Using unidim rather than omega in estimating undimensionality

William Revelle[1] and David Condon[2]
Northwestern University University of Oregon

A simple index of the unidimensionality of a scale, u, is introduced. u is just the product of two other indices: τ (a measure of τ equivalence) and ρ_c (a measure of congeneric fit). Simulations of u across scales ranging from 3 to 24 items with various levels of factor homogeneity, and demonstrations of its performance on 45 different personality and ability measures are shown. Comparisons with traditional measures (e.g., $\omega_h, \alpha, \omega_t$, CFI, ECV) show greater sensitivity to unidimensional structure and less sensitivity to the number of items in a scale. u is easily calculated with open source statistical packages and is relatively robust to sample sizes ranging from 100 to 5,000.

Public Significance

How to evaluate whether a psychologocial scale measures just one construct is a recurring problem in assessment. We present an intuitively easy to understand and easily calculated new index of undimensionality, u. We compare this index to conventional measures with simulated and real data sets.

Evaluating the dimensionality of a measure has been an ongoing challenging for many years. Ever since (Walker, 1931, p 75) assessed the degree of 'higgledypiggledyness,' researchers have attempted to assess the unidimensionality of scales. Hattie (1985) reviewed and then evaluated (Hattie, 1984) 87 ways to assess unidimensionality and stated that "Unidimensionality can be rigorously defined as the existence of one latent trait underlying the set of items" (Hattie, 1985, p 152). Although a test with more than 3 items can never be truly unidimensional, the question becomes how to assess how *close* a test is to unidimensionality (Ten Berge & Socan, 2004). Unidimensionality is a necessary condition for the application of Item Response Theory (IRT) models, even though these models can also be fit to multidimensional models as well (e.g., MIRT, Chalmers, 2012). If a multi-item scale is unidimensional, scores on that scale will reflect one underlying construct. If not, then scores reflect the underlying construct as well as other, extraneous sources of variance.

In this article, we suggest yet one more index, the u measure of unidimensionality which we believe compares favorably to some of the more popular alternatives. u is both easy to understand and easy to calculate and does not share the weaknesses of several popular alternatives (e.g., α, ω_t, CFI, or Explained Common Variance).

To understand the challenge of evaluating unidimensionality, consider a multi-item scale designed to measure a single construct (T) with inevitable random error (E). That is, each item x_i contributes construct-specific variance but is "befuddled by error" (McNemar, 1946, p 246). In terms from classical test theory (Spearman, 1904; Lord & Novick, 1968; McDonald, 1999), this may be represented as

$$x_i = \lambda_i \tau + \epsilon_i \tag{1}$$

with variances

$$\sigma_i^2 = \lambda_i^2 \sigma_\tau^2 + \sigma_\epsilon^2 \tag{2}$$

In the unlikely case that the λ_i are equal for all items and the variances of the ϵ_i are all equal, the items are said to be *parallel*. That is, all items contribute equally with respect to the construct measured by the scale. Similarly, subsets of the items from the scale would represent *parallel forms*. If the λ_i are equal for all items but the variances of the ϵ_i are unequal, the items are said to be τ *equivalent*. That is, each item has the same *True* score but different error variances. If the λ_i are unequal (as is the typical case for most measures used in psychology), the items are said to be *congeneric*. For parallel items, all the correlations and covariances among items will be identical. For τ equivalent items, the covariances will be identical but the correlations will not. For congeneric items, neither the covariances nor the correlations need to be identical. In all three cases, the values of λ_i fit a one factor latent variable model (Lucke, 2005; McDonald, 1999).

We thank Richard Zinbarg for helpful comments on an earlier draft.
contact: William Revelle revelle@northwestern.edu
Draft version of March 18, 2024

The relationships among these values are often used to evaluate the internal consistency of scales as an estimate of reliability as with, for example, Cronbach's α (aka Guttman's λ_3) (Cronbach, 1951; Guttman, 1945). For a scale (X) that is scored as the sum of its items ($X = \Sigma_{x_i}$), internal consistency reliability is estimated as the proportion of construct-specific variance to total observed variance. That is,

$$\rho_{xx} = \frac{\sigma^2_\tau}{\sigma^2_X}$$

Although most frequently used as an index of reliability, α is sometimes used — confusingly, in our opinion — as a means of assessing unidimensionality as well. α is a particularly poor measure because it assumes (without testing) the case of τ equivalence (Cronbach, 1951). In other words, all of the items are presumed to have the same (true score) relationship to the measured construct. This means α is just a function of the number of items (k) and the average covariance of the items:

$$\alpha = \frac{\sigma^2_X - \Sigma\sigma^2_i + k\bar{\sigma}_{ij}}{\sigma^2_X} = \frac{k}{k-1}\frac{k\bar{\sigma}_{ij}}{(\bar{\sigma^2_i} + (k-1)\bar{\sigma}_{ij})}. \quad (3)$$

Before the introduction of modern computers, the advantage of α was that it could be calculated from just the test variances (σ^2_X) and the sum of the item variances ($\Sigma\sigma^2_i$), and it did not require examining the internal structure of the test. This simplicity of calculation and subsequent introduction into popular proprietary software packages probably accounts for the widespread use of α (Cho & Kim, 2015),

An alternative to α is omega-total (ω_t). Based upon a factorial model of the item covariances (McDonald, 1999), ω_t is found by replacing the item variances (σ^2_i) with the amount of common variance (h^2_i) for each item:

$$\omega_t = \frac{\sigma^2_X - \Sigma\sigma^2_i + \Sigma h^2_i}{\sigma^2_X}. \quad (4)$$

Importantly, ω_t reflects the total amount of common variance among the items rather than the amount due to any single factor or dimension, and this limits its usefulness as a measure of unidimensionality. The difficulty is that ω_t does not indicate the extent to which items co-vary on underlying dimensions beyond the primary or "general" factor.

In fact, α and ω_t are similar in that both are the ratio of a reduced variance $\sigma^2_X - \delta$ to the total variance σ^2_X (where $\delta = \Sigma\sigma^2_i - k\bar{\sigma}_{ij}$ for α or $\delta = \Sigma\sigma^2_i - \Sigma h^2_i$ for ω_t). Thus, in both cases, the numerator increases linearly by the number of items, but the denominator by the square of the number items. Importantly, both coefficients will increase asymptotically to 1 as the number of items increases.

McDonald (1999) simultaneously introduced another coefficient which he also referred to as ω, but which Zinbarg, Revelle, Yovel, & Li (2005) refer to as hierarchical ω (ω_h). ω_h may be found by a hierarchical factoring of the original data followed by a subsequent Schmid-Leiman Transformation (Schmid & Leiman, 1957) or by a bifactor solution with a general factor and a number of group factors.

