Estimating Generalizability to a Latent Variable Common to All of a Scale's Indicators: A Comparison of Estimators for ω_h

Richard E. Zinbarg, Northwestern University
   and The Family Institute at Northwestern University
Iftah Yovel, Northwestern University
William Revelle, Northwestern University
Roderick P. McDonald, University of Illinois

The extent to which a scale score generalizes to a latent variable common to all of the scale's indicators is indexed by the scale's general factor saturation. Seven techniques for estimating this parameter—omega$_{hierarchical}$ (ω_h)—are compared in a series of simulated data sets. Primary comparisons were based on 160 artificial data sets simulating perfectly simple and symmetric structures that contained four group factors, and an additional 200 artificial data sets confirmed large standard deviations for two methods in these simulations when a general factor was absent. Major findings were replicated in a series of 40 additional artificial data sets based on the structure of a real scale widely believed to contain three group factors of unequal size and less than perfectly simple structure. The results suggest that alpha and methods based on either the first unrotated principal factor or component should be rejected as estimates of ω_h. *Index terms: generalizability, alpha, omega, factor analysis, measurement, reliability.*

Many scales are assumed by their developers and users to be primarily a measure of one latent variable. When it is also assumed that the scale conforms to the effect indicator model of measurement (as is almost always the case in psychological assessment), it is important to support such an interpretation with evidence regarding the internal structure of that scale (Bollen & Lennox, 1991). In particular, it is important to examine two related properties pertaining to the internal structure of such a scale. The first property relates to whether all the indicators forming the scale measure a latent variable in common.

The second internal structural property pertains to the proportion of variance in the scale scores (derived from summing or averaging the indicators) accounted for by this latent variable that is common to all the indicators (Cronbach, 1951; McDonald, 1999; Revelle, 1979). That is, if an effect indicator scale is primarily a measure of one latent variable common to all the indicators forming the scale, then that latent variable should account for the majority of the variance in the scale scores. Put differently, this variance ratio provides important information about the sampling fluctuations when estimating individuals' standing on a latent variable common to all the indicators arising from the sampling of indicators (i.e., when dealing with either Type 2 or

Type 12 sampling, to use the terminology of Lord, 1956). That is, this variance proportion can be interpreted as the square of the correlation between the scale score and the latent variable common to all the indicators in the infinite universe of indicators of which the scale indicators are a subset. Put yet another way, this variance ratio is important both as reliability and a validity coefficient. This is a reliability issue as the larger this variance ratio is, the more accurately one can predict an individual's relative standing on the latent variable common to all the scale's indicators based on his or her observed scale score. At the same time, this variance ratio also bears on the construct validity of the scale given that construct validity encompasses the internal structure of a scale.

An index corresponding to this variance ratio can be defined in a model labeled by McDonald (1999) as the hierarchical factor model, in which it is assumed that observed scores may be decomposed into four parts:

$$x = \mathbf{G}g + \mathbf{AF} + \mathbf{DS} + \mathbf{E}, \tag{1}$$

where g is a general factor (i.e., a factor common to all k scale indicators), $\mathbf{G}$ is a $k \times 1$ vector of unstandardized general factor loadings (with each unstandardized loading being equal to the item's standardized loading times its *SD*), $\mathbf{F}$ is an $r \times 1$ vector of group factors (i.e., factors that are common to some but not all k indicators) with $r < k$, $\mathbf{A}$ is the $k \times r$ matrix of unstandardized group factor loadings, $\mathbf{S}$ is the $k \times 1$ vector of specific factors unique to each item, $\mathbf{D}$ is the $k \times k$ diagonal matrix of unstandardized factor loadings on the item specific factors, and $\mathbf{E}$ is the $k \times 1$ vector of random error scores as above. The model expressed in (1) assumes also that all factors (g, $\mathbf{F}$, and $\mathbf{S}$) are uncorrelated with each other, and with $\mathbf{E}$, all the errors (in $\mathbf{E}$) are uncorrelated with each other, without loss of generality that the variance of each of the common factors (g and those in $\mathbf{F}$) equals 1. In addition, each of the group factors (in $\mathbf{F}$) is required to have nonzero loadings on at least three indicators to identify the model in the absence of additional constraints.

Given the model in (1), a coefficient relating to an important psychometric property—omega$_{\text{hierarchical}}$ (ω_h)—is defined as

$$\omega_h = \frac{1'\mathbf{GG}'1}{1'\mathbf{S}1}, \tag{2}$$

where $\mathbf{S}$ is the observed variance/covariance matrix among the k indicators comprising the scale, and ω_h is the proportion of variance in the scale scores accounted for by a general factor. Of course, (2) still holds even in the special case when there are no group factors in which (1) reduces to $x = \mathbf{G}g + \mathbf{DS} + \mathbf{E}$.[1] ω_h can be interpreted as the square of the correlation between the scale score and the latent variable common to all the indicators in the infinite universe of indicators of which the scale indicators are a subset (McDonald, 1999). Thus, the larger ω_h is, the more strongly one's scale scores are influenced by a latent variable common to all the indicators, and the more the observed scale scores can be said to generalize to scores on that latent variable.

The first psychometric property described above, that there is a latent variable common to all of the scale's indicators, is clearly a requisite assumption for the definition of ω_h. Whenever possible, this assumption should be formally tested via the appropriate use of confirmatory factor analysis (CFA) or via Stout's methods for evaluating whether a scale is essentially unidimensional. The primary focus in this article, however, is on methods for estimating ω_h rather than on methods for conducting formal tests of this assumption. The reasons for this are fourfold. First, there are no published studies comparing the several different methods that one might use to estimate ω_h. In contrast, there are already many excellent discussions of the use of CFA for testing the assumptions underlying one's measurement model; for some examples, see Bollen (1989), Drewes

(2000), Jöreskog (1971, 1974), McDonald (1981, 1999), Raykov, (1997), and Reuterberg and Gustafsson (1992). (For examples of CFA tests of the assumption that all of a scale's indicators measure a latent variable in common and of the use and interpretation of ω_h with real data sets, see Mohlman & Zinbarg, 2000; Zinbarg, Barlow, & Brown, 1997).

Second, there are likely to be times in practice when the model testing approach fails. For example, failure to include appropriate comparison models could lead an investigator to accept a measurement model containing a general factor when in fact there is no general factor. In such cases, it would be important to know whether an ω_h estimate based on this erroneously accepted model is likely to produce a value reasonably close to zero.

Third, there will be situations in which scale users might not have strong expectations regarding the number of group factors to include in the measurement model and/or which indicators load on each of the group factors. In such cases, Fabrigar, Wegener, MacCallum, and Strahan's (1999) suggestion that exploratory factor analysis (EFA) is a more sensible approach than CFA seems reasonable, and EFA does not permit formal testing of the assumption of a latent variable common to all of a scale's indicators. Despite the inability to conduct a formal test of this assumption in such cases, estimating ω_h can still provide useful information regarding the justifiability of aggregating across each member of a set of indicators to derive a scale score.

Finally, either a CFA test of the assumption that there is a latent variable common to all of the scale's indicators or a test of essential unidimensionality results in a binary decision: Either there is a general factor present or not; either the scale is essentially unidimensional or not. In contrast, ω_h provides a quantitative estimate indexing how well the scale measures its general factor. A scale can be essentially unidimensional but still not measure its dominant factor very well if, for example, it contains few indicators that are each only weakly saturated with the underlying common factor. Conversely, a scale can be clearly multidimensional but still have a single factor that is common to all the scale's indicators and that accounts for a large proportion of the variance in scale scores. That is, whereas essential unidimensionality is an important and interesting concept in its own right, it can be seen that essential unidimensionality is neither necessary nor sufficient for ω_h to be high. Put differently, ω_h provides important quantitative information about the reliability and validity of a scale that cannot be obtained from either a CFA inferential hypothesis testing procedure or an inferential test of essential unidimensionality.

Thus, the primary purpose of this article is to address the gap in the literature regarding the relative performance of different methods of estimating ω_h. A discussion of the relationship between ω_h and McDonald's (1985, 1999) omegas and the importance of ω_h is presented, followed by a discussion of several methods of estimating ω_h. Finally, a Monte Carlo study of the performance of these different methods of estimating ω_h is reported.

Relationship With McDonald's Omegas and the Importance of ω_h

Equation (2) is algebraically equivalent to one of the expressions that McDonald (1985, 1999) has labeled omega (ω). Specifically, if λ_{gj} is the loading of indicator j on the general factor, then equation (2) is equivalent to McDonald's (1985, p. 217) equation (7.3.8) or to McDonald's (1999, p. 89) equation (6.20a), which—using our notation of ω_h—reads

$$\omega_h = \frac{(\sum_{i=1}^{k} \lambda_{gj})^2}{1'S1}.$$ (3)

Of course, (3) is simpler and more efficient to use than (2) when calculating ω_h by hand.

