

CRONBACH'S α , REVELLE'S β , AND MCDONALD'S ω_H :
THEIR RELATIONS WITH EACH OTHER AND TWO ALTERNATIVE
CONCEPTUALIZATIONS OF RELIABILITY

RICHARD E. ZINBARG

NORTHWESTERN UNIVERSITY,
THE FAMILY INSTITUTE AT NORTHWESTERN UNIVERSITY

WILLIAM REVELLE

NORTHWESTERN UNIVERSITY

IFTAH YOVEL

NORTHWESTERN UNIVERSITY

WEN LI

NORTHWESTERN UNIVERSITY

We make theoretical comparisons among five coefficients — Cronbach's α , Revelle's β , McDonald's ω_h , and two alternative conceptualizations of reliability. Though many end users and psychometricians alike may not distinguish among these five coefficients, we demonstrate formally their nonequivalence. Specifically, whereas there are conditions under which α , β , and ω_h are equivalent to each other and to one of the two conceptualizations of reliability considered here, we show that equality with this conceptualization of reliability and between α and ω_h holds only under a highly restrictive set of conditions and that the conditions under which β equals ω_h are only somewhat more general. The nonequivalence of α , β , and ω_h suggests that important information about the psychometric properties of a scale may be missing when scale developers and users only report α as is almost always the case.

1. Introduction

In this article, we make theoretical comparisons among five coefficients — Cronbach's α , Revelle's β , McDonald's ω_h , and two alternative conceptualizations of reliability — indexing internal psychometric properties of scale scores applicable with effect indicator scales when we are interested in sampling fluctuations resulting from the sampling of items (Lord, 1955).¹ As our general model, we adopt the hierarchical factor model (Schmid & Leiman, 1957; Tucker, 1940; Wherry, 1959). We chose this model because one of the coefficients included in our comparisons is defined only within the hierarchical factor model and each of the other four indices can under many conditions be expressed in terms of the parameters of this model. Though many

Preparation of this article was supported by the Patricia M Nielsen Research Chair of the Family Institute at Northwestern University.

We thank Lewis R. Goldberg, Win Hill, Dan McAdams, Tony Z. Tang and especially Roderick P. McDonald for their comments on earlier drafts of portions of this article.

Requests for reprints should be sent to Richard Zinbarg, 102 Swift Hall, 2029 Sheridan Rd., Northwestern University, Evanston, IL 60208-2710, USA. E-mail: rzinbarg@northwestern.edu.

¹As defined by Bollen and Lennox (1991) an effect indicator scale is comprised of indicators that are dependent on the latent variable they are measuring. In contrast, Bollen and Lennox define a causal indicator scale as being comprised of indicators that cause the latent variable they are measuring.

end users and psychometricians alike do not appear to distinguish among the five coefficients compared here, we will demonstrate formally their nonequivalence in this article. Whereas there are conditions under which α , β , and ω_h are equivalent to each other and to one of the two conceptualizations of reliability considered here, we will show that equality with this conceptualization of reliability and between α and ω_h holds only under a highly restrictive set of conditions and that the conditions under which β equals ω_h are only somewhat more general.

Though our results are related to ones reported in earlier articles, there are some important differences. Whereas Lord and Novick (1968) and Novick and Lewis (1967) compared α with reliability, they did not consider β and ω_h as we have done. McDonald (1970, 1999) considered the relation between α and ω_h only and did not consider conditions of multidimensionality as we have done. Revelle (1979) provided examples in which α can be high when β (and ω_h) equals 0. As his examples covered only a limited set of conditions, however, the generality of his results was unclear unlike the more general results of our formal comparison which shows that α can be either less than or greater than ω_h depending on several factors which we explicate. Finally, models having unequal factor loadings and accounting for error covariances — which bear some similarities to the group factors we consider in that both error covariances and group factors account for covariances among items that cannot be accounted for by a general factor — have been considered by Green and Hershberger (2000), Komaroff (1997), and Raykov (1997, 1998a, 2001a, b). However, each of these authors has concluded that there are conditions in which α overestimates reliability in contrast to our conclusion that α is a lower bound to reliability. The reason for this conflict is that we consider group factor variance to be a component of true scores and thus a contributor to reliable variance whereas Green and Hershberger (2000), Komaroff (1997), and Raykov (1997, 1998a, b, 2001a, 2001b) do not consider what they call error covariance to be a component of true scores. An advantage of our approach of using group factors to account for item covariances that cannot be accounted for by a general factor is that it will often be more parsimonious. One has to estimate $[m(m - 1)]/2$ paths when accounting for these item covariances using error covariances to model the associations among the m items loading on a particular group factor that are modeled via estimating m paths using our group factor approach.² In addition, Green and Hershberger (2000), Komaroff (1997), and Raykov (1997, 1998a, b, 2001a, 2001b) do not consider β and ω_h .