When considering the case of ω_h, one might expand equation 1 to account for the relationship of any item to (multiple) group factors in addition to the general and specific/error terms[3]. Thus, the observed score for a particular subject on an item is the sum of the products of factor scores ($\boldsymbol{g}, \boldsymbol{f}, \boldsymbol{s}, \boldsymbol{e}$) and loadings ($\boldsymbol{c}, \boldsymbol{A}, \boldsymbol{D}$) on these factors:

$$\boldsymbol{x} = \boldsymbol{cg} + \boldsymbol{Af} + \boldsymbol{Ds} + \boldsymbol{e} \quad (5)$$

ω_h is found by summing the loadings on the general factor and comparing the square of their sum to the test variance:

$$\omega_h = \frac{(\Sigma\lambda_i)^2}{\sigma^2_X} = \frac{\boldsymbol{1cc'1'}}{\sigma^2_X}. \quad (6)$$

As ω_h represents the proportion of general factor variance to total test variance, it directly addresses the difficulty introduced by using ω_t (the proportion of all common variance to total test variance) as an estimate of unidimensionality. In cases where one or more subset(s) of items share group variance that is not fully explained by variance on the general factor among all items, ω_h will be a smaller proportion than ω_t and a much better estimate of unidimensionality. In addition, because the relative contribution of the variance of a single item to the total variance decreases as the the number of items increases, the asymptotic value of ω_h is

$$\omega_{h_{limit}} = (1cc'1')/(1cc'1' + 1AA'1').$$

Although Revelle & Condon (2019) have previously recommended reporting ω_h, α, and ω_t for all scales in order to estimate reliability and ω_h for unidimensionality, unfortunately ω_h is not appropriate for very short scales. This is because ω_h requires at least 2 (and preferably ≥ 3) factors in order to find a *hierarchical* solution or even a bifactor structure. Since the degrees of freedom for a factor model with k factors and n variables is:

$$\frac{n(n-1)}{2} - nk + \frac{k(k-1)}{2}$$

the minimum number of variables needed for a 2 factor model to be defined is 5. Note that this case does not allow for a proper hierarchical solution using a Schmid-Leiman transformation for ω_h, as this requires 3 lower

[3]Specific and error are confounded unless using repeated measures to assess specific variance.

level factors (Zinbarg, Yovel, Revelle, & McDonald, 2006). The minimum number of variables needed to properly estimate a hierarchical solution is 6. Thus, in the case of $n \leq 4$ $\omega_h = \omega_t$.

Another index of the percent of general factor variance is *Explained Common Variance* (ECV) which compares the amount of variance extracted by the first factor to the amount explained by all factors (Rodriguez, Reise, & Haviland, 2016; Sijtsma, 2009; Ten Berge & Socan, 2004). ECV as discussed by Ten Berge & Socan (2004) compares the size of the eigenvalues extracted using *minimium rank factor analysis* (Shapiro & ten Berge, 2002; Ten Berge & Kiers, 1991). A nice set of examples comparing α to the ECV is given by Sijtsma (2009).

Yet one more way to test for unidimensionality is to ask how many factors best fit a scale. If the best estimate is greater than one, clearly the scale is not unidimensional. However, the number of factors problem is notoriously difficult (Horn & Engstrom, 1979) and variations on the method of parallel analysis Horn (1965) have spawned a small industry (Lorenzo-Seva, Timmerman, & Kiers, 2011; Revelle & Rocklin, 1979; Timmerman & Lorenzo-Seva, 2011; Velicer, 1976; Zwick & Velicer, 1986). However, even if these methods suggest one factor, they do not distinguish between the quality of the scale. Thus a completely homogeneous scale of six items with all correlations of .16 and and ECV of 1.00 (table 6 from Sijtsma, 2009) is seen as no different than the six political items from Ten Berge & Socan (2004) with an ECV of .82. Both appear to be one factor using parallel analysis with minimum rank factor analysis.

Here we introduce a very simple alternative test for unidimensionality — u, which may be found using the `unidim` function in the *psych* package in R — and examine its properties using both simulated and real data.

Unidim: a test for unidimensionality

The logic is deceptively simple: Unidimensionality implies that a one factor model of the data fits the covariances of the data. If this is the case, then the factor model implies that $R = \lambda\lambda' + U^2$ will have residuals of 0. That is, that the observed correlations will equal the model. To evaluate this, find the residuals of the observed correlations minus the modeled correlations, sum the off-diagonal elements of these squared residuals ($F_m = \sum_{i \neq j}(R_{ij} - \lambda\lambda)^2$), and compare this to the sum of off-diagonal elements of the squared original correlations ($F_o = \sum_{i \neq j} R_{ij}^2$). This is the measure of congeneric fit:

$$\rho_c = 1 - \frac{F_m}{F_o} = \frac{F_o - F_m}{F_o}.$$

When fitting a one factor model, ρ_c is a direct measure of the fit of a congeneric model to the observed correlations/covariances.

A statistically more elegant estimate is that of the Compararative Fit Index (Bentler, 1990), which compares a fit statistic (i.e., χ^2) of the model less the expected value ($F_m = \chi^2_m - dof_m$) to that of the original data ($F_o = \chi^2_o - dof_o$). Constraining both of the fits to be positive leads to the CFI:

$$CFI = \frac{max(\chi^2_0 - df_0), 0) - max(\chi^2_m - df_m), 0)}{max(\chi^2_0 - df_0), 0)}. \quad (7)$$

Clearly ρ_c is a direct index of the fit of a model, whereas the CFI is a comparison of the fit statistics. Although very similar in their values, we prefer the simplicity of ρ_c. In the following tables, we compare these two indices as well as a larger set of fit statistics.

ρ_c works well, but when some of the loadings are very small or differ drastically in their magnititude, it is probably not a good idea to think of the items as forming a unidimensional scale. Thus, an alternative model (the τ statistic) compares the observed correlations (r_{ij}) to the mean correlation ($\bar{r_{ij}}$) and considers 1 - the ratio of the sum of the squared residuals to the sum of the squared correlations:

$$\tau = 1 - \frac{\sum_{i \neq j}(r_{ij} - \bar{r_{ij}})^2}{\sum_{i \neq j} r_{ij}^2} \quad (8)$$

τ will achieve a maximum if the item covariances are all identical (a *tau* equivalent model, McDonald, 1999). Indeed, for dichotomous data, given the equivalence of factor analytic models using tetrachoric correlations with 2PL IRT models (Kamata & Bauer, 2008; Takane & de Leeuw, 1987), the Rasch (1966) model — a one parameter IRT model with the assumption that all items have equally good discrimination — is functionally a τ equivalent model.

The product of ρ_c and τ is the measure of unidimensionality, u. That is, congeneric fit x tau equivalent fit as a measure of unidimensionality. In the following tables, we show how u behaves in various simulations as well as with real data. We also show the behavior of the u statistic as a function of sample size and compare the standard errors as a function of sample sizes (Figure 1). These demonstrations use the functions `unidim` and `omega` as implemented in the *psych* package (Revelle, 2024a) for the open source statistical system R (R Core Team, 2023). Output from both functions is also shown in the `reliability` function.

Tests with simulations

To demonstrate the `unidim` function, we simulated four one-factor models for 3, 6, 12 and 24 items. Loadings for each model were specified for a three-item model as large (.7, .6, .5), medium (.6, .5, .4), small (.5,

.4, .3), or mixed (.7, .5, .3) loadings. To form 6, 12, or 24 item structure, the basic loadings were repeated 2, 4, and 8 times. We used the `sim` and `sim.minor` functions to generate the data. Both functions generate a latent variable model by multiplying the factor loading matrix by a matrix of random normal deviates and then adding normally distributed error. `sim.minor` follows the advice of MacCallum & Tucker (1991) who distinguished between the factor model we want (pure factors) and a generating model of pure factors with a number of smaller, nuisance factors. (For further examples of large and small factors, see Lorenzo-Seva et al., 2011; Timmerman & Lorenzo-Seva, 2011).

For ease of simulation, we formed data based upon the first four columns (pure) or all eight columns (minor) of Table 2. By partitioning the resulting correlation matrix appropriately, we are thus able to generate 16 different one-factor models. We show loadings for each of the models. The first set (column 1) contain large loadings (.7, .6, .5), the second medium (.6, .5, .4) loadings, and the third small (.5, .4, .3) loadings. The fourth column shows mixed loadings of .7, .5, and .3. The final four columns were used when generating minor factors, with loadings of ±.2 randomly assigned to variables. The partitioning of the overall 96 x 96 correlation matrix resulted in e.g., a 3 x 3 correlation matrix with large loadings (R[1:3,1:3], medium loadings (R[4:6;4:6], or a 6 x 6 with mixed loadings (R[c(10:12,22:24), c(10:12,22:24)]), etc.