However, McDonald (1985, equation (7.3.9); 1999, equation (6.2.1)) also identified another expression and used the same label of ω for it and (3). McDonald's (1985, p. 217) equation (7.3.9) or McDonald's (1999, p. 89) equation (6.2.1) reads as follows:

$$\omega = 1 - \frac{\sum_{i=1}^{k} u_i^2}{1'\mathbf{S}1}, \tag{4}$$

where the $d_i s_i$ and e_i components from (1) combine to form u_i—the item "uniqueness"

$$u_i = d_i s_i + e_i. \tag{5}$$

Equation (4) is algebraically equivalent to

$$\omega = \frac{1'\mathbf{G}\mathbf{G}'1 + 1'\mathbf{A}\mathbf{A}'1}{1'\mathbf{S}1}. \tag{6}$$

It is readily apparent from (2) and (6) that ω and ω_h are different parameters, with $\omega \geq \omega_h$ and equality holding in the case of a unidimensional scale only (McDonald, 1985, 1999; Zinbarg, Revelle, Yoveld, & Li, 2005). In differentiating the labels for these two, Zinbarg et al. (2005) retained McDonald's (1985, 1999) label of ω for (4) and (6) but gave (2) and (3) the new label of ω_h. The primary consideration underlying this decision was that the parameter expressed by (2)—or, equivalently, by (3)—is specific to the hierarchical factor model, whereas (4) may be defined in either the hierarchical factor model or what McDonald (1999) labels the independent clusters factor model. In addition, other references to McDonald's ω have tended to focus on the parameter equivalent to (4) or (6).

Before discussing the different methods for estimating ω_h, it might be useful to briefly discuss why ω_h is important. One might wonder, for example, if one knows that ω or a conventional reliability estimate is high for a particular scale, why should one care about the value of ω_h for that scale? If ω_h were low for such a scale, it might indicate that (a) the universe from which one's scale indicators are sampled is multifaceted, and (b) the scale scores are a result of the largely independent contributions of what is unique to several of these facets, without much of a contribution from a latent construct that is common to all the facets. If so, it would be unclear (a) which facet is accounting for a given relationship between the scale scores and some other variable and (b) which facets are elevated among individuals who obtain moderately high scale scores. In such cases, it would be more appropriate for scale users to derive two or more subscale scores from the instrument rather than a single scale score to clarify theoretical understanding of relations with other variables and to avoid equating individuals who have very different profiles across the multiple facets (for related discussions, see Kelley, 1942; McNemar, 1946; Mosier, 1936). The other possible implication of a low value of ω_h when ω or a conventional reliability estimate is high for a particular scale is that there are several indicators that should be eliminated because they do not load on the dominant factor that is loaded on by most of the indicators.

Methods for Estimating ω_h

In practice, of course, one never knows the loadings of a scale's indicators on a factor but can only estimate these loadings and thus can only estimate ω_h. This study chose to focus on those methods that appear to be most commonly used by applied researchers for extracting a general factor and estimating its loadings. Finally, it is suspected that coefficient alpha is frequently used by researchers as an estimate of ω_h; therefore, alpha was also included in the comparisons.

Does alpha estimate ω_h? In fact, McDonald (1970, 1985, 1999) has proven that coefficient alpha is a special case of ω_h when the indicators conform to the classical test-theory assumptions underlying alpha—that is, when the indicators are unidimensional and essentially tau-equivalent and that measurement errors are uncorrelated. However, alpha should provide biased estimates of ω_h when these assumptions are violated, with the direction of the bias depending on the assumption or assumptions violated (Zinbarg et al., in press; for related discussions, see Komaroff, 1997; Zimmerman, Zumbo, & Lalonde, 1993). On one hand, alpha will tend to underestimate ω_h for a unidimensional set of indicators that are not essentially tau-equivalent (e.g., McDonald, 1970, 1985, 1999; Raykov, 1997). On the other hand, several authors have noted that alpha tends to overestimate the proportion of variance due to a general factor when the indicators are multidimensional (e.g., Cronbach, 1951; Revelle, 1979; Schmitt, 1996).

In fact, many scales used in psychological and educational measurement are multidimensional. As one example, consider the various IQ scales that have been hypothesized to consist of various primary mental abilities at the first-order level, higher order group factors, such as the well-known Verbal and Performance factors at the second-order level and a general factor at the highest order level (e.g., Vernon, 1969). Many other examples of hierarchical/higher order structures could be given from other domains, including personality scales (e.g., Costa & McCrae, 1985; Waller, Lilienfeld, Tellegen, & Lykken, 1991; Zinbarg, Mohlman, & Hong, 1999), affect scales (e.g., Tellegen, Watson, & Clark, 1999; Watson & Clark, 1992), and psychopathology scales (e.g., Beck & Steer, 1993a, 1993b; Foa, Riggs, Dancu, & Olasov-Rothbaum, 1993; Zinbarg & Barlow, 1996). Given that (a) multidimensional scales are common and (b) the problem of alpha overestimating ω_h for a multidimensional scale has received less theoretical and empirical attention than the problem of alpha underestimating ω_h for unidimensional indicators that are not essentially tau-equivalent, the focus in the remainder of this article is on the multidimensional case in which alpha should overestimate ω_h.

Although the discussion of alpha primarily focuses on the tendency of alpha to overestimate ω_h for multidimensional scales, there is another problem associated with alpha in the multidimensional case. In the words of Osburn (2000), there is a tendency for "coefficient alpha to underestimate the true reliability when the data are multidimensional" (p. 343), where by true reliability, Osburn must have been referring to the proportion of variance due to all common factors. In fact, there is a large literature on alpha's tendency to underestimate the proportion of variance due to all common factors for a multidimensional scale, and several indices to overcome this limitation, such as stratified alpha, maximal reliability, or Guttman's (1945) Lambda 2, have been studied fairly extensively (e.g., Cronbach, Schoneman, & McKie, 1965; Feldt & Brennan, 1989; Kristof, 1974; Li, Rosenthal, & Rubin, 1996; Osburn, 2000). Moreover, the problem of alpha underestimating the proportion of variance due to all common factors is important as it will lead to overcorrection when alpha is used to correct for attenuation. This problem has been much more widely studied, however, than that of alpha overestimating ω_h for a multidimensional scale.

There are additional complexities associated with traditional reliability estimates that are illuminated by generalizability theory and should be considered in relation to ω_h. The generalizability theory distinction between absolute error and relative error and analysis of the effects of sampling scheme on error variance clearly show that one's research purpose (i.e., whether one is interested in making "absolute" vs. "relative" interpretations of test scores) and design can affect error variances. In this regard, it is important to note that ω_h is a generalizability coefficient that employs relative error variance rather than absolute error variance. Sampling scheme, however, will typically not have an effect on ω_h.

Factor-analytic methods. Four methods derived from factor analysis are considered and compared in the remainder of this article: (a) perform an EFA using the principal factor method of extraction (PF) and use the loadings on the first unrotated PF as estimates of the general factor

loadings (first PF method) (see Goldberg, 1999, for an example of the application of this technique for the purpose of estimating general factor loadings); (b) perform a higher order EFA analysis using the PF method of extraction (i.e., extract two or more obliquely transformed first-order principal factors and then use PF to extract higher order factors based on the correlations among the obliquely transformed first-order principal factors), relate the indicators directly to the highest order PF (using the transformation originally described by Schmid & Leiman, 1957; also see Loehlin, 1998, pp. 225-228), and use the loadings on the highest order PF as estimates of the general factor loadings (HO-PF method); (c) perform a CFA of a higher order factor model and relate the indicators directly to the highest order factor to estimate the loadings of the indicators on the general factor (HO-CF method); and (d) perform a CFA of a hierarchical factor model similar in form to equation (1), in which each indicator loads directly on a general factor in addition to loading on at least one group factor (Hi-CF method).

Principal components analysis (PCA) methods. PCA is another procedure that some researchers use to represent the structure of the correlations among a set of variables. Although the theoretical model underlying PCA is inconsistent with the factor-analytic model underlying ω_h (e.g., Bentler & Kano, 1990; Gorsuch, 1983, 1990; McArdle, 1990; Mulaik, 1990; Snook & Gorsuch, 1989; Widaman, 1990, 1993), a series of investigations by Velicer (1974, 1976, 1977) and his associates (Velicer & Fava, 1987; Velicer & Jackson, 1990a, 1990b; Velicer, Peacock, & Jackson, 1982) has shown that PCA and EFA produce highly similar solutions under many conditions. In addition, although Snook and Gorsuch (1989) and Widaman (1993) have identified conditions in which there are reliable differences between the two methods, Widaman found that PCA produces larger estimates of loadings on primary dimensions and smaller estimates of intercorrelations among the primary dimensions than EFA under these conditions. These two biases may tend to offset each other when computing loadings on second-order dimensions (especially if one follows the recommendation of Gorsuch, 1983, pp. 243–244, and uses PF for the second-order factor analysis of the correlations among the obliquely transformed first-order principal components). That is, even in the conditions in which Widaman found reliable differences between the lower order estimates produced by PCA and EFA, the two methods may produce essentially equivalent estimates of second-order loadings. In addition, PCA is the default option in many of the major statistical program packages and has been reported in the past to be more widely used than EFA (Pruzek & Rabinowitz, 1981; also see Lilly, Hoaglin, & Anderson-Kulman, 1989, cited in Velicer & Jackson, 1990a). Thus, it is important to examine ω_h estimates based on PCA to see if ω_h gives reasonable results, even when it is estimated in a manner that conflicts with the theoretical model that permitted its derivation in the first place.

Two methods derived from PCA analysis were therefore also considered: (a) perform a PCA and use the loadings on the first unrotated PC as estimates of the general factor loadings (first PC) (again see Goldberg, 1999, for an example of the application of this technique for the purpose of estimating general factor loadings), and (b) perform a higher order PCA (i.e., extract two or more obliquely transformed first-order principal components and then, following the recommendation of Gorsuch, 1983, use PF to extract higher order factors based on the correlations among the obliquely transformed first-order principal components), relate the indicators directly to the highest order dimension, and use the loadings on the highest order dimension as estimates of the general factor loadings (HO-PC).