2. The Hierarchical Factor Model and McDonald's ω_h

Before defining ω_h , it is useful to review the basic structure of the hierarchical factor model. Within the hierarchical factor model, it is assumed that observed scores may be decomposed into four parts

$$\mathbf{x} = \mathbf{c}g + \mathbf{A}\mathbf{f} + \mathbf{D}\mathbf{s} + \mathbf{e}, \quad (1)$$

where \mathbf{x} is the $k \times 1$ vector of observed scores on the k scale items, g is a general factor (i.e., a factor that is common to all k scale items), \mathbf{c} is a $k \times 1$ vector of unstandardized general factor loadings, \mathbf{f} is a $r \times 1$ vector of group factors (i.e., factors that are common to some but not all k items) 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 that are 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. The

²It is obvious that $[m(m - 1)]/2 \geq m$ only when $m \geq 3$. 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, our 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 items loading on a particular group factor.

model expressed in (1) assumes also that all factors (g , f and s) are uncorrelated with each other and with e , and without loss of generality that the variance of each of the common factors (g and those in f) equals 1. In addition, each of the group factors (in f) are required to have non-zero loadings on at least three indicators to identify the model in the absence of additional constraints (e.g., Bollen, 1989).

Given the model in (1), the proportion of variance in the scale scores accounted for by a general factor³ — $\omega_{\text{hierarchical}}$ (ω_h) — is defined as

$$\omega_h = \frac{\mathbf{1}'\mathbf{c}\mathbf{c}'\mathbf{1}}{\text{Var}(X)}, \tag{2}$$

where $\text{Var}(X)$ is the variance in the scale scores obtained by summing the k indicators comprising the scale.

The parameter expressed in (2) was originally identified by McDonald (1985, 1999). Specifically, if λ_{gi} is the loading of indicator i on the general factor then (2) is algebraically equivalent to McDonald’s (1985) Equation 7.3.8 or to McDonald’s (1999) Equation 6.20a that is more efficient to use when making calculations by hand and, using our notation of ω_h , reads

$$\omega_h = \frac{(\sum_{i=1}^k \lambda_{gi})^2}{\text{Var}(X)}. \tag{3}$$

However, McDonald (Equation 7.3.9., 1985; Equation 6.2.1, 1999) also identified an expression that is algebraically equivalent to the reliability coefficient expressed in (12) below and used the same label of ω for it and (3). In differentiating the labels for these two parameters to prevent confusing them in the future, two considerations led us to retain McDonald’s (1985, 1999) label of ω for (12) but give (2) the new label of ω_h . The primary consideration was that the parameter expressed by (2) — or equivalently by (3) — is specific to the hierarchical factor model whereas (12) is not (McDonald, personal communication, October 25, 2002). In addition, other references in the literature to McDonald’s ω have tended to focus on expressions equivalent to (12) (e.g., Raykov & Shrout, 2002).

3. Revelle’s (1979) β

Revelle (1979) proposed an index labeled coefficient beta (β) that he showed equals the proportion of variance in scale scores accounted for by a general factor under more general conditions than α . As we shall demonstrate, however, those conditions are only somewhat more general than is the case for α . The formal definition of β is $\beta = k^2 \bar{\sigma}_{ij'}/\text{Var}(X)$ where $\bar{\sigma}_{ij'}$ is the average covariance between items in different halves for the worst split (i.e., the split that minimizes this average covariance but the split “halves” can be of unequal lengths). The relation between α and β is readily apparent from an examination of Cronbach’s (1951, p. 304) Equation 16 for coefficient α

$$\alpha = \frac{k^2 \bar{\sigma}_{ij}}{\text{Var}(X)}, \tag{4}$$

where $\bar{\sigma}_{ij}$ is the average interitem covariance. Thus, it is obvious that $\beta \leq \alpha$.