Table 1

Four simulated loadings matrices. For each model, one of the first 4 columns was combined with columns 5-8. The loadings represent large, medium and small loadings, as well as a mixed set. The minor factors had loadings of ±.2 for four nuisance factors. To simulate an (e.g.) six item problem with medium loadings, the first six rows of the second and 5-8th columns were used.

Item	large	middle	small	mixed	m1	m2	m3	m4
1	0.7	0.6	0.5	0.7	0.0	0.0	0.2	-0.2
2	0.6	0.5	0.4	0.5	0.0	0.2	0.0	0.0
3	0.5	0.4	0.3	0.3	-0.2	0.0	0.0	0.2
4	0.7	0.6	0.5	0.7	0.2	0.2	0.2	0.2
5	0.6	0.5	0.4	0.5	0.0	-0.2	0.0	0.0
6	0.5	0.4	0.3	0.3	0.0	0.2	0.0	0.2
	...	...	...	..				
22	0.7	0.6	0.5	0.7	-0.2	0.0	-0.2	0.2
23	0.6	0.5	0.4	0.5	0.0	0.0	0.2	-0.2
24	0.5	0.4	0.3	0.3	-0.2	0.0	0.0	0.0

Data were generated for 500 simulated subjects using both the "pure" (just one of the first four columns) and the "noisy" (one + four noise factors) model. Solutions for 3, 6, 12, and 24 items per scale for 500 simulated participants are shown for pure factors in Table 3 and for noisy data in Table 4. The last four rows reflect scales formed from the first and second factors for 6, 10, 12, and 24 items[4]. These are clearly not unidimensional. Several things to note in these tables: Following the Spearman-Brown equation (Brown, 1910; Spearman, 1910), α and ω_t increase with the number of items in the scale. Neither statistic flags the scales formed from two orthogonal factors as poor fits. Because a hierarchical model is not identified for three item scales, ω_h was forced to a one factor solution for those scales but properly identifies the last four scales as having low values for general factor saturation. The behavior of the u statistic is very gratifying, in that it does not increase with the number of items per scale, varies as a function of the the range of loadings, and correctly identifies the last four scales as non-unidimensional. For these examples, the performance of the CFI and ECV statistics showed similar patterns: high for unidimensional scales, lower for multidimensional scales. This is in striking contrast to ω_t or α which show very "reasonable" values for these non-unidimensional scales.

Applying `unidm` to real data

The simulations were done with continuous item scores. However, most real ability and personality items are categorical. `unidim` can be applied to such data using either tetrachoric or more generally polychoric correlations. In addition to showing results with simulated continuous and categorical data (Table 5) we show the utility of the unidim statistic with real categorical data from three data sets available in the *psychTools* package (Revelle, 2024b) for the R statistical system. For reasons discussed in Chalmers (2017) and Revelle & Condon (2019), conventional reliability indices were found using Pearson correlations. However, the unidimensional estimates were found from factoring the tetrachoric or polychoric correlation matrices.

Of these data sets, the first, the `ability` data set includes 16 items for 1,525 participants from the International Cognitive Ability Resource (Condon & Revelle, 2014), which represents 4 lower level factors and one higher level factor (Table 6 part 1). The items are dichotomous. The second dataset, `bfi`, contains data from 2,800 participants who responded to 25 Likert-like items with six response choices ranging from "very inaccurate" to "very accurate." Although normally scored using only five separate constructs (the familiar Big Five traits shown in Table 6 part 2), composite scales were also formed here, for demonstration purposes, from two (E+O), three (A+C+N) or five (all) of these constructs. This is a particularly nice example of the advantage of the

[4]Because we also are finding split half estimates, we limited our examples to 24 items to allow for finding all split half values from the 2,704,156 possible splits.

Table 2

The simulated loadings matrix. Rows 1-12 were repeated 8 times to generate the 96 item loadings. The loadings represent large, medium and small loadings, as well as a mixed set. The minor factors had loadings of ±.2 for four nuisance factors. The resulting 96 96 correlation matrix is then partitioned into unifactorial subsets. (E.g., R[1:3,1:3] represents the correlation matrix of 3 items with large correlations, R[4:6,4:6] the medium sized correlation matrix.*

Variable	Large	Medium	Small	Mixed	m1	m2	m3	m4
V1	0.7	0.0	0.0	0.0	0.0	0.0	0.2	-0.2
V2	0.6	0.0	0.0	0.0	0.0	0.2	0.0	0.0
V3	0.5	0.0	0.0	0.0	-0.2	0.0	0.0	0.2
V4	0.0	0.6	0.0	0.0	0.2	0.2	0.2	0.2
V5	0.0	0.5	0.0	0.0	0.0	-0.2	0.0	0.0
V6	0.0	0.4	0.0	0.0	0.0	0.2	0.0	0.2
V7	0.0	0.0	0.5	0.0	0.2	0.2	0.0	0.0
V8	0.0	0.0	0.4	0.0	-0.2	0.0	-0.2	0.2
V9	0.0	0.0	0.3	0.0	0.0	0.2	0.0	0.2
V10	0.0	0.0	0.0	0.7	-0.2	0.0	0.2	0.0
V11	0.0	0.0	0.0	0.5	0.0	0.0	0.2	-0.2
V12	0.0	0.0	0.0	0.3	-0.2	0.0	0.0	0.0
	...	...	...	...	...	...	...	...
V85	0.7	0.0	0.0	0.0	0.0	0.0	0.2	0.0
V86	0.6	0.0	0.0	0.0	0.0	0.0	0.0	-0.2
V87	0.5	0.0	0.0	0.0	0.2	0.2	0.0	0.0
V88	0.0	0.6	0.0	0.0	-0.2	0.2	0.0	0.0
V89	0.0	0.5	0.0	0.0	-0.2	0.0	0.0	0.2
V90	0.0	0.4	0.0	0.0	0.2	0.0	-0.2	-0.2
V91	0.0	0.0	0.5	0.0	0.0	0.0	0.2	-0.2
V92	0.0	0.0	0.4	0.0	-0.2	0.0	0.0	0.0
V93	0.0	0.0	0.3	0.0	0.2	-0.2	0.2	0.0
V94	0.0	0.0	0.0	0.7	0.0	0.2	0.2	0.0
V95	0.0	0.0	0.0	0.5	0.0	0.2	0.0	0.0
V96	0.0	0.0	0.0	0.3	0.0	0.0	-0.2	0.2

Table 3

Various estimates of unidimensionality and reliability for 500 simulated participants for scales formed from 3, 6, 12, and 24 items with factor loadings as specified in Table 2. The last four rows report results for scales formed from two orthogonal subscales. u *is the unidimensionality statistic, τ and ρ_C are the τ and congeneric fits, ω_h and ω_t are the two omega statistics, α is the traditional estimate. Max split and min split represent the maximum and minimum split half reliabilities found by complete sampling of all $C^k_{k/2}$ possible split half coefficients. $\bar{r}$ reports the mean correlation in the scale. Median r is just the median correlation among the items. CFI is the comparative fit index, ECV is the Explained Common Variance.*