Study Questions

This study compared the usefulness of (a) the two methods for estimating ω_h derived from EFA (first PF and HO-PF), (b) the two methods derived from CFA (HO-CF and Hi-CF), (c) the two

methods derived from PCA (first PC and HO-PC), and (d) alpha. These were first applied to artificial data sets that simulated perfectly simple and symmetric structures containing four group factors but that varied in terms of the presence and strength of a general factor (the intended meaning of perfectly simple and symmetric in this context is that each item had loadings of zero on all but one group factor, and the group factors were exactly equal in size). In the first general factor loading pattern condition, the general factor loadings were set equal to 0 to simulate conditions in which a general factor is absent. In the second general factor loading pattern condition, the general factor loadings were set equal to .274 to simulate a relatively weak general factor. In the third general factor loading pattern condition, the general factor loadings were set equal to .274 for the indicators defining two of the group factors and .500 for the indicators defining the remaining two group factors to simulate a moderately strong general factor with some degree of heterogeneity in the general factor loadings. In the fourth general factor loading pattern condition, the general factor loadings were set equal to .500 to simulate a relatively strong general factor. A second independent variable that was systematically varied was the sample size, n. This study chose two levels of n: 100 and 200. The third and final independent variable was the number of indicators or scale length, k. Two levels of k were chosen: 12 indicators (3 indicators per group factor) and 20 indicators (5 indicators per group factor). This variable was included to study the effects of scale length on the various ω_h estimates because the bias inherent in alpha tends to increase with scale length, whereas it has been noted that differences between PF and PC analyses tend to diminish with increased scale length (e.g., Gorsuch, 1983, Nunnally, 1978; Snook & Gorsuch, 1989; Widaman, 1993).

Certain findings to follow are reasonably well documented in the literature; for example, the first unrotated factor or component often indicates the presence of a strong general dimension, even when none exists (Jensen & Weng, 1994; Nunnally, 1978). Thus, a relatively brief amount of space is devoted to findings that are additional demonstrations of such effects. Of main interest is an examination of whether the HO-PF, HO-CF, HO-PC, and Hi-CF methods of estimating ω_h are more accurate than alpha and the first PF and first PC methods. A good estimator should be (a) unbiased and (b) relatively efficient, and (c) its accuracy should be stable across different general factor loading patterns, sample sizes, and scale lengths.

To begin to test the generalizability of the results from the initial series of 160 artificial data sets, 40 additional artificial data sets were generated based on the structure of widely used scales in the measurement of anxiety—the Anxiety Sensitivity Index (ASI; Reiss, Peterson, Gursky, & McNally, 1986). The structure of the ASI is useful for testing the generalizability of the major findings from the initial series of 160 simulated data sets as it is thought to differ in several ways from the perfectly simple and symmetric four-group factor structures simulated in the initial series of 160 artificial data sets. That is, the ASI has been consistently found to contain three group factors, with one group factor much larger than the other two and with several indicators having group factor cross-loadings (e.g., Mohlman & Zinbarg, 2000; Zinbarg et al., 1997; Zinbarg et al., 1999). Thus, of the 16 indicators comprising the ASI, 8 have their primary loadings on the first group factor, and the remaining 8 indicators are evenly divided between the second and third group factors. In addition, using 0.15 as a cutoff for cross-loadings, Zinbarg et al. (1997) found that 4 ASI indicators have a cross-loading on a second group factor, in addition to their primary group factor loading and one item loaded on all three group factors.

Method

For all analyses, artificial data sets were generated using EQS (Bentler, 1989). In the initial series of simulations, 10 artificial data sets were created within each of the 16 conditions resulting

from the 4 (type of higher order structure: higher order factor absent, weak higher order factor loadings, mixed higher order factor loadings, and strong higher order factor loadings) $\times$ $2(n : 100 \text{ vs. } 200) \times 2(k, 12 \text{ vs. } 20)$ design. In each simulation, the indicators were divided into four equal-sized sets of indicators (thus there were three indicators per set in the $k = 12$ condition and five indicators per set in the $k = 20$ condition).

The higher order factor model representations of the models underlying these simulations were as follows. In all of the models, every indicator within each of the four sets of indicators had a loading of .707 on that set's group factor. For the 40 simulations that did not contain a general factor, all four group factors had loadings of 0 on the second-order factor. For the 40 simulations in the weak general factor loadings condition, the four group factors had loadings of .387 on the second-order factor. For the 40 simulations in the mixed general factor loadings condition, the first two group factors had loadings of .387 on the second-order factor, whereas the last two group factors had second-order factor loadings of .707. For the 40 simulations in the strong general factor loadings condition, the four group factors had loadings of .707 on the second-order factor. These models produced the following population correlation matrices. In each condition, indicators loading on the same group factor correlated .50. In the general factor absent condition, indicators loading on different group factors correlated 0. In the weak general factor loading condition, indicators loading on different group factors correlated .075. In the mixed general factor loading condition, indicators loading on Group Factor 1 correlated .075 with indicators loading on Group Factor 2, indicators loading on Group Factors 1 and 2 correlated .137 with indicators loading on Group Factors 3 and 4, and indicators loading on Group Factor 3 correlated .25 with indicators loading on Group Factor 4. Finally, in the strong general factor loadings condition, indicators loading on different group factors correlated .25.

The factors and errors and hence measured variables in each simulation were multivariate normally distributed. Given that many measurement scales are composed of indicators that yield nonnormal and/or categorical data, the choice to model multivariate normally distributed indicators limits the generality of the results reported here and therefore deserves comment. The reasoning behind this choice was prosaic yet may be sensible nevertheless. First, the methods employed here often produce results when applied to nonnormal and/or categorical data that converge very strongly with the results of methods that were explicitly developed for nonnormal and/or categorical data (e.g., factor loadings are often correlated .90 or higher with item slope estimates from IRT analyses). Second, it could be argued that a reasonable strategy for an initial test of different methods for estimating ω_h is to conduct this test using conditions that maximize the likelihood that at least some of the methods would perform well. That is, before undertaking the more difficult task of testing these methods in simulations of categorical variables, one might want to know if there are any conditions in which ω_h could be estimated in a relatively unbiased and efficient manner. Finally, ω_h could be easily adapted for use with the nonlinear factor analysis model developed by McDonald (1963, 1965, 1999) explicitly for use with categorical data. (A detailed description of checks on the adequacy of the data generation procedure and more details regarding the simulation methods are available upon request from Richard E. Zinbarg.) Coefficient alpha was computed for each artificial data set using the SPSS (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975) reliability procedure. PCAs and EFAs were conducted using the SPSS (Nie et al., 1975) factor procedure with oblimin rotation for the higher order solutions and iterative estimates of communality using the squared multiple correlation (SMC) as the initial estimate for PF extraction. Each higher order EFA was conducted by extracting four first-order factors, rotating them using oblimin (with delta set to 0), and then subjecting the intercorrelation matrix of the first-order factors to a factor analysis and extracting a single second-order factor (note that one could use other methods of oblique rotation other than oblimin but could not use an orthogonal method such as varimax, which is the default option for many computer packages). CFAs were conducted using EQS (Bentler, 1989)

and maximum likelihood estimation. Each higher order CFA fit a model in which there were four group factors that were each loaded on by one fourth of the indicators, and each of the group factors loaded on a single second-order factor. Each hierarchical CFA fit a model consisting of one general factor, which was loaded on by all indicators, and four group factors, which each were loaded on by one fourth of the indicators. It should be noted that the dimensionality of the higher order EFA, PCA, and CFA models equaled four, whereas the dimensionality of the hierarchical CFAs equaled five. This is because in the higher order models, the higher order dimension lies entirely in the space defined by the primary four dimensions, whereas in the hierarchical CFA models, the general factor does not lie in the space defined by the group factors given that all five factors were constrained to be mutually orthogonal. ω_h estimates from all seven methods (first PF, first PC, HO-PF, HO-PC, HO-CF, Hi-CF, and coefficient alpha) for each sample were compared to the population ω_h value by subtracting the population value from each of the estimates. The resulting difference scores provided a direct measure of the bias of the ω_h estimates and alpha as compared to the population ω_h values. This yielded seven sets of difference scores, which were analyzed by a repeated-measures multivariate analysis of variance (MANOVA) in which the type of higher order structure was a between-groups factor with four levels, number of indicators was a between-groups factor with two levels, and sample size was a between-groups factor with two levels. The repeated-measure factor of estimate type had seven levels and was decomposed into a set of five planned contrasts. The first contrast was designed to test the hypotheses of primary interest—namely, that the HO-PAF, HO-PC, HO-CF, and Hi-CF methods (the putatively accurate methods) are better than alpha and the first PF and first PC methods (the putatively inaccurate methods). Thus, Contrast 1 contrasts the four putatively accurate methods with the three putatively inaccurate methods. Contrast 2 tests for differences among the putatively inaccurate methods; specifically, it contrasts alpha with the first PF and first PC estimators. Contrasts 3, 4, and 5 were designed to test for differences among the four putatively accurate methods. Contrast 3 contrasts the two exploratory methods (HO-PF and HO-PC) with the two CFA methods (HO-CF and Hi-CF). Contrast 4 contrasts the HO-PF method with the HO-PC method. Contrast 5 contrasts the HO-CF method with the Hi-CF method. Based on the findings of Jensen and Weng (1994), it was predicted that any differences involving Contrasts 3, 4, and 5 would be trivially small.