³Though expressed in terms corresponding to a scale obtained by summing the k indicators comprising the scale, this expression for ω_h also provides the proportion of variance in the scale score accounted for by a general factor for scale scores obtained by averaging the k items. Please contact Richard E. Zinbarg for an algebraically equivalent expression to (2) that is expressed in terms of the variance in scale scores obtained by averaging rather than summing the k items.

4. Two Factor Analytic Conceptualizations of Reliability: ρ and ω

Before discussing relations with reliability, we must first define precisely what we mean by reliability. As noted by several authors (e.g., Bollen, 1989, Joreskog, 1974; Lord & Novick, 1968), fundamental to understanding the link between the factor analytic model and the classical test theory concept of reliability is the relation of the classical test theory notion of true score to the latent factors of factor analysis. In classical test theory, observed scores on the k items of a scale are decomposed into two parts $x = t + e$ where t is the $k \times 1$ vector of true scores and e is the $k \times 1$ vector of random error scores that are uncorrelated with t . Given the hierarchical factor model in (1), one can define two alternative reliability indices depending upon whether one believes that the vector of specific factors, s , can be meaningfully differentiated from e and included as part of the definition of true score. One position is that this distinction can be made in principle in that the specific factors are systematic and can contribute to correlations between the items and other (non-test) variables whereas the components of e are random errors that can only attenuate correlations with other variables. Advocates of this position would also argue that whereas specific factors and random errors are typically confounded empirically, they can sometimes be differentiated from each other with the use of appropriate longitudinal designs and analyses.⁴ Following Lord and Novick (1968, p. 535) and McDonald (1970), when we relate t to the factors in a hierarchical factor model we obtain

$$t = cg + Af + Ds. \quad (5)$$

Given the definition of true score expressed in (5), reliability is given by

$$\rho = \frac{\mathbf{1}'\mathbf{c}\mathbf{c}'\mathbf{1} + \mathbf{1}'\mathbf{A}\mathbf{A}'\mathbf{1} + \mathbf{1}'\mathbf{D}\mathbf{D}'\mathbf{1}}{\text{Var}(X)}. \quad (6)$$

An alternative position is that there is no good ground for distinguishing specific factors from the components of e (e.g., Bollen, 1989; McDonald, 1985, 1999). Thus, it has been argued that both s and e are errors in x with respect to measuring the factors that are common to the items (Bollen, 1989). In addition, it has been argued there will often be difficulties in estimating the specific factors and distinguishing them from the components of e even in a longitudinal design depending on the temporal stability of the specific factors and the retest intervals chosen (McDonald, 1985, 1999). According to this position, we should rewrite (1) as $x = cg + Af + u$, where s and e combine to form u — the “unique” component of x , $u = Ds + e$, and when t is related to the factors in a hierarchical factor model, we obtain

$$t = cg + Af. \quad (7)$$

Given the definition of true score expressed in (7), reliability is given by

$$\omega = \frac{\mathbf{1}'\mathbf{c}\mathbf{c}'\mathbf{1} + \mathbf{1}'\mathbf{A}\mathbf{A}'\mathbf{1}}{\text{Var}(X)}. \quad (8)$$

It is readily apparent from (6) and (8) that $\omega \leq \rho$ with equality holding if and only if there are no item specific factors as noted by McDonald (1970).

The square root of ρ provides an upper bound to the extent to which the scale can correlate with another measure. Perhaps the most compelling motivation for ω was provided by McDonald (1978) who showed that if the scale's indicators are representative of the domain from which they are sampled then ω equals the squared multiple correlation between the scale score and

⁴For detailed discussions of this point see Blalock (1970), Heise (1969) and Wiley and Wiley (1970).

the domain score (defined as the sum of all the indicators in the infinite domain remaining after selecting the scale’s indicators). Rather than choose between these two alternative conceptualizations of reliability, we will include both ρ as expressed in (6) and ω as expressed in (8) in our comparisons to allow adherents of either position to see how α , β , and ω_h relate to their preferred conceptualization of reliability.