hline Variable	u	τ	ρ_C	ω_h	α	ω_t	Max split	Min split	$\bar{r}$	Median r	CFI	ECV	N items
high.3	0.99	0.99	1.00	0.61	0.61	0.61	0.58	0.52	0.34	0.36	1.00	0.98	3
med.3	0.99	0.99	1.00	0.54	0.54	0.54	0.50	0.45	0.28	0.28	1.00	0.98	3
low.3	0.97	0.97	1.00	0.34	0.33	0.34	0.32	0.26	0.14	0.14	1.00	0.97	3
mixed.3	0.92	0.92	1.00	0.45	0.44	0.45	0.44	0.31	0.21	0.18	1.00	0.97	3
high.6	0.96	0.96	0.99	0.69	0.76	0.78	0.78	0.73	0.35	0.33	0.99	0.74	6
med.6	0.91	0.92	0.99	0.41	0.67	0.74	0.70	0.65	0.26	0.26	1.00	0.73	6
low.6	0.85	0.88	0.97	0.21	0.52	0.62	0.56	0.47	0.15	0.15	1.00	0.69	6
mixed.6	0.73	0.74	0.98	0.26	0.61	0.71	0.66	0.55	0.21	0.18	0.99	0.55	6
high.12	0.94	0.95	0.99	0.76	0.87	0.87	0.89	0.84	0.35	0.35	0.99	0.90	12
med.12	0.88	0.90	0.98	0.38	0.78	0.81	0.82	0.75	0.23	0.23	0.97	0.74	12
low.12	0.78	0.83	0.94	0.48	0.66	0.68	0.71	0.57	0.14	0.14	0.99	0.76	12
mixed.12	0.77	0.78	0.98	0.26	0.77	0.81	0.82	0.71	0.22	0.20	0.99	0.73	12
high.24	0.95	0.95	1.00	0.11	0.93	0.93	0.95	0.92	0.36	0.35	1.00	0.93	24
med.24	0.90	0.92	0.98	0.64	0.88	0.88	0.91	0.85	0.23	0.23	0.99	0.88	24
low.24	0.81	0.85	0.95	0.66	0.82	0.82	0.86	0.77	0.16	0.16	0.99	0.84	24
mixed.24	0.78	0.79	0.98	0.36	0.88	0.89	0.91	0.82	0.24	0.21	0.99	0.90	24
F2.6	0.32	0.50	0.63	0.11	0.50	0.61	0.59	0.12	0.14	0.07	0.61	0.54	6
F2.10	0.30	0.49	0.62	0.07	0.66	0.73	0.75	0.09	0.16	0.06	0.61	0.46	10
F2.12	0.12	0.19	0.63	0.05	0.62	0.73	0.73	0.09	0.12	0.05	0.62	0.50	12
F2.24	0.23	0.34	0.68	0.08	0.79	0.84	0.87	0.32	0.14	0.07	0.65	0.58	24

Table 4

Various estimates of unidimensionality and reliability for 500 simulated participants for scales formed from 3, 6, 12, and 24 items with major and minor factor loadings as specified in Table 2. The last four rows report results for scales formed from two orthogonal subscales. Column headings are the same as in Table 3. The addition of a small amount of noise makes very little difference (compare with Table 3).

Variable	u	τ	ρ_C	ω_h	α	ω_t	Max split	Min split	$\bar{r}$	Median r	CFI	ECV	N items
high.3	0.98	0.98	1.00	0.59	0.58	0.59	0.56	0.47	0.32	0.30	1.00	0.98	3
med.3	0.99	0.99	1.00	0.53	0.53	0.53	0.51	0.44	0.27	0.29	1.00	0.97	3
low.3	0.90	0.90	1.00	0.43	0.39	0.43	0.43	0.28	0.18	0.19	1.00	0.91	3
mixed.3	0.91	0.91	1.00	0.54	0.52	0.54	0.53	0.36	0.27	0.22	1.00	0.98	3
high.6	0.93	0.93	1.00	0.43	0.76	0.82	0.78	0.74	0.35	0.34	1.00	0.74	6
med.6	0.92	0.94	0.97	0.55	0.67	0.70	0.71	0.61	0.26	0.24	0.96	0.76	6
low.6	0.82	0.84	0.97	0.39	0.54	0.62	0.59	0.47	0.16	0.15	0.99	0.61	6
mixed.6	0.80	0.81	0.99	0.56	0.68	0.72	0.71	0.61	0.26	0.22	0.99	0.79	6
high.12	0.93	0.94	0.99	0.65	0.86	0.87	0.89	0.83	0.34	0.34	0.97	0.88	12
med.12	0.91	0.94	0.97	0.66	0.80	0.81	0.84	0.74	0.25	0.25	0.94	0.81	12
low.12	0.79	0.84	0.95	0.49	0.69	0.70	0.75	0.61	0.16	0.16	0.96	0.77	12
mixed.12	0.81	0.83	0.99	0.30	0.81	0.83	0.85	0.72	0.26	0.23	0.99	0.75	12
high.24	0.93	0.94	0.99	0.60	0.92	0.93	0.95	0.89	0.34	0.34	0.95	0.89	24
med.24	0.89	0.91	0.97	0.64	0.89	0.89	0.92	0.83	0.25	0.25	0.95	0.86	24
low.24	0.77	0.83	0.93	0.58	0.82	0.83	0.87	0.74	0.16	0.16	0.90	0.79	24
mixed.24	0.80	0.82	0.98	0.67	0.89	0.90	0.92	0.82	0.26	0.23	0.98	0.86	24
F2.6	0.06	0.11	0.57	0.02	0.46	0.56	0.56	0.03	0.12	0.05	0.57	0.52	6
F2.10	0.30	0.48	0.63	0.07	0.65	0.73	0.76	0.09	0.16	0.09	0.62	0.52	10
F2.12	0.31	0.47	0.64	0.07	0.68	0.74	0.77	0.10	0.15	0.10	0.63	0.53	12
F2.24	0.30	0.47	0.62	0.06	0.81	0.84	0.88	0.29	0.15	0.11	0.58	0.53	24

Table 5

Comparing unidim estimates for continuous and item (categorical) data for pure and minor simulated factors. The categorical data were generated with six categories using the `sim` and `sim.minor` functions in psych.

Variable	pure continuous	pure items	minor continuous	minor items
high.3	0.99	0.98	0.98	0.97
med.3	0.99	0.97	0.99	0.96
low.3	0.97	0.83	0.90	0.94
mixed.3	0.92	0.87	0.91	0.86
high.6	0.96	0.96	0.93	0.94
med.6	0.91	0.95	0.92	0.95
low.6	0.85	0.84	0.82	0.81
mixed.6	0.73	0.84	0.80	0.76
high.12	0.94	0.94	0.93	0.93
med.12	0.88	0.92	0.91	0.93
low.12	0.78	0.79	0.79	0.83
mixed.12	0.77	0.81	0.81	0.73
high.24	0.95	0.95	0.93	0.94
med.24	0.90	0.90	0.89	0.91
low.24	0.81	0.79	0.77	0.82
mixed.24	0.78	0.80	0.80	0.74
F2.6	0.32	0.06	0.06	0.32
F2.10	0.30	0.09	0.30	0.27
F2.12	0.12	0.10	0.31	0.28
F2.24	0.23	0.22	0.30	0.31

u statistic as contrasted with the more conventional α and ω_t statistics, for the latter show quite reasonable values (.71 - .84) for scales that are not unidimensional. u and ω_h on the other hand, show clear evidence (.30 - .41) for multidimensionality (Table 6 part 3). What is interesting is that for these nominally unifactorial scales, the CFI and ECV statistics were quite low. The ECVs for the single construct `bfi` scales ranged from .79 to .87, values, which, although larger than for the multi-construct scales (.52 - .61), do not suggest strong unidimensionality.

The third dataset uses the 135 items of the SAPA Personality Inventory (Condon, 2018) for 4,000 participants (`spi`) to form 27 lower level scales; these also had six response options ("very inaccurate" to "very accurate") (Table 7). In this case, the u values for the SPI-27 reflect the higher degree of unidimensionality expected from brief 5-item scales with a median of .94 and ranging form .83 to .99. By contrast, α values are relatively lower (again, as expected, given the short scale lengths), and the ω_h values are difficult to interpret due to inadequate degrees of freedom. In contrast to the findings with the `bfi` scales, these 27 scales show higher levels of ECV with a median of .91 and ranging from .77 to .96.