Results

Table 1 presents the correlations among the population values of ω_h and the seven estimators of ω_h across all 160 simulations. Table 2 presents the population values of ω_h and the mean difference scores for the seven methods of estimating ω_h across the four types of structures and two levels of the number of indicators simulated in this study. The standard errors were very small for almost all of the comparisons owing to the repeated-measures design and the very high correlations among most of the estimators. Therefore, it is important to focus heavily on effect size estimates in addition to statistical significance. The results were collapsed across the two different levels of sample size as almost all effects involving sample size were nonsignificant, all were trivially small (η^2 values $\leq .006$), and the standard deviations were not appreciably smaller for the larger sample sizes for all but one of the seven estimators (the ratio of the standard deviation in the larger sample size condition divided by the standard deviation in the smaller sample size condition equaled 1.001, 1.005, 1.050, 0.935, 0.911, 0.935, and 0.786, respectively, for alpha and the first PF, first PC, HO-PF, HO-PC, HO-CF, and Hi-CF methods).

Contrast 1: The four putatively accurate methods versus the three putatively inaccurate methods. The main effect of Contrast 1 was significant ($F_{1,144} = 1680.12$, $p \leq .001$; $\eta^2 = .65$), indicating that

Table 1

Intercorrelations Among Population Values and the Seven Methods of Estimating ω_h

Estimator	ω_h	α	First PF	First PC	HO-PF	HO-PC	HO-CF
α	.84						
First PF	.82	.83					
First PC	.78	.77	.99				
HO-PF	.95	.88	.85	.81			
HO-PC	.95	.88	.85	.81	.99		
HO-CF	.94	.86	.80	.75	.97	.97	
Hi-CF	.93	.83	.75	.70	.96	.96	.95

Note. ω_h = population value of ω_h, α = coefficient alpha; first PF = ω_h estimate based on the first unrotated principal factor; first PC = ω_h estimate based on the first unrotated principal component; HO-PF = ω_h estimate based on a higher order exploratory factor analysis; HO-PC = ω_h estimate based on a higher order principal components analysis; HO-CF = ω_h estimate based on a higher order confirmatory factor analysis; Hi-CF = ω_h estimate based on a hierarchical confirmatory factor analysis.

the average of the four putatively accurate methods was more accurate (less positively biased) than the average of the three putatively inaccurate methods when aggregated across all conditions ($M = .023, SD = .086$ vs. $M = .356, SD = .153$). Importantly, although this main effect was moderated by two interactions discussed in more detail below, the average of the four putatively accurate methods was less positively biased than the average of the putatively inaccurate methods in every condition in these simulations.

The two-way interactions of Contrast 1 × the type of general factor loading pattern ($F_{3,144} = 44.11$, $p \leq .001, \eta^2 = .05$) and Contrast 1 × the number of indicators ($F_{1,144} = 4.99$, $p \leq .05; \eta^2 = .002$) were significant. The significant interaction of Contrast 1 × the type of general factor loading pattern resulted from the fact that the performance of the four putatively accurate methods varied less across the different general factor loading patterns (general factor absent condition: $M = .10$, weak higher order loadings: $M = -.02$, mixed higher order loadings: $M = .04$, and strong higher order loadings: $M = -.04$) than was the case for the three putatively inaccurate methods. That is, whereas the three putatively inaccurate methods were positively biased regardless of the type of general factor loading pattern, they were not as positively biased in the strong higher order loadings condition as in the other three general factor loading conditions (general factor absent condition: $M = .38$, weak higher order loadings: $M = .46$, mixed higher order loadings conditions: $M = .38$, and strong higher order loadings condition: $M = .20$). The two-way interaction of Contrast 1 × the number of indicators was associated with such a trivially small effect size that its practical significance is highly questionable and will not be discussed further. No other effects involving Contrast 1 were significant.

To summarize the results for Contrast 1, compared with the three putatively inaccurate methods of estimating ω_h, the four putatively accurate methods were more accurate (less positively biased). The four putatively accurate methods were also more stable across general factor loading patterns.

Contrast 2: First PF and first PC versus alpha. The interaction of Contrast 2 by the type of general factor loading pattern was significant ($F_{3,144} = 109.46, p \leq .001, \eta^2 = .28$), as was the interaction of Contrast 2 × the number of indicators ($F_{3,144} = 6.64, p \leq .05, \eta^2 = .01$). The significant interaction of Contrast 2 × the type of general factor loading pattern resulted from alpha being significantly more positively biased in the general factor absent condition (first PF and first PC: $M = .27$; alpha: $M = .61$) but significantly less positively biased in the weak higher order loadings (first PF and first PC: $M = .50$; alpha: $M = .38$), mixed higher order loadings (first PF and

Table 2

Mean Differences Between Estimated and Population Values of ω_h: Four Group Factor Simulations

k	ω_h	α	First PF	First PC	HO-PF	HO-PC	HO-CF	Hi-CF
				$\lambda_G = 0$				
12	.00	.534	.191	.245	.079	.078	.156	.069
		(.057)	(.232)	(.284)	(.066)	(.065)	(.114)	(.070)
20	.00	.691	.297	.335	.094	.095	.170	.089
		(.056)	(.286)	(.328)	(.082)	(.082)	(.106)	(.071)
				Weak λ_G				
12	.34	.336	.406	.568	−.035	−.036	.019	−.067
		(.036)	(.157)	(.189)	(.074)	(.072)	(.079)	(.107)
20	.36	.432	.464	.574	−.012	−.015	.044	−.055
		(.026)	(.094)	(.102)	(.076)	(.077)	(.062)	(.133)
				Mixed λ_G				
12	.43	.333	.410	.546	.076	.073	.129	.127
		(.030)	(.027)	(.020)	(.057)	(.057)	(.057)	(.061)
20	.58	.272	.317	.409	−.048	−.043	.026	.004
		(.012)	(.012)	(.010)	(.039)	(.041)	(.031)	(.035)
				Strong λ_G				
12	.71	.123	.176	.291	−.071	−.082	−.006	−.012
		(.025)	(.019)	(.002)	(.054)	(.055)	(.050)	(.049)
20	.74	.159	.188	.257	−.059	−.066	.008	.001
		(.014)	(.009)	(.002)	(.049)	(.049)	(.039)	(.052)
		.360	.306	.403	.003	.001	.068	.020
		(.181)	(.178)	(.214)	(.088)	(.091)	(.098)	(.100)

Note. k = the number of indicators; ω_h = population value of ω_h, α = coefficient alpha; first PF = ω_h estimate based on the first unrotated principal factor; first PC = ω_h estimate based on the first unrotated principal component; HO-PF = ω_h estimate based on a higher order exploratory factor analysis; HO-PC = ω_h estimate based on a higher order principal components analysis; HO-CF = ω_h estimate based on a higher order confirmatory factor analysis; Hi-CF = ω_h estimate based on a hierarchical confirmatory factor analysis; λ_G = population values of loadings of indicators on a general factor. Standard deviations are in parentheses. As the data were aggregated across the two levels of sample size, each of the eight conditions resulting from the 2 (k: 12 vs. 20) × 4 (type of general factor loading pattern) displayed in this table contained 20 replications. The bottom two rows display the results aggregated across all 160 replications.

first PC: $M = .42$; alpha: $M = .30$), and strong higher order loadings (first PF and first PC: $M = .23$; alpha: $M = .14$) conditions. The two-way interaction of Contrast 2 × the number of indicators resulted from the number of indicators having a greater effect on alpha ($\eta^2 = .03$) than on the first PF and first PC methods ($\eta^2 = .00$), with alpha becoming more positively biased with more indicators ($k = 12 : M = .33$ vs. $k = 20, M = .39$), whereas the positive bias in the first PF and first PC methods did not get worse with more indicators ($k = 12 : M = .35$ vs. $k = 20 : M = .36$). No other effects involving Contrast 2 were significant.

To summarize the results for Contrast 2, when aggregated across all conditions, alpha ($M = .36$) and the first PF and first PC methods ($M = .35$) were comparably positively biased. However, alpha was much more positively biased when a general factor was absent and somewhat less positively biased in the weak, mixed, and strong higher order loadings conditions. Finally, the bias associated with alpha did increase with scale length, whereas the bias associated with the first PF and first PC methods did not.

Contrast 3: Exploratory versus confirmatory methods. The main effect of Contrast 3 was significant ($F_{1,144} = 119.21$, $p \le .001$) though very small ($\eta^2 = .05$), indicating that, even though the average bias associated with both types of methods was small, the average of the HO-PF and HO-PC methods ($M = .002$, $SD = .090$) was slightly less positively biased than the average of the HO-CF and Hi-CF methods ($M = .044$, $SD = .090$) when aggregated across all conditions. Not only was this main effect associated with a very small effect size, but it was also moderated by a significant interaction of Contrast 3 by type of general factor loading pattern ($F_{3,144} = 10.90$, $p \le .001$). Although this interaction was associated with a trivially small effect size ($\eta^2 = .01$), the interaction resulted from inconsistency across the general factor loading pattern conditions in terms of whether the CFA or exploratory approaches were less biased. That is, the putatively accurate CFA approaches were nonsignificantly less negatively biased in the weak higher order loadings condition ($M = -.015$) and the strong higher order loadings condition ($M = -.002$) than the putatively accurate exploratory approaches (weak higher order loadings condition: $M = -.024$, strong higher order loadings condition: $M = -.070$), whereas the exploratory approaches were significantly less positively biased in the general factor absent (confirmatory: $M = .121$, exploratory: $M = .087$) and mixed higher order loadings (confirmatory: $M = .071$, exploratory: $M = .014$) conditions (a pattern of interaction that makes the importance of the main effect questionable). Finally, the interaction of Contrast 3 × the sample size was significant ($F_{1,144} = 13.22$, $p \le .001$; $\eta^2 = .006$) but was associated with such a trivially small effect size that its practical significance is highly questionable and will not be discussed further. There were no other significant effects involving Contrast 3.

