}$ for any model of the form (14) whose null hypothesis is accepted, we also have $\hat{\Sigma}_C$ and $\hat{\Sigma}$, and we can use our basic internal consistency definition (2) to compute

$$\hat{\rho}_{xx} = \frac{1' \hat{\Sigma}_C 1}{1' \hat{\Sigma} 1} = 1 - \frac{1' \hat{\Psi} 1}{1' \hat{\Sigma} 1}. \quad (15)$$

Bentler (2007) shows that (15) can equivalently be conceptualized as the coefficient for a unidimensional subspace of $\hat{\Sigma}_C$. Although we consider this coefficient to be one of the most defensible general-purpose internal consistency estimators, it was not considered by Sijtsma. Hopefully, the coefficient (15) is just an ancillary measure to the more important objective of actually understanding the structure of the instrument based on (14).

5. Coefficients for Binary and Ordinal Variables

Sijtsma mentions item response theory, but does not give concrete guidelines on how to obtain internal consistency reliability estimates under this framework. Here, we note that this is easy to do if one accepts the notion that any observed binary or ordered categorical variable X_i represents a discretization of an underlying continuous normal variable Z_i . The sum of the observed item scores, $X = X_1 + X_2 + \dots + X_p$ is no longer of interest, rather the sum of the underlying normal variables, that is, $Z = Z_1 + Z_2 + \dots + Z_p$. The internal consistency reliability of the composite Z can be obtained once an estimate of the correlation matrix of the Z_i , known as the polychoric and polyserial correlation matrix, is available. Such estimates can be obtained via standard SEM packages based, e.g., on Lee, Poon, and Bentler (1995) or Muthén (1984). Any of the previously described internal consistency coefficients can then be computed.

6. Minor Issues and Conclusion

The decompositions we have used to develop internal consistency coefficients also allow us to address some minor points in Sijtsma's overview. In his real data example on coping, Sijtsma uses principal components analysis (PCA) to get "factor" loadings. However, nothing in the basic definitions related to reliability summarized above has anything to do with PCA. PCA is not relevant to the model, so whatever PCA tells us is irrelevant unless PCA and FA loading matrices are identical, which would only occur as the number of items becomes arbitrarily large (e.g., Bentler & Kano, 1990). Second, Sijtsma analyzes a correlation matrix. However, our development shows that the properties of the total score X depend on the covariance matrix of the parts, not on their correlation matrix. Reliability coefficients defined for continuous X are not identical to those that might be defined for $Z = Z_1 + Z_2 + \dots + Z_p$ whose components are in a z-score or correlation metric. Third, in association with $\rho_{g|b}$ Sijtsma gives an example of a 3×3 covariance matrix (4) and draws conclusions related to it. However, this covariance matrix is not full rank, and hence in our view it violates a basic regularity assumption of classical reliability theory, namely, that no variable is measured perfectly without any random error.

Almost all of the coefficients described in this paper have been available in the EQS program (Bentler, *in press*) for many years. The ECV measure is not currently computed.

We conclude by making an observation on a typically-overlooked point. Sample covariance matrix based estimates such as $\hat{\alpha}$ or $\hat{\rho}_{g|b}$ and even the model-based coefficients reviewed above are not necessarily lower-bounds to population reliability. Thirty years ago, Woodward and Bentler (1978) developed a probabilistic or statistical lower bound to population reliability. It would seem that the concept of a statistical lower bound remains to be developed for several of the coefficients discussed above.

References

- Bentler, P. M. (1964). Generalized classical test theory error variance. *American Psychologist*, *19*, 548.
- Bentler, P. M. (1968). Alpha-maximized factor analysis (Alphamax): Its relation to alpha and canonical factor analysis. *Psychometrika*, *33*, 335–345.
- Bentler, P. M. (1972). A lower-bound method for the dimension-free measurement of internal consistency. *Social Science Research*, *1*, 343–357.
- Bentler, P. M. (2003, July). *Should coefficient alpha be replaced by model-based reliability coefficients?* Invited paper presented at International Meetings of the Psychometric Society, Cagliari, IT.
- Bentler, P. M. (2007). Covariance structure models for maximal reliability of unit-weighted composites. In S.-Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 1–19). Amsterdam: North-Holland.
- Bentler, P. M. (in press). *EQS 6 structural equations program manual*. Encino: Multivariate Software. www.mvsoft.com.
- Bentler, P. M., & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research*, *25*, 67–74.
- 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.
- Bollen, K. A. (1980). Issues in the comparative measurement of political democracy. *American Sociological Review*, *45*, 370–390.
- della Riccia, G., & Shapiro, A. (1982). Minimum rank and minimum trace of covariance matrices. *Psychometrika*, *47*, 443–448.
- Heise, D. R., & Bohrnstedt, G. W. (1970). Validity, invalidity, and reliability. In E. F. Borgatta (Ed.), *Sociological methodology 1970* (pp. 104–129). San Francisco: Jossey-Bass.
- Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, *36*, 109–133.
- Kano, Y., & Azuma, Y. (2003). Use of SEM programs to precisely measure scale reliability. In H. Yanai, A. Okada, K. Shigemasa, & Y. Kano (Eds.), *New developments in psychometrics* (pp. 141–148). Tokyo: Springer.
- Lee, S.-Y., Poon, W.-Y., & Bentler, P. M. (1995). A two-stage estimation of structural equation models with continuous and polytomous variables. *British Journal of Mathematical and Statistical Psychology*, *48*, 339–358.
- Li, L., & Bentler, P. M. (2004). *The greatest lower bound to reliability: Corrected and resampling estimators*. Paper presented at Symposium on Recent Developments in Latent Variables Modeling. Tokyo: Japanese Statistical Association and Japanese Behaviormetric Society.
- McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale: Erlbaum.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah: Erlbaum.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115–132.
- Raykov, T. (2001). Bias of coefficient α for fixed congeneric measures with correlated errors. *Applied Psychological Measurement*, *25*, 69–76.
- Shapiro, A. (1982). Rank reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis. *Psychometrika*, *47*, 187–199.
- 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.
- Sijtsma, K. (in press). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*. doi:10.1007/s11336-008-9101-0.
- 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.
- 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*, 201–213.
- 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.
- Woodward, J. A., & Bentler, P. M. (1978). A statistical lower-bound to population reliability. *Psychological Bulletin*, *85*, 1323–1326.
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α , Revelle's β , and McDonald's ω_H : Their relations with each other and two alternate conceptualizations of reliability. *Psychometrika*, *70*, 1–11.