Sensitivity to sample size

For practical purposes, we addressed the question of the effect of sample size on the u statistic. We simulated 200, 500, 1,000 and 5,000 participants using the factor structure shown in Table 2. For each simulation we performed 100 replications. We also examined the effect of sample size on ω_t, CFI and ECV. It is quite clear (Figure 1) that even for samples as small as 200, the u statistic could distinguish between unidimensional scales versus multidimensional scales. The pattern of results show that u is, in contrast to ω_t, not sensitive to the number of items in the scale, but is sensitive to unidimensionality. This is evident from comparisons of the first 16 to the last four

Table 6

Unidim coefficients and multiple reliability measures for the `ability` *data set with 16 items and 1,525 participants and the* `bfi` *data set with 2,800 participants for 25 items. The first three columns report* `unidim` *statistics* u, τ, ρ_c, *the next three columns the conventional* ω_h, α, *and* ω_t *internal consistency estimates, and the next two report the maximum and minimum split half reliabilities based upon a split half decomposition of the scales. The next two columns (*$\bar{r}$ *and Median* r*) are the mean and median within scale correlations, respectively. The last three columns are the CFI and ECV, followed by the number of items per scale. The ICAR16 represents a higher order factor composed of four lower level factors. Although the* `bfi` *is scored as five unidimensional constructs, the last three lines represent scales formed from two (E+O), three (A+C+N) or five (all) constructs. For the latter three scales, note how the high value of* α *and* ω_t *conflict with low values of* u *and* ω_h *which are better indicators of unidimensionality.*

Variable	u	τ	ρ_c	ω_h	α	ω_t	Max split	Min split	$\bar{r}$	Median r	CFI	ECV	N items
ICAR dataset													
ICAR16	0.90	0.93	0.96	0.56	0.83	0.85	0.87	0.73	0.23	0.21	0.76	0.78	16
reasoning	0.98	0.98	1.00	0.05	0.65	0.65	0.65	0.64	0.31	0.31	1.00	0.99	4
letters	0.99	0.99	1.00	0.65	0.67	0.68	0.68	0.66	0.34	0.33	1.00	0.98	4
matrix	0.93	0.95	0.98	0.37	0.54	0.59	0.58	0.49	0.23	0.22	0.95	0.83	4
rotate	1.00	1.00	1.00	0.71	0.77	0.78	0.78	0.74	0.45	0.44	0.99	0.96	4
bfi dataset: single construct scales													
extraversion	0.96	0.97	0.99	0.55	0.76	0.82	0.78	0.66	0.39	0.38	0.97	0.86	5
openness	0.89	0.91	0.98	0.39	0.61	0.70	0.66	0.52	0.24	0.23	0.94	0.79	5
agreeableness	0.90	0.91	0.99	0.64	0.71	0.75	0.75	0.62	0.33	0.34	0.96	0.86	5
conscientious	0.95	0.97	0.98	0.53	0.73	0.77	0.74	0.64	0.35	0.34	0.93	0.84	5
neuroticism	0.94	0.95	0.98	0.71	0.81	0.85	0.83	0.69	0.47	0.41	0.90	0.87	5
bfi multi construct scales													
EO	0.49	0.62	0.79	0.36	0.71	0.77	0.79	0.38	0.20	0.21	0.64	0.61	10
ACN	0.43	0.63	0.68	0.33	0.78	0.81	0.87	0.47	0.19	0.14	0.44	0.52	15
All	0.38	0.56	0.68	0.30	0.82	0.84	0.89	0.58	0.15	0.13	0.40	0.56	25

Table 7

Unidim coeffficients and multiple reliability measures of the `spi-135` *(Condon, 2018). The SAPA Personality Inventory (Condon, 2018) has five higher order scales assessing the "Big Five" and 27 lower level scales assessing other aspects of personality. The 27 lower level measures have 5 items each. The* `spi` *data set in the psychTools has 4,000 observations on these 135 items plus 10 criteria/demographic variables. The columns are the same as in Table 6.*

Variable	u	τ	ρ_c	ω_h	α	ω_t	Max split	Min split	$\bar{r}$	Median r	CFI	ECV	N items
Compassion	0.99	0.99	1.00	0.80	0.88	0.89	0.87	0.82	0.59	0.58	0.99	0.94	5
Trust	0.99	0.99	1.00	0.80	0.87	0.89	0.87	0.81	0.58	0.58	0.98	0.94	5
Honesty	0.96	0.97	0.99	0.71	0.81	0.84	0.83	0.70	0.46	0.46	0.94	0.88	5
Conservatism	0.84	0.91	0.92	0.56	0.78	0.85	0.84	0.61	0.41	0.35	0.66	0.76	5
Authoritarianism	0.89	0.93	0.96	0.63	0.81	0.86	0.85	0.63	0.46	0.46	0.83	0.81	5
EasyGoingness	0.90	0.92	0.98	0.45	0.68	0.76	0.73	0.58	0.29	0.29	0.94	0.78	5
Perfectionism	0.83	0.84	0.99	0.34	0.70	0.74	0.72	0.53	0.31	0.33	0.97	0.87	5
Order	0.93	0.94	0.99	0.62	0.81	0.85	0.83	0.66	0.46	0.42	0.93	0.86	5
Industry	0.99	0.99	1.00	0.72	0.84	0.86	0.84	0.76	0.52	0.50	0.98	0.94	5
Impulsivity	0.98	0.98	1.00	0.72	0.87	0.90	0.87	0.80	0.58	0.58	0.99	0.96	5
SelfControl	0.91	0.94	0.96	0.49	0.76	0.83	0.80	0.60	0.39	0.36	0.87	0.79	5
EmotionalStability	0.99	0.99	1.00	0.65	0.85	0.89	0.84	0.76	0.52	0.50	0.99	0.93	5
Anxiety	0.99	0.99	1.00	0.83	0.90	0.91	0.89	0.83	0.64	0.62	0.97	0.95	5
Irritability	0.98	0.99	0.99	0.78	0.89	0.91	0.89	0.79	0.61	0.60	0.94	0.91	5
WellBeing	0.99	0.99	1.00	0.80	0.90	0.92	0.90	0.81	0.63	0.63	0.95	0.92	5
EmotionalExpressiveness	0.93	0.94	0.99	0.73	0.80	0.83	0.83	0.68	0.45	0.43	0.95	0.90	5
Sociability	0.97	0.98	0.99	0.66	0.85	0.89	0.85	0.75	0.53	0.50	0.95	0.88	5
Adaptability	0.93	0.94	0.99	0.62	0.80	0.84	0.82	0.68	0.44	0.42	0.95	0.88	5
Charisma	0.94	0.96	0.98	0.67	0.82	0.86	0.84	0.72	0.47	0.43	0.88	0.85	5
Humor	0.94	0.94	0.99	0.68	0.78	0.82	0.81	0.64	0.42	0.40	0.97	0.91	5
AttentionSeeking	0.93	0.93	0.99	0.80	0.88	0.90	0.89	0.77	0.58	0.67	0.94	0.92	5
SensationSeeking	0.97	0.98	0.99	0.77	0.86	0.89	0.87	0.77	0.55	0.54	0.92	0.91	5
Conformity	0.90	0.94	0.96	0.67	0.82	0.87	0.85	0.67	0.47	0.47	0.82	0.83	5
Introspection	0.94	0.95	0.99	0.56	0.78	0.84	0.81	0.68	0.41	0.41	0.94	0.87	5
ArtAppreciation	0.90	0.90	1.00	0.68	0.80	0.83	0.81	0.65	0.44	0.46	0.98	0.92	5
Creativity	0.97	0.98	1.00	0.70	0.85	0.86	0.85	0.77	0.52	0.53	0.98	0.94	5
Intellect	0.99	0.99	1.00	0.81	0.86	0.87	0.84	0.78	0.54	0.52	0.99	0.96	5

columns of each panel. The CFI statistic was not sensitive to sample size nor range of factor loadings for one factor data but did drop when given two factor data. The ECV was sensitive to sample size, increasing with larger samples, and to factor structure (decreasing from factors with large to medium to low loadings). In contrast to u, with the exception of 3 item scales, ECV was sensitive to the number of items per factor, increasing from 6 to 12 to 24.