To summarize the results for Contrast 3, the results reversed from one type of general factor loading pattern to the next in terms of whether the exploratory procedures or the CFA procedures were less biased. Importantly, both the putatively accurate exploratory procedures and the putatively accurate confirmatory procedures tended to perform well in most of the conditions studied here.

Contrast 4: HO-PF versus HO-PC. The main effect of Contrast 4 was significant ($F_{1,144} = 9.98$, $p \le .005$) though trivially small ($\eta^2 = .0002$). Whereas the average bias associated with these two estimators was small, the average of the HO-PC estimator ($M = .001$, $SD = .091$) was slightly less positively biased than the average of the HO-PF estimator ($M = .003$, $SD = .088$). However, not only was the effect size for this main effect trivially small, but it was also moderated by a significant interaction of Contrast 4 by the type of general factor loading pattern ($F_{3,144} = 7.29$, $p \le .001$; $\eta^2 = .0004$). Although this interaction itself was associated with a trivially small effect size, the interaction resulted from inconsistency across the general factor loading pattern conditions in terms of whether the HO-PF or HO-PC approaches were less biased. That is, the interaction resulted from the HO-PF estimator being nonsignificantly more positively biased in the general factor absent condition (HO-PF: $M = .087$, HO-PC: $M = .086$) but nonsignificantly less negatively biased in the weak higher order loadings (HO-PF: $M = -.023$, HO-PC: $M = -.025$) and strong higher order loadings (HO-PF: $M = -.065$, HO-PC: $M = -.074$) conditions and nonsignificantly less positively biased in the mixed higher order loadings condition (HO-PF: $M = .014$, HO-PC: $M = .015$). There were no other significant effects involving Contrast 4.

To summarize the results for contrast 4, although there were significant differences in bias between the HO-PC and HO-PF estimators when averaged across all conditions, the magnitude of

this difference was trivially small, and the trivially small differences in bias between them were not consistent across the types of general factor loading patterns. Most important, both estimators performed well in all of the conditions studied here.

Contrast 5: HO-CF versus Hi-CF. The main effect of Contrast 5 was significant ($F_{1,144} = 75.64$, $p \leq .001$) though very small ($\eta^2 = .06$), indicating that, even though the average bias associated with both estimators was reasonably small, the Hi-CF estimator ($M = .020$, $SD = .100$) was slightly less positively biased than the HO-CF estimator ($M = .068$, $SD = .098$) when aggregated across all conditions. Not only was this main effect associated with a very small effect size, but it was also moderated by a significant interaction of Contrast 5 by the type of general factor loading pattern ($F_{3,144} = 16.72$, $p \leq .001$; $\eta^2 = .04$). This interaction resulted from the HO-CF estimator being significantly more positively biased in the general factor absent (HO-CF: $M = .163$, Hi-CF: $M = .079$) and mixed higher order loadings (HO-CF: $M = .078$, Hi-CF: $M = .065$) conditions but significantly less biased in the weak higher order loadings (HO-CF: $M = .032$, Hi-CF: $M = -.061$) and strong higher order loadings (HO-CF: $M = .001$, Hi-CF: $M = -.006$) conditions. Finally, the interaction of Contrast $5 \times$ the type of general factor loading pattern times the sample size was significant ($F_{3,144} = 3.11$, $p \leq .05$; $\eta^2 = .007$) but was associated with such a trivially small effect size that its practical significance is highly questionable and will not be discussed further. There were no other significant effects involving Contrast 5.

To summarize the results for Contrast 5, although there were significant differences between the HO-CF and Hi-CF estimators on average, the magnitude of this difference was very small, and the differences in bias were not consistent across the general factor loading pattern conditions. Moreover, both estimators performed well in most of the conditions studied here.

Efficiency. Table 2 shows that both alpha and the four putatively accurate methods were associated with relatively small standard deviations in almost every condition. In contrast, the first PF and first PC methods were associated with large standard deviations in the general factor absent condition and also tended to have somewhat elevated standard deviations in the weak higher order loadings condition.

Are the large standard deviations in the general factor absent condition due to random sampling error or actual variation? The results presented thus far are consistent with the predictions that the HO-PF, HO-PC, HO-CF, and Hi-CF methods would outperform alpha and the first PC and first PF methods. However, the large standard deviations observed for the first PC and first PF methods in the general factor absent condition complicate the interpretation of the results. Given the relatively small number of replications per condition, it is difficult to know whether these large standard deviations are due to random sampling error or actual variation. To more conclusively resolve this issue, more replications for the general factor absent condition were conducted. Fifty simulated samples were created within each of the four conditions resulting from the 2×2 crossing of sample size (n: 100 vs. 200) and scale length (k, 12 vs. 20); thus, a total of an additional 200 artificial data sets in the general factor absent condition were created. In addition to estimating ω_h using the first PF and first PC methods, for comparison purposes, ω_h was also estimated using the putatively accurate method with the largest standard deviation in the general factor absent condition—the HO-CF method—and alpha for each of these data sets.

Again, the results were collapsed across the two different levels of sample size as all effects involving sample size were trivially small ($\eta^2 \leq .002$). Compared with alpha ($k = 12$: $SD = .067$; $k = 20$: $SD = .037$) and the HO-CF ($k = 12$: $SD = .113$; $k = 20$: $SD = .107$) method, the standard deviations associated with the first PF method ($k = 12$: $SD = .162$; $k = 20$: $SD = .187$) were moderately large, and those associated with the first PC method ($k = 12$: $SD = .273$; $k = 20$, $SD = .260$) were very large. This evidence strongly suggests that the large standard deviations shown in Table 2 for the first PF and especially for the first PC methods reflect actual variation rather than random

sampling error. (To save space, the mean difference scores for the four methods of estimating ω_h in these simulations and the results of the inferential statistical comparing the methods in terms of their bias are not reported here but are available upon request from Richard E. Zinbarg.)

Do the results generalize beyond the case of four equal-sized group factors with perfectly simple structure? Additional limitations of the results presented thus far are that it is not clear (a) if they hold only in the case of four group factors and (b) if they hold only in the case of scales with perfectly simple and symmetric structure. As noted above, to begin to address these issues, some additional simulations were conducted based on the structure of the ASI (Reiss et al., 1986). Ten simulated samples were created within each of the four conditions resulting from the 2×2 crossing of sample size (*n*: 50 vs. 200) and type of general factor loading pattern (relatively weak vs. relatively strong); thus, a total of an additional 40 artificial data sets were created based on the structure of the ASI. In all conditions, the models used to generate the artificial data sets were created by assigning indicators their standard deviations from Zinbarg et al.'s (1997) Table 1 and their standardized group factor loadings from Zinbarg et al.'s Table 2, including cross-loadings that were greater than .15. This resulted in 11 indicators loading on only one of the three group factors, four indicators having a cross-loading on a second group factor, and one item loading on all three group factors. In the relatively strong general factor loading condition, indicators were assigned their standardized general factor loadings from Zinbarg et al.'s Table 2. This resulted in an average standardized general factor loading of $.49(SD = .13)$. In the relatively weak general factor loading condition, the standardized general factor loadings were reduced by an average of .35, such that the average standardized general factor loading equaled $.14(SD = .05)$.

ω_h estimates from all seven methods (first PF, first PC, HO-PF, HO-PC, HO-CF, Hi-CF, and coefficient alpha) for each sample were again compared to the population ω_h value by subtracting the population value from each of the estimates. For these simulations, the dimensionality of the estimated models equaled three for the higher order EFA, PCA, and CFA models and four for the hierarchical CFAs. Table 3 presents the population values of ω_h and the mean difference scores for the seven methods of estimating ω_h across the two types of structures and two levels of sample size in the set of simulations based on the structure of the ASI. The repeated-measure factor of estimate type was again decomposed into the set of five planned contrasts described above.

Contrast 1: The four putatively accurate methods versus the three putatively inaccurate methods. The main effect of Contrast 1 was significant ($F_{1,36} = 1041.60$, $p \le .001$; $\eta^2 = .68$), indicating that the average of the four putatively accurate methods was more accurate (less positively biased) than the average of the three putatively inaccurate methods when aggregated across all conditions ($M = .022, SD = .115$ vs. $M = .441, SD = .172$). Importantly, although this main effect was moderated by two interactions (discussed below), the average of the four putatively accurate methods was more accurate (less positively biased) than the average of the three putatively inaccurate methods in every condition.

The two-way interaction of Contrast 1 $\times$ the type of general factor loading pattern ($F_{1,36} = 70.30$, $p \le .001$, $\eta^2 = .05$) and the three-way interaction of Contrast 1 $\times$ the type of general factor loading pattern times the sample size interaction ($F_{1,36} = 5.18$, $p \le .05$; $\eta^2 = .004$) were significant. The significant interaction of Contrast 1 $\times$ the type of general factor loading pattern resulted from the type of general factor loading pattern having a greater effect on the putatively inaccurate methods ($\eta^2 = .89$) than on the putatively accurate methods ($\eta^2 = .21$), with the putatively inaccurate methods being even more positively biased with a weak general factor ($M = .60$) than with a relatively strong general factor ($M = .28$). The three-way interaction of Contrast 1 $\times$ the type of general factor loading pattern times the sample size was associated with such a trivially small effect size that its practical significance is highly questionable and will not be discussed further. No other effects involving Contrast 1 were significant.