5. Relations Among α , β , ω_h and Reliability

Given (2) and (6),

$$\rho - \omega_h = \frac{\mathbf{1}'\mathbf{A}\mathbf{A}'\mathbf{1} + \mathbf{1}'\mathbf{D}\mathbf{D}'\mathbf{1}}{\text{Var}(X)}, \tag{9}$$

and, given (2) and (8), $\omega - \omega_h = \mathbf{1}'\mathbf{A}\mathbf{A}'\mathbf{1}/\text{Var}(X)$. That is, regardless of which of the two alternative conceptualizations of reliability that one adopts — ρ or ω , reliability will always be greater than ω_h when group factors are present. Further, the differences between ω_h and reliability can be substantial when group factors are present. In fact, as has been demonstrated many times (e.g., Cronbach, 1951; Green, Lissitz, & Mulaik, 1977; Revelle, 1979), reliability can be high even in the special case of a scale that does not tap a general factor in which (6) reduces to $\rho = (\mathbf{1}'\mathbf{A}\mathbf{A}'\mathbf{1} + \mathbf{1}'\mathbf{D}\mathbf{D}'\mathbf{1})/\text{Var}(X)$ and (8) reduces to $\omega = \mathbf{1}'\mathbf{A}\mathbf{A}'\mathbf{1}/\text{Var}(X)$. In contrast, ω_h equals zero by definition in such cases.

The relations among α and β with ω_h and with the two reliability indices discussed above are complex and vary depending on both (a) whether the scale in question is multidimensional or unidimensional and (b) whether the loadings on the general factor are all equal.⁵ Thus, below we discuss the relations among these indices under each of the four conditions produced by the crossing of dimensionality (multidimensional vs. unidimensional) and equality of general factor loadings (all are equal vs. not all are equal). The relations of α , β , ω_h and reliability across these four conditions are summarized in Table 1.

5.1. Case I: Multidimensional, Unequal General Factor Loadings

In what is likely to be the most general case of a multidimensional scale with unequal general factor loadings, and expressing alpha in terms of (4) and the parameters of the hierarchical factor model given by (1) leads to

$$\alpha = \frac{k}{k - 1} \left(\frac{\mathbf{1}'\mathbf{c}\mathbf{c}'\mathbf{1} - \mathbf{c}'\mathbf{c} + \mathbf{1}'\mathbf{A}\mathbf{A}'\mathbf{1} - \text{Tr}(\mathbf{A}\mathbf{A}')}{\text{Var}(X)} \right). \tag{10}$$

Given (6) and (10),

$$\rho - \alpha = \frac{\mathbf{1}'\mathbf{D}\mathbf{D}'\mathbf{1}}{\text{Var}(X)} + \frac{k^2}{(k - 1)\text{Var}(X)} \left\{ \left(\frac{\mathbf{c}'\mathbf{c}}{k} - \frac{\mathbf{1}'\mathbf{c}\mathbf{c}'\mathbf{1}}{k^2} \right) + \left(\frac{\text{Tr}(\mathbf{A}'\mathbf{A})}{k} - \frac{\mathbf{1}'\mathbf{A}\mathbf{A}'\mathbf{1}}{k^2} \right) \right\}, \tag{11}$$

and, given (8) and (10),

$$\omega - \alpha = \frac{k^2}{(k - 1)\text{Var}(X)} \left\{ \left(\frac{\mathbf{c}'\mathbf{c}}{k} - \frac{\mathbf{1}'\mathbf{c}\mathbf{c}'\mathbf{1}}{k^2} \right) + \left(\frac{\text{Tr}(\mathbf{A}'\mathbf{A})}{k} - \frac{\mathbf{1}'\mathbf{A}\mathbf{A}'\mathbf{1}}{k^2} \right) \right\}. \tag{12}$$

Thus, α is less than reliability when group factors are present and there is variance in the general factor loadings regardless of which of the two alternative conceptualizations of reliability that one

⁵One anonymous reviewer found our use of the term unidimensional to be unclear. To avoid confusion on this important point, by unidimensional we mean that the scale has only 1 common factor.