Comparison of fit statistics

A reasonable question to ask is what is the relationship between these various estimates. As one reviewer questioned, is our ρ_c measure any different from ECV? Conceptually our proposed measure, u, based upon the product of τ and ρ_c would seem to be similar to the goodness of fit of a one factor model (CFI) or the Explained Common Variance. To address this question, we report the correlations of these measures for two our data sets (Table 8): the results for 500 simulated subject across 24 different conditions shown in Table 3 and for the 27 5 item scales for 4,000 real subjects from the `spi` data set (Table 7). For the simulated data, ρ_c correlated .99 and .77 with CFI and ECV, while for the `spi` data, the correlations were .97 and .87. The correlations with the unidim u statistic were .95 and .83 for the simulated data and .61 and .75 for the `spi` data.

Table 8

Correlations of fit statistics: Lower off diagonal is for the analysis of the `spi` data. Upper off diagonal is for the simulation of pure data with 500 subjects.

Variable	Uni	τ	ρ_c	ω_h	CFI	ECV
Uni		0.99	0.96	0.71	0.95	0.83
τ	0.95		0.92	0.71	0.92	0.81
ρ_c	0.70	0.44		0.66	0.99	0.77
ω_h	0.76	0.76	0.45		0.66	0.65
CFI	0.61	0.34	0.97	0.30		0.76
ECV	0.75	0.56	0.87	0.70	0.82	

Summary and Conclusion

The problem of assessing the unidimensionality of a scale has been a challenge for many years, and many solutions have been offered. Here we have suggested a simple index, u, which is the the product of indices of τ equivalance and ρ_c or congeneric equivalence. u is simple to calculate (e.g., the `unidim` or `reliability` functions in the *psych* package for R). We compare the u index to five popular indices of scale quality, α, ω_h ω_t, CFI, and ECV show that unlike α, ω_t, or ECV, u is insensitive to the number of items in a scale and unlike the CFI is sensitive to factor structure. u is robust across sample sizes from 200 to 5000 in identifying unidimensionality.

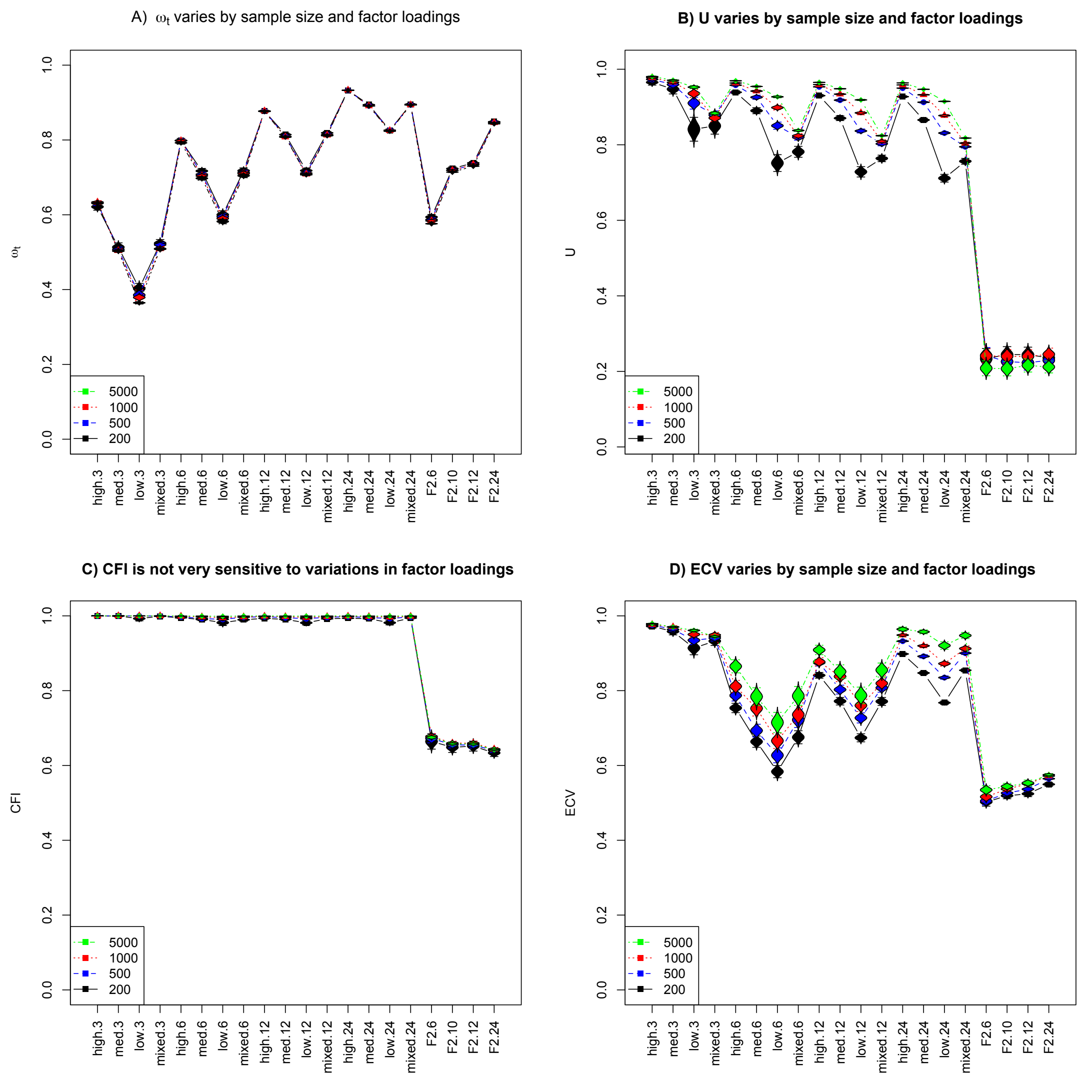

Figure 1

The ω_t, u, CFI *and* ECV *statistics behave very differently across the number of items per factor and sample size. Panel A shows how ω_t increases with the number of items/factor and varies by the size of the factor loadings. It is relatively insensitive to the non-unidimensionality of scales formed from multiple factors and to sample size. Panel B shows how the* u *statistic does not increase with the number of items per scale, is sensitive to the the size of the factor loadings, and is very sensitive to the degree of non-unidimensionality (the four right hand observations in both panels). Panel C shows that the* CFI *varies only slightly by sample or range of factor loadings. Panel D shows that ECV varies by sample size and factor loadings. Sample sizes in all panels range from 200 (black) to 500, 1000, and 5000 (green) simulated participants. The cats-eyes shapes display standard deviations of 100 replications for each sample size.*

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin*, *107*(2), 238-246.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, *3*(3), 296-322. doi: 10.1111/j.2044-8295.1910.tb00207.x

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. doi: 10.18637/jss.v048.i06

Chalmers, R. P. (2017). On misconceptions and the limited usefulness of ordinal alpha. *Educational and Psychological Measurement*, *78*(6), 1-16. doi: 10.1177/0013164417727036