Table 3

Mean Differences Between Estimated and Population Values of ω_h:
Simulations Based on the Anxiety Sensitivity Index

n	ω_h	α	First PF	First PC	HO-PF	HO-PC	HO-CF	Hi-CF
				Weak λ_G				
50	.127	.559	.521	.715	.130	.120	.118	−.053
		(.049)	(.109)	(.148)	(.153)	(.140)	(.244)	(.096)
200	.127	.529	.503	.780	.127	.151	.048	−.050
		(.031)	(.064)	(.087)	(.121)	(.122)	(.142)	(.056)
				Strong λ_G				
50	.643	.231	.244	.343	−.038	−.072	−.015	−.226
		(.026)	(.032)	(.020)	(.114)	(.107)	(.142)	(.210)
200	.643	.252	.259	.352	.073	.050	.076	−.082
		(.007)	(.007)	(.003)	(.065)	(.057)	(.061)	(.086)
		.393	.382	.548	.073	.062	.057	−.103
		(.156)	(.146)	(.220)	(.132)	(.138)	(.162)	(.142)

Note. n = sample size; ω_h = population value of ω_h, α = coefficient alpha; first PF = ω_h estimate based on the first unrotated principal factor; first PC = ω_h estimate based on the first unrotated principal component; HO-PF = ω_h estimate based on a higher order exploratory factor analysis; HO-PC = ω_h estimate based on a higher order principal components analysis; HO-CF = ω_h estimate based on a higher order confirmatory factor analysis; Hi-CF = ω_h estimate based on a hierarchical confirmatory factor analysis; λ_G population values of loadings of indicators on a general factor. Standard deviations are in parentheses. As the data were aggregated across the two levels of sample size, each of the eight conditions resulting from the 2 (k: 12 vs. 20) × 4 (type of general factor loading pattern) displayed in this table contained 20 replications. The bottom two rows display the results aggregated across all 160 replications.

Contrast 2: First PF and first PC versus alpha. The main effect of Contrast 2 ($F_{1,36} = 82.82$, $p \leq .001$, $\eta^2 = .04$) and the three-way interaction of Contrast 2 × the type of general factor loading pattern times the sample size ($F_{1,36} = 3.98$, $p \leq .05$, $\eta^2 = .002$) were significant. The significant main effect of Contrast 2 resulted from the first PF and first PC being even more biased than alpha aggregated across all conditions (first PF and first PC: $M = .46$; alpha: $M = .39$). The three-way interaction of Contrast 2 × the type of general factor loading pattern times the sample size was associated with such a trivially small effect size that its practical significance is highly questionable and will not be discussed further. No other effects involving Contrast 2 were significant.

Contrast 3: Exploratory versus confirmatory methods. The main effect of Contrast 3 was significant ($F_{1,36} = 27.14$, $p \leq .001$, $\eta^2 = .11$), indicating that the average of the confirmatory methods ($M = .003$) were even less biased than the average of the exploratory methods ($M = .132$). There were no other significant effects involving Contrast 3.

Contrast 4: HO-PF versus HO-PC. The interaction of Contrast 4 × the type of general factor loading pattern was significant ($F_{1,36} = 6.55$, $p \leq .05$) though trivially small ($\eta^2 = .007$). There were no other significant effects involving Contrast 4.

Contrast 5: HO-CF versus Hi-CF. The main effect of Contrast 5 was significant ($F_{1,36} = 30.18$, $p \leq .001$, $\eta^2 = .22$), indicating that, even though the average bias associated with

both estimators was reasonably small, the Hi-CF estimator ($M = -.10$, $SD = .14$) tended to be negatively biased, whereas the HO-CF estimator ($M = .057$, $SD = .16$) tended to be positively biased. There were no other significant effects involving Contrast 5.

To summarize the results from Contrasts 1 through 5 for the simulations based on the structure of the ASI, perhaps the most important result was that the four putatively accurate methods were again more accurate (less positively biased) and more stable across general factor loading patterns than the three putatively inaccurate methods. In addition, on average, the first PF and first PC methods were even more positively biased than alpha. There were also some differences among the four putatively accurate methods. Thus, the HO-CF and Hi-CF methods were even more accurate (less positively biased) on average than the HO-PF and HO-PC methods, and the HO-CF method tended to be slightly positively biased, whereas the Hi-CF tended to be slightly negatively biased. However, it is important to note that the HO-PF, HO-PC, HO-CF, and Hi-CF methods all performed reasonably well in the ASI simulations.

Efficiency. Table 3 shows that alpha was associated with relatively small standard deviations in each of the four conditions included in the simulations based on the structure of the ASI. In contrast, the putatively accurate CFA methods were associated with large standard deviations in the $n = 50$ condition. In addition, the first PF, first PC, HO-PF, and HO-PC also tended to have somewhat elevated standard deviations in the $n = 50$ condition (especially when also in the weak higher order loadings condition).

Discussion

The predictions that the HO-PF, HO-PC, HO-CF, and Hi-CF methods of estimating ω_h would perform better than coefficient alpha and the first PF and first PC methods received strong support. In every single condition studied here, the HO-PF, HO-PC, HO-CF, and Hi-CF methods of estimating ω_h were more accurate than alpha and the first PF and first PC methods. Moreover, with a few exceptions, the HO-PF, HO-PC, HO-CF, and Hi-CF methods all performed well (within .10 of the population value of ω_h, on average) to very well (within .05 of the population value of ω_h, on average) in terms of bias. The HO-PF, HO-PC, HO-CF, and Hi-CF methods were also reasonably efficient in almost every condition, except the $n = 50$ condition, in the simulations based on the ASI in which the CFA methods were associated with especially large standard deviations. It is also important to note that (a) differences in the size and variance of the general factor loadings, (b) variation in sample size ranging from 50 to 200, (c) increasing scale length from 12 to 20 indicators, (d) whether there were four group factors present versus three, and (e) the presence versus absence of item cross-loadings by and large had trivial effects on the performance of the HO-PF, HO-PC, HO-CF, and Hi-CF methods of estimating ω_h. Although the HO-PF, HO-PC, HO-CF, and Hi-CF methods tended to outperform alpha and the first PF and first PC methods, even with sample sizes of 50 in the simulations based on the structure of the ASI, their performance did deteriorate somewhat (especially in terms of efficiency), suggesting caution in their use and interpretation with such small sample sizes.

Although it is reasonable to ask whether there is a best method for estimating ω_h among these four methods, any differences that did emerge among them in terms of both bias and efficiency tended to be quite small. Moreover, these differences were not consistent across conditions as each of the main effects (exploratory vs. confirmatory, HO-PF vs. HO-PC, and HO-CF vs. Hi-CF) involved was moderated by interactions in which the results reversed from one condition to the next in terms of which method was even less biased than another. Furthermore, the estimates derived from these four methods were almost perfectly correlated with each other and the population values of ω_h. Thus, none of these four methods is recommended over the other three at this time on empirical grounds. However, it seems

reasonable to argue that the HO-PF, HO-CF, and Hi-CF methods should be preferred over the HO-PC method on the grounds that the theoretical model underlying PCA analysis is inconsistent with the factor-analytic model underlying ω_h. Fortunately, the results suggest that the accuracy of estimating ω_h will not suffer when researchers who do not share this preference use the HO-PC method instead.

The use of the two CFA methods (HO-CF and Hi-CF) over the accurate EFA method (HO-PF) is recommended whenever the investigator has clear theoretical or empirical expectations regarding the number of group factors to include in the measurement model and the pattern of loadings of the indicators on the common factors. The advantage of the CFA methods over the EFA method is that in the process of fitting one's CFA model to obtain the parameter estimates necessary to calculate ω_h, one can (and should) test the assumptions required for the derivation of ω_h. However, the results suggest that when investigators do not have clear expectations regarding the number of factors and/or the pattern of the loadings of the indicators on the common factors and therefore use EFA rather than CFA, the accuracy of estimating ω_h does not have to suffer in the multidimensional case as long as they use the HO-PF method rather than the first PF method.

A recommendation that should be heeded, regardless of the method chosen to estimate ω_h, is to always examine the pattern of the estimated general factor loadings prior to estimating ω_h. Such an examination constitutes an informal test of the assumption that there is a latent variable common to all of the scale's indicators that can be conducted even in the context of EFA. If the loadings were salient for only a relatively small subset of the indicators, this would suggest that there is no true general factor underlying the covariance matrix. Just such an informal assumption test would have afforded a great deal of protection against the possibility of misinterpreting the misleading ω_h estimates occasionally produced in the simulations reported here. For example, Tables 4 and 5 show the ''general'' factor loadings from the HO-CF solutions taken from the eight samples in the general factor absent condition that produced ω_h estimates of .28 or higher. The pattern of loadings on the ''general'' factor in all but the last of these cases was clearly representing one of the group factors and should be interpreted as evidence that there was no true general factor and that ω_h should therefore be set to equal 0.

It is also worth noting that the results provide yet another demonstration of the well-documented problem with coefficient alpha as an estimate of the proportion of scale variance due to a general factor. In each of the conditions studied here, alpha was a highly positively biased estimator of ω_h. The first PC and first PF methods of estimating ω_h share coefficient alpha's limitations in this regard. In fact, the first PC and first PF methods were even somewhat more positively biased than alpha in these simulations whenever a general factor was present, although they were less positively biased than alpha when a general factor was absent. In addition, the bias associated with alpha did increase with scale length as expected. The use of any of these three methods is not recommended for estimating ω_h when group factors are present, as was the case in each of the simulations studied here.

The conclusion that alpha provides a highly positively biased estimate of ω_h for multidimensional data might appear to conflict with the large literature demonstrating that alpha tends to underestimate reliability. The key to understanding this seeming contradiction lies in the distinction between a general factor and group factors made within the hierarchical factor model expressed in equation (1). Maximal reliability, stratified alpha, and related coefficients—such as McDonald's (1985, 1999) omega expressed equivalently by either equation (4) or (6)—are related to the proportion of scale variance associated with all common factors, including not only the general factor (if one is present) but also the group factors. As Cronbach (1951) noted, alpha tends to fall somewhere in between the proportion of scale variance due to a general factor and the proportion due to all common factors (i.e., alpha tends to overestimate the former while underestimating the latter).

Table 4

Sample General Factor Loading Patterns From Higher Order Confirmatory Factor Model Solutions
That Grossly Overestimated ω_h When No General Factor Was Present With 12 Indicators

	Sample			
Indicator	Sample 3 ($n = 100$)	Sample 4 ($n = 100$)	Sample 9 ($n = 200$)	Sample 10 ($n = 200$)
1	−.15	.07	.13	.06
2	−.10	0.07	.11	.05
3	−.12	0.07	.12	.06
4	.12	0.13	.08	.24
5	.11	0.16	.08	.21
6	.11	0.20	.07	.23
7	**.82**	.02	**.80**	**.49**
8	**.72**	.02	**.70**	**.46**
9	**.82**	.02	**.73**	**.52**
10	.16	**.74**	.01	.17
11	.19	**.70**	.01	.18
12	.17	**.78**	.01	.19
HO-CF	.36	.35	.30	.28
α	.51	.58	.61	.64

Note. n = sample size; HO-CF = sample estimate of ω_h using the higher order confirmatory factor analysis method; α = coefficient alpha. Factor loadings $\geq |.30|$ are listed in boldface.

Relationship to Alternative Studies Demonstrating a Positive Bias in Alpha

Models accounting for error covariances—which bear some similarities to the group factors considered here—have been considered by Green and Hershberger (2000), Komaroff (1997), and Raykov (1998a). Moreover, each of these authors has concluded that there are conditions involving correlated errors in which α has a positive bias. There are some important differences in the contributions made by these earlier articles and the current article. First, Green and Hershberger, Komaroff, and Raykov were concerned with the bias in alpha as an estimate of a reliability coefficient closer in meaning to ω rather than as an estimate of ω_h. Raykov, for example, examined cases in which a scale measured two separate, but possibly intercorrelated, factors and was interested in alpha as an estimate of the proportion of variance in scale scores due to both of these factors. There will be conditions—such as when Raykov's two factors are uncorrelated, the factor loadings are heterogeneous, and the error covariances are zero—in which alpha would continue to overestimate ω_h but would underestimate the proportion of variance in scale scores due to both of these factors that Raykov focuses on. In addition, there are at least two potential advantages of the group factor approach over the correlated errors approach. The first advantage is conceptual in that if one allows for correlated errors, it would appear to be difficult to maintain a distinction between common factors and errors/uniquenesses (for a related discussion, see Green & Hershberger, 2000, p. 253). The second advantage is that the group factor approach will often be more parsimonious. That is, one has to estimate $\frac{m(m-1)}{2}$ paths in the error covariance approach to model the associations among the m indicators loading on a particular group factor that are modeled via the estimation of m paths using the group factor approach.

Table 5

Sample General Factor Loading Patterns From Higher Order Confirmatory Factor Model Solutions
That Grossly Overestimated ω_h When No General Factor Was Present With 20 Indicators

	Sample			
Indicator	Sample 3 ($n = 100$)	Sample 4 ($n = 100$)	Sample 9 ($n = 200$)	Sample 10 ($n = 200$)
1	.16	.11	.14	.13
2	.14	.12	.12	.13
3	.16	.12	.11	.11
4	.15	.11	.12	.12
5	.14	.11	.13	.12
6	.02	**.71**	.09	.18
7	.01	**.71**	.09	.19
8	.02	**.69**	.08	.18
9	.01	**.75**	.08	.17
10	.02	**.65**	.09	.18
11	**.80**	.21	−.03	**.37**
12	**.66**	.21	−.04	**.41**
13	**.81**	.24	−.04	**.39**
14	**.67**	.21	−.03	**.40**
15	**.71**	.22	−.03	**.36**
16	.07	−.07	**.61**	**.30**
17	.08	−.07	**.70**	**.30**
18	.08	−.07	**.67**	.27
19	.07	−.07	**.67**	.28
20	.06	−.07	**.73**	**.31**
HO-CF	.35	.40	.28	.32
α	.74	.70	.73	.78

Note. $n =$ sample size; HO-CF $=$ sample estimate of ω_h using the higher order confirmatory factor analysis method; $\alpha =$ coefficient alpha. Factor loadings $\leq |.30|$ are listed in boldface.

Limitations, Future Directions, and Conclusion

This study has at least three strengths. First, a wide range of methods of estimating ω_h are evaluated for multidimensional scales, as are often encountered in practice. Second, ω_h estimates are investigated under various levels of a number of conditions, including strength of general factor loadings, scale length, and sample size. Third, the major findings were replicated in a set of simulations based on the structure of a widely used instrument, suggesting that the generality of the results is not limited to the four-factor case or to structures demonstrating perfectly simple structures.

However, this study also has limitations. First, only multivariate normal distributions and continuous indicators are used, whereas nonnormal and/or categorical data are often encountered in practice. Second, although sample size and the number of indicators were varied, further work should be done to examine whether larger effects for these variables would emerge if more levels and more extreme values were used for them.

Third, with the exception of the first PF and first PC methods, all of the EFA, CFA, and PCA models were correctly specified. That is, the CFA models correctly specified the number of group factors, and the indicators that should load on the group factors and the HO-PF and HO-PC models

correctly specified the number of first-order and second-order factors. Further work should be done to examine the extent to which estimates of ω_h are affected by model misspecification such as underfactoring and overfactoring. Such work should certainly include the case of overfactoring in the unidimensional situation, which is an important special case given that the unidimensional situation often represents the ideal of measurement. The first PF and first PC methods would be expected to perform better in the unidimensional case than in the multidimensional situations studied here, but it is unclear how much the performance of the HO-PF, HO-PC, HO-CF, and Hi-CF methods will deteriorate in the unidimensional case due to the effects of overfactoring.

Fourth, point estimation only was considered; no consideration was given to the important topic of using sample statistics to derive interval estimates for the population value of ω_h. Although further discussion of this topic is beyond the scope of this article, it is important to note that when estimating ω_h, good statistical practice would dictate that point estimation should be supplemented by interval estimation. Certainly, interval estimation of ω_h is a topic that should be studied in detail in future investigations. Bootstrapping and analytic techniques similar to thos'e described by Raykov (1998b) and Raykov and Shrout (2002) should also be useful for obtaining a standard error and confidence interval when estimating ω_h.

Fifth, all simulations included four or three group factors and therefore do not address a difficulty that arises in the estimation of ω_h when only two group factors are present. In the two–group factor case, the higher order portion of any higher order model will be underidentified as there will be only one observed correlation between the two lower order factors to estimate the two higher order factor loadings. Thus, further work is needed to examine the extent to which different identifying constraints introduce bias into the process of estimating ω_h in the two–group factor case.

Finally, it is important to note that if ω_h is applied in a mechanical fashion without careful thought, an investigator might be led to discard a potentially useful item set. For example, imagine a 20-item, effect-indicator instrument with a factor structure consisting of a very weak general factor and three group factors that are loaded on by 13, 4, and 3 indicators, respectively. The population value of ω_h would be low in this instance, and the sample estimates for the accurate methods of estimating ω_h (e.g., HO-PF, HO-PC, HO-CF, Hi-CF) should be correspondingly low. It is not difficult to imagine some scale users or developers deciding to abandon the entire item set as a result. Such a decision, however, would represent an instance of "throwing out the baby with the bath water," as it is likely that the 13 indicators loading on the first group factor might form a scale with a more than adequate value of ω_h. Depending on the size of the loadings, even the four indicators loading on the second group factor might form a scale with an adequate ω_h, as might the three indicators loading on the third group factor. If so, one might be able to derive three reasonable subscales from these 20 indicators, even though a single total score based on all 20 indicators would be ill advised.

Thus, it is important to keep in mind that a small value of ω_h merely indicates that the empirical justification for aggregating across all indicators to derive a single, effect-indicator total score is weak. Moreover, the proper interpretation of ω_h must take into account not only the estimated value of ω_h but also the matrix of factor loadings. When ω_h estimates are relatively high, one needs to pay attention to both the general factor loadings (as discussed earlier to protect against the possibility of overestimating ω_h when a general factor is not actually present) and the group factor loadings (to see if [a] subscale scores might be considered in addition to a total scale score or [b] there are indicators that might simply be dropped because they do not have a salient loading on either the general factor or one of the group factors). When ω_h estimates are relatively low, one needs to pay particular attention to the group factor loadings to see if (a) reasonable subscales might be created, even though a single total score would not be well justified; (b) some indicators might be

dropped, leaving a shorter instrument that yields a single, well-justified total score; or (c) additional indicators are needed to produce an instrument that yields several subscale scores.