Cho, E., & Kim, S. (2015). Cronbach's coefficient alpha: Well known but poorly understood. *Organizational research methods*, *18*(2), 207-230. doi: 10.1177/1094428114555994

Condon, D. M. (2018). The SAPA Personality Inventory: An empirically-derived, hierarchically-organized self-report personality assessment model. *PsyArXiv*. doi: 10.31234/osf.io/sc4p9

Condon, D. M., & Revelle, W. (2014). The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. *Intelligence*, *43*, 52-64. doi: 10.1016/j.intell.2014.01.004

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334. doi: 10.1007/BF02310555

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255–282. doi: 10.1007/BF02288892

Hattie, J. (1984). An empirical study of various indices for determining unidimensionality. *Multivariate Behavioral Research*, *19*(1), 49-78. doi: /10.1207/s15327906mbr1901_3

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and ltems. *Applied Psychological Measurement*, *9*, 139-164. doi: 10.1177/014662168500900204

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. doi: 10.1007/BF02289447

Horn, J. L., & Engstrom, R. (1979). Cattell's scree test in relation to Bartlett's chi-square test and other observations on the number of factors problem. *Multivariate Behavioral Research*, *14*(3), 283-300. doi: 10.1207/s15327906mbr1403_1

Kamata, A., & Bauer, D. J. (2008). A note on the relation between factor analytic and item response theory models. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(1), 136-153. doi: 10.1080/10705510701758406

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

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The hull method for selecting the number of common factors. *Multivariate Behavioral Research*, *46*(2), 340-364. doi: 10.1080/00273171.2011.564527

Lucke, J. F. (2005). "rassling the hog": The influence of correlated item error on internal consistency, classical reliability, and congeneric reliability. *Applied psychological measurement*, *29*(2), 106-125. doi: 10.1177/0146621604272

MacCallum, R. C., & Tucker, L. R. (1991). Representing sources of error in the common-factor model: Implications for theory and practice. *Psychological Bulletin*, *109*(3), 502-511. doi: 10.1037/0033-2909.109.3.502

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, N.J.: L. Erlbaum Associates.

McNemar, Q. (1946). Opinion-attitude methodology. *Psychological Bulletin*, *43*(4), 289-374. doi: 10.1037/h0060985

R Core Team. (2023). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rasch, G. (1966). An item analysis which takes individual differneces in to account. *British Journal of Mathematical and Statistical Psychology*, *19*(1), 49–57. doi: 10.1111/j.2044-8317.1966.tb00354.x

Revelle, W. (2024a). psych:procedures for psychological, psychometric, and personality research (2.4.3 ed.) [Computer software manual]. https://CRAN.r-project.org/package=psych. (R package version 2.4.3)

Revelle, W. (2024b). psychTools tools to accompany the psych package for psychological research [Computer software manual]. psychTools. (R package version 2.4.3)

Revelle, W., & Condon, D. M. (2019). Reliability from α to ω: A tutorial. *Psychological Assessment.*, *31*(12), 1395-1411. doi: 10.1037/pas0000754

Revelle, W., & Rocklin, T. (1979). Very Simple Structure - alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research*, *14*(4), 403-414. doi: 10.1207/s15327906mbr1404_2

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological methods*, *21*(2), 137-150. doi: 10.1037/met0000045

Schmid, J. J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*(1), 83-90. doi: 10.1007/BF02289209

Shapiro, A., & ten Berge, J. M. (2002). Statistical inference of minimum rank factor analysis. *Psychometrika*, *67*(1), 79-94. doi: 10.1007/BF02294710

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107-120. doi: 10.1007/s11336-008-9101-0

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *American Journal of Psychology*, *15*(2), 201-292. doi: 10.2307/1412107

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, *3*(3), 271-295. doi: 10.1111/j.2044-8295.1910.tb00206.x

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*, 393-408. (10.1007/BF02294363) doi: 10.1007/BF02294363

Ten Berge, J. M. F., & Socan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, *69*(4), 613-625. doi: 0033-3123/2004-4/2002-0997-A

Ten Berge, J. M., & Kiers, H. A. (1991). A numerical approach to the approximate and the exact minimum rank of a covariance matrix. *Psychometrika*, *56*, 309–315. doi: 10.1007/BF02294464

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological methods*, *16*(2), 209. doi: 10.1037/a0023353

Velicer, W. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327. doi: 10.1007/BF02293557

Walker, D. A. (1931). Answer-pattern and score-scatter in tests and examinations. *British Journal of Psychology. General Section*, *22*(1), 73-86. doi: 10.1111/j.2044-8295.1931.tb00615.x

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 alternative conceptualizations of reliability. *Psychometrika*, *70*(1), 123-133. doi: 10.1007/s11336-003-0974-7

Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ω_h. *Applied Psychological Measurement*, *30*(2), 121-144. doi: 10.1177/0146621605278814

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*(3), 432 - 442. doi: 10.1037/0033-2909.99.3.432