In conclusion, four methods of estimating ω_h (Hi-CF, HO-PF, HO-PC, and HO-CF) were identified that clearly outperformed alpha and the first PF and first PF methods. In addition, the accuracy of these four methods of estimating ω_h was stable across the different general factor loading patterns, with sample sizes ranging from 50 to 200 and scale lengths ranging from 12 to 20 indicators. Based on theoretical considerations in addition to performance in these simulations, it is recommended that the HO-CF and/or the Hi-CF methods be used to estimate ω_h whenever the investigator has a clear a priori measurement model that can be tested via CFA. When the investigator does not have a clear a priori measurement model, the HO-PF method is recommended, although the HO-PC method will probably produce similar results. Regardless of which of these methods an investigator uses to estimate ω_h, in practice, the proper interpretation of the implications of ω_h for scale development and use requires an examination of the pattern of group factor loadings. Finally, one should always assess the reasonableness of the assumption that all of the scale's indicators measure a latent variable in common prior to estimating ω_h. One informal test of this assumption that can be performed in either a CFA or an EFA context is to examine the pattern of estimated general factor loadings. Whenever possible, this assumption should also be tested formally via the appropriate use of CFA.

Notes

1. That is, there will be no group factors when the scale is unidimensional. In this case, the vector of unstandardized general factor loadings is given by the vector of unstandardized loadings on the single common factor.
2. It is obvious that $\frac{m(m-1)}{2} \geq m$ only when $m > 2$. What may be less obvious is that even when $m = 2$, the number of group factor loadings estimated will equal the number of error covariances estimated as a constraint has to be imposed on one of the two group factor loadings for identification purposes. Thus, the group factor approach will not only usually be more parsimonious but will always be at least as parsimonious as an error covariance approach that models the associations among the m indicators loading on a particular group factor.

References

Beck, A. T., & Steer, R. A. (1993a). *The Beck Anxiety Inventory manual*. San Antonio, TX: The Psychological Corporation.

Beck, A. T., & Steer, R. A. (1993b). *The Beck Depression Inventory manual*. San Antonio, TX: The Psychological Corporation.

Bentler, P. M. (1989). *EQS structural equations program manual*. Los Angeles: BMDP Statistical Software.

Bentler, P. M., & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research*, *25*, 67-74.

Bollen, K. (1989). *Structural equations with latent variables*. New York: John Wiley.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*, 305-314.

Costa, P. J., & McCrae, R. R. (1985). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.

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

Cronbach, L. J., Schoneman, P., & McKie, D. (1965). Alpha coefficients for stratified tests. *Educational and Psychological Measurement*, *25*, 291-312.

Drewes, D. W. (2000). Beyond the Spearman-Brown: A structural approach to maximal reliability. *Psychological Methods*, *5*, 214-227.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272-299.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105-146). New York: Macmillan.

Foa, E. B., Riggs, D. S., Dancu, C. V., & Olasov-Rothbaum, B. (1993). Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *Journal of Traumatic Stress*, *6(4)*, 459-473.

Goldberg, L. R. (1999). The curious experiences survey, a revised version of the dissociative experiences scale: Factor structure, reliability and relations to demographic and personality variables. *Psychological Assessment*, *11*, 134-145.

Gorsuch, R. L. (1983). *Factor analysis*. Hillsdale, NJ: Lawrence Erlbaum.

Gorsuch, R. L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research*, *25*, 33-39.

Green, S. B., & Hershberger, S. L. (2000). Correlated errors in true score models and their effect on coefficient alpha. *Structural Equation Modeling. 7*, 251-270.

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

Jensen, A. R., & Weng, L. (1994). What is a good *g*? *Intelligence*, *18*, 231-258.

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

Jöreskog, K. G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In R. C. Atkinson, D. H. Krantz, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (pp. 1-56). San Francisco: Freeman.

Kelley, T. L. (1942). The reliability coefficient. *Psychometrika*, *7*, 75-83.

Komaroff, E. (1997). Effect of simultaneous violations of essential Tau-equivalence and uncorrelated error on coefficient alpha. *Applied Psychological Measurement*, *21*, 337-348.

Kristof, W. (1974). Estimation of reliability and true score variance from a split of a test into three arbitrary parts. *Psychometrika*, *39*, 491-499.

Li, H., Rosenthal, R., & Rubin, D. B. (1996). Reliability measurement in psychology: From Spearman-Brown to maximal reliability. *Psychological Methods*, *1*, 98-107.

Loehlin, J. (1998). *Latent variable models: An introduction to factor, path and structural analysis* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Lord, F. (1956). Sampling error due to choice of split in split-half reliability coefficients. *Journal of Experimental Education*, *24*, 245-249.

McArdle, J. J. (1990). Principles versus principals of structural factor analyses. *Multivariate Behavioral Research*, *25*, 81-87.

McDonald, R. P. (1963). A general approach to nonlinear factor analysis. *Psychometrika*, *27*, 397-415.

McDonald, R. P. (1965). Difficulty factors and nonlinear factor analysis. *British Journal of Mathematical and Statistical Psychology*, *18*, 11-23.

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

McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, *34*, 100-117.

McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Lawrence Erlbaum.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.

McNemar, Q. (1946). Opinion-attitude methodology. *Psychological Bulletin*, *43*, 289-374.

Mohlman, J. & Zinbarg, R. (2000). The structure and correlates of anxiety sensitivity in the elderly. *Psychological Assessment*, *12*, 440-446.

Mosier, C. I. (1936). A note on indicator analysis and the criterion of internal consistency. *Psychometrika*, *1*, 275-282.

Mulaik, S. (1990). Blurring the distinctions between component analysis and common factor analysis. *Multivariate Behavioral Research*, *25*, 53-59.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. (1975). *SPSS: Statistical package for the social sciences* (2nd ed.). New York: McGraw-Hill.

Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

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

Pruzek, R., & Rabinowitz, S. (1981). A class of simple methods for exploratory structural analysis. *American Educational Research Journal*, *18*, 173-189.

Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential Tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research*, *32*, 329-353.

Raykov, T. (1998a). Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement, 22*, 375-385.

Raykov, T. (1998b). A method for obtaining standard errors and confidence intervals of composite reliability for congeneric items. *Applied Psychological Measurement, 22*, 369-374.

Raykov, T., & Shrout, P. E. (2002). Reliability of scales with general structure: Point and interval estimation using a structured equation modeling approach. *Structural Equation Modeling, 9*, 195-212.

Reiss, S., Peterson, R. A., Gursky, D. M., & McNally, R. J. (1986). Anxiety sensitivity, anxiety frequency, and the prediction of fearfulness. *Behaviour Research and Therapy*, *24*, 1-8.

Reuterberg, S.-E., & Gustafsson, J.-E. (1992). Confirmatory factor analysis and reliability: Testing measurement model assumptions. *Educational and Psychological Measurement*, *52*, 795-811.

Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research*, *14*, 57-74.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*, 53-62.

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

Snook, S. C., & Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin*, *106*, 148-154.

Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science*, *10*, 297-303.

Velicer, W. F. (1974). A comparison of the stability of factor analysis, principal components analysis, and rescaled image analysis. *Educational and Psychological Measurement*, *34*, 563-572.

Velicer, W. F. (1976). The relation between factor score estimates, image scores and principal component scores. *Educational and Psychological Measurement*, *36*, 149-159.

Velicer, W. F. (1977). An empirical comparison of the similarity of principal component, image, and factor patterns. *Multivariate Behavioral Research*, *12*, 3-22.

Velicer, W. F., & Fava, J. L. (1987). An evaluation of the effects of variable sampling on component, image and factor analysis. *Multivariate Behavioral Research*, *22*, 193-209.

Velicer, W. F., & Jackson, D. N. (1990a). Component analysis versus common factor analysis: Some further observations. *Multivariate Behavioral Research*, *25*, 97-114.

Velicer, W. F., & Jackson, D. N. (1990b). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, *25*, 1-28.

Velicer, W. F., Peacock, A. C., & Jackson, D. N. (1982). A comparison of component and factor patterns: A Monte Carlo approach. *Multivariate Behavioral Research*, *17*, 371-388.

Vernon, P. E. (1969). *Intelligence and cultural environment*. London: Methuen.

Waller, N., Lilienfeld, S., Tellegen, A., & Lykken, D. (1991). The Tridimensional Personality Questionnaire: Structural validity and comparison with the Multidimensional Personality Questionnaire. *Multivariate Behavioral Research*, *26*, 1-23.

Watson, D., & Clark, L. A. (1992). Affects separable and inseparable: On the hierarchical arrangement of the negative affects. *Journal of Personality & Social Psychology*, *62*, 489-505.

Widaman, K. (1990). Bias in pattern loadings represented by common factor analysis and component analysis. *Multivariate Behavioral Research*, *25*, 89-95.

Widaman, K. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, *28*, 263-311.

Zimmerman, D. W., Zumbo, B. D., & Lalonde, C. (1993). Coefficient alpha as an estimate of test reliability under violation of two assumptions. *Educational and Psychological Measurement*, *53*, 33-49.

Zinbarg, R. E., & Barlow, D. H. (1996). The structure of anxiety and the anxiety disorders: A hierarchical model. *Journal of Abnormal Psychology*, *105*, 181-193.

Zinbarg, R. E., Barlow, D. H., & Brown, T. A. (1997). The hierarchical structure and general factor saturation of the Anxiety Sensitivity Index: Evidence and implications. *Psychological Assessment*, *9*, 277-284.

Zinbarg, R., Mohlman, J., & Hong, N. (1999). Dimensions of anxiety sensitivity. In S. Taylor (Ed.), *Anxiety sensitivity: Theory, research, and treatment of the fear of anxiety* (pp. 83-113). Mahwah, NJ: Lawrence Erlbaum.

Zinbarg, R., 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*, 122-133.

Acknowledgments

Author's Address

Address correspondence to Richard E. Zinbarg, Psychology Department, Northwestern University, 2029 Sheridan Rd., Evanston, IL 60208-2710; e-mail: rzinbarg@northwestern.edu.