Appendix: R code

R code

```
library(psych)
library(psychTools)
#requires psych version 2.4.3 or above
#The basic data structure

fx4 <- matrix(c(.7,.6,.5,rep(0,12),.6,.5,.4,rep(0,12) ,.5,.4,.3,rep(0,12),.7,.5,.3),ncol=4)
fx96 <- rbind(fx4,fx4,fx4,fx4,fx4,fx4,fx4,fx4)
rownames(fx96) <- paste0("V",1:96)
colnames(fx96) <- paste0("F",1:4)
 keys96 <- factor2cluster(fx96)
 rownames(keys96) <- paste0("V",1:96)
 keys.96 <- keys2list(keys96)
 keys.12 <-keys2list(keys96[1:12,])
keys.24 <- keys2list(keys96[1:24,])
 keys.48 <- keys2list(keys96[1:48,])
# names(keys.12) <- paste0("F",1:4,".12")
names(keys.12) <- paste0(cs(high, med, low, mixed),".",3)
names(keys.24) <- paste0(cs(high, med, low, mixed),".",6)
 names(keys.48) <- paste0(cs(high, med, low, mixed),".",12)
names(keys.96) <- paste0(cs(high, med, low, mixed),".",24)
keys.6.2.factors <- list(c(keys.12$high,keys.12$med))
keys.10.2.factors <- list(c(keys.24$high[1:5],keys.24$med[1:5]))
keys.12.2.factors <- list(c(keys.48$high[1:6],keys.48$med[1:6]))
keys.24.2.factors <- list(c(keys.96$high[1:12],keys.96$med[1:12]))
names(keys.6.2.factors) <- "F2.6"
names(keys.10.2.factors) <- "F2.10"
names(keys.12.2.factors) <- "F2.12"
names(keys.24.2.factors) <- "F2.24"
#names(keys.48.2.factors) <- "F2.48"
 keys <- c(keys.12,keys.24,keys.48,keys.96,keys.6.2.factors,keys.10.2.factors,
       keys.12.2.factors,keys.24.2.factors)

#one run to create Tables 3 and 4
n.obs=500
set.seed(42)  #for reproducible results
sim.96 <- sim(fx96,n=n.obs)  #generate n.obs participants for 96 variables
minor.96  <- sim.minor(fbig=fx96,n=n.obs, n.small=4) #include nuisance factors
#create item like data
sim.96.item <- sim(fx96,n=n.obs, items=TRUE)  #generate n.obs participants for 96 variables
minor.96.item  <- sim.minor(fbig=fx96,n=n.obs, n.small=4,items=TRUE) #include nuisance factors

reliability.pure.500 <- reliability(keys, sim.96$observed) #
reliability.minor.500 <- reliability(keys, minor.96$observed,n.sample=50000)


uni.pure.cont <- unidim(keys,sim.96$observed)
uni.minor.cont <-  unidim(keys,minor.96$observed)
#do it for categoriical data
uni.pure.item <- unidim(keys,sim.96.item$observed,cor="poly")
uni.minor.item  <- unidim(keys,minor.96.item$observed,cor="poly")

uni.sum <- data.frame(pure.cont=uni.pure.cont$uni[,1:3], pure.item=uni.pure.item$uni[,1:3],
            minor.cont = uni.minor.cont$uni[,1:3], minor.item = uni.minor.item$uni[,1:3])

 #now, a cleaner table, just the u values
 uni.sum.1<- data.frame(pure.cont=uni.pure.cont$uni[,1], pure.item=uni.pure.item$uni[,1],
            minor.cont = uni.minor.cont$uni[,1], minor.item = uni.minor.item$uni[,1])

matPlot(uni.sum.1, xlas=3 ,legend=3,
       main="Unidim as a function of number of items, factor loadings and item type")

#now try to get confidence intervals on the basic stats
# we do 100 runs for each condition

# We create a short function to simulate the data
#It uses the sim functions from psych
#also the reliability and unidim functions
#use 10 replications for testing purposes

simulate.unidim <- function(fx,n.obs=1000, n.replications = 10, keys=NULL, minor=FALSE, items=FALSE) {
        results <- list()
        uni.results <- list()
        vnames <- cs(omega.h ,alpha, omega.tot,  ECV, Uni, r.fit ,fa.fit ,n.items)
        if(is.null(keys)) {nvar <- 1} else {nvar  <-  length(keys)}

for (i in (1: n.replications)) {  # short loop
        if(minor) {sim.data <- sim.minor(fbig=fx, n=n.obs, n.small=4,items=items) } else {sim.data <- sim(fx,n=n.obs, items=items
```

```
        if(items) {R <- lowerCor(sim.data$observed, cor="poly", show=FALSE)$rho } else {R <- lowerCor(sim.data$observed,show=FALS

        temp  <- reliability(keys,R,split=FALSE,n.obs=n.obs)$result.df
        uni.temp <- unidim(keys,R, n.obs=n.obs)$fa.stats
        results[[i]] <- temp
uni.results[[i]] <- uni.temp
        }  #end of loop


#organize the results
result.df <- matrix(unlist(results),byrow=TRUE,ncol=13*nvar)
uni.df <- matrix(unlist(uni.results), byrow=TRUE, ncol=9*nvar)
return(result.df)
}  # the end of our short simulation function


#now use this function to generate the replications of the data
set.seed(42)         #allows for reproducible results

 sim100 <- simulate.unidim(fx96,n.obs=100,n.replications=100,keys=keys)   # n = 100
 sim200 <- simulate.unidim(fx96,n.obs=200,n.replications=100,keys=keys)
  sim500 <- simulate.unidim(fx96,n.obs=500,n.replications=100, keys=keys)
  sim1k <- simulate.unidim(fx96,n.obs=1000,n.replications=100, keys=keys)
   sim5k <- simulate.unidim(fx96,n.obs=5000,n.replications=100, keys=keys)


 #now combine the u and omega_t estimates across data sets to graph them

 total.sims <- data.frame(N=c(rep(200,100), rep(500,100), rep(1000,100),rep(5000,100)),
      rbind(sim200[,41:80], sim500[,41:80], sim1k[,41:80], sim5k[,41:80]),
     rbind(sim200[,221:260], sim500[,221:260] , sim1k[,221:260], sim5k[,221:260]))

 #clean up the names for the graphics
temp <- gsub("omega.tot","",colnames(total.sims))
 temp <- gsub("Uni","u",temp)
  temp[18:21] <- c("F1+F2 6","F1+F2 10","F1+F2 12","F1+F2 24")
temp[38:41] <- c("F1+F2 6","F1+F2 10","F1+F2 12","F1+F2 24")
colnames(total.sims) <- temp

#n = 100

 error.bars.by(total.sims[,22:41],total.sims[,1],by.var=FALSE,v.labels=colnames(total.sims[2:21]),
      ylab="U",xlab="", las=3,main="B) U varies by sample size and factor loadings",legend=3,
      labels =c(200,500,1000,5000), lty=1:4,col =cs(black, blue,red, green))


error.bars.by(total.sims[,2:21],total.sims[,1],by.var=FALSE,v.labels=colnames(total.sims[2:21]),
    ylab=expression(omega[t]), las=3,xlab="",ylim=c(0,1),main=expression(paste("A)  ",
       omega[t]," varies by sample size and factor loadings")),
       legend=3,     labels =c(100,500,1000,5000), lty=1:4,col =cs(black, blue,red, green))

error.bars.by(total.sims2[,42:61],total.sims2[,1],by.var=FALSE, v.labels=names(keys),
    ylab="CFI", las=3,xlab="", ylim=c(0,1),
    main=" C) CFI is not very sensitive to variations in factor loadings",
    labels =c(200,500,1000,5000), lty=1:4,col =cs(black, blue,red, green))

 error.bars.by(total.sims2[,62:81],total.sims2[,1],by.var=FALSE,v.labels=names(keys),
    ylab="ECV",  las=3,xlab="",ylim=c(0,1),
    main=" D) ECV varies by sample size and factor loadings",
      labels =c(200,500,1000,5000), lty=1:4,col =cs(black, blue,red, green))


   #min n =200

     total.sims <- data.frame(N=c(rep(100,100), rep(500,100), rep(1000,100),rep(5000,100)),
      rbind(sim200[,41:80], sim500[,41:80], sim1k[,41:80], sim5k[,41:80]),
     rbind(sim200[,221:260], sim500[,221:260] , sim1k[,221:260], sim5k[,221:260]))

 error.bars.by(total.sims[,2:21],total.sims2[,1],by.var=FALSE,v.labels=names(keys),
    ylab=expression(omega[t]), las=3,xlab="",ylim=c(0,1),main=expression(paste("A)  ",
       omega[t]," varies by sample size and factor loadings"))

 #      legend=3,     labels =c(200,500,1000,5000), lty=1:4,col =cs(black, blue,red, green))

      error.bars.by(total.sims[,22:41],total.sims2[,1],by.var=FALSE,v.labels=names(keys),
      ylab="U",xlab="", las=3,main="B) U varies by sample size and factor loadings",
      labels =c(5000,1000,500,200), lty=1:4,col =cs(green, red, blue, black))


  #do legends separately to have more control
```

```
 error.bars.by(total.sims[,22:41],total.sims2[,1],by.var=FALSE,v.labels=names(keys),
      ylab="U",xlab="", las=3,main="B) U varies by sample size and factor loadings",
      labels =c(5000,1000,500,200), lty=1:4,col =cs(black, blue,red, green))
legend("bottomleft",lty=4:1,legend=c(5000,1000,500,200),col=cs(green, red, blue, black) ,pch=15)


#score
#

bfi.keys.plus <-
  list(extraversion=c("-E1","-E2","E3","E4","E5"),
  openness = c("O1","-O2","O3","O4","-O5"),
  agree=c("-A1","A2","A3","A4","A5"),
  conscientious=c("C1","C2","C3","-C4","-C5"),
neuroticism=c("N1","N2","N3","N4","N5"),

EO = c("-E1","-E2","E3","E4","E5","O1","-O2","O3","O4","-O5"),
ACN =c("-A1","A2","A3","A4","A5","C1","C2","C3","-C4","-C5","N1","N2","N3","N4","N5"),
All = c("-E1","-E2","E3","E4","E5","O1","-O2","O3","O4","-O5","-A1","A2","A3","A4","A5",'
      "C1","C2","C3","-C4","-C5","N1","N2","N3","N4","N5"))
bfi.rel <- reliability(bfi.keys.plus,bfi,n.sample=5200300)

}
```