TABLE 1.

Summary of relations among α , β , ω_h and reliability as a function of scale dimensionality and equality of general factor loadings

	Unequal general factor loadings	Equal general factor loadings
Multidimensional	Case I $\{\beta < \alpha < \omega \leq \rho$ $\omega_h < \omega \leq \rho$	Case II Case II $\beta = \omega_h < \alpha < \omega \leq \rho$
Unidimensional	Case III $\beta < \alpha < \omega_h = \omega \leq \rho$	Case IV (Essential tau equivalence) $\alpha = \beta = \omega_h = \omega \leq \rho$

Note: The relations between α and ω_h and between β and ω_h depend on the specifics of the factor pattern for Case I structures. Though the Case IV result that $\alpha \leq \rho$ might appear to contradict the proof originally published by Novick and Lewis (1967) that α is a lower bound of reliability with equality holding under conditions of essential tau equivalence, it must be remembered that our ρ reflects a different definition of true score than that used by Novick and Lewis. The Case IV result that $\alpha = \omega$ does accord with the Novick and Lewis proof.

adopts. Note that since it is always true that $\beta \leq \alpha$, it follows that β is also less than reliability in this case also. Furthermore, (11) and (12) show that the extent to which α is less than reliability is a positive function of (a) the variability in the general factor loadings and (b) the variability in the group factor loadings. Equation (11) also shows that if one defines true scores as including specific factors as in (5) then the proportion of variance in scale scores due to item specific factors is a third factor that determines the extent to which α is less than reliability.

Given (2) and (10),

$$\omega_h - \alpha = \frac{k^2}{(k - 1)\text{Var}(X)} \left\{ \left(\frac{c'c}{k} - \frac{(1'c)^2}{k^2} \right) + \left(\frac{\text{Tr}(AA')}{k} - \frac{1'AA'1}{k^2} \right) \right\} - \left\{ \frac{1'AA'1}{\text{Var}(X)} \right\}. \quad (13)$$

Though the relation between α and ω_h is complex in this case, (13) shows that α can be greater than, equal to or less than ω_h when group factors are present and there is variance in the general factor loadings depending on the group factor saturation (i.e., the proportion of variance in scale scores due to group factors), the variance in the general factor loadings, and the variance in the group factor loadings. More specifically, (13) reveals that increasing group factor saturation will tend to elevate α relative to ω_h whereas increases in the variance of either the general factor loadings and/or the group factor loadings will tend to depress the value of α relative to ω_h . ω_h will be greater than α only when

$$\frac{k^2}{(k - 1)} \left\{ \left(\frac{c'c}{k} - \frac{(1'c)^2}{k^2} \right) + \left(\frac{\text{Tr}(AA')}{k} - \frac{1'AA'1}{k^2} \right) \right\} > 1'AA'1,$$

that is only when $k^2/(k - 1)$ times the sum of the variance in the general factor loadings plus the sum of the variance in the loadings on each of the group factors exceeds the group factor saturation. Whereas the variance in the group factor loadings tends to increase for a fixed value of k as the amount of variance in scale scores due to group factor increases, these two parameters can vary at least somewhat independently from each other and from the variance in general factor loadings. Thus, for example, when a scale has a factor pattern with a lot of variability in the general factor loadings and relatively weak group factors, alpha can be less than ω_h . For example, consider scale A that consists of six items with the first three items loading 0.2 on a general factor and the last three items loading 0.9 on the general factor. In addition, the first three items load 0.2 on one group factor and the last three items load 0.2 on a second group factor. If all of the items have unit variance then $\omega_h = 0.73$ whereas $\alpha = 0.71$ (and $\beta = 0.44$).

On the other hand, when there is little variability in the general factor loadings and relatively strong group factors, alpha will tend to be greater than ω_h . In fact, alpha can be considerably greater than ω_h in such cases. Consider another example of a six-item scale — scale B. For scale B, the first three items load 0.4 on a general factor whereas the last three items load 0.5 on the general factor. In addition, the first three items load 0.6 on one group factor and the last three items load 0.6 on a second group factor. If all of the items have unit variance then $\omega_h = 0.45$ whereas $\alpha = 0.76$ (and $\beta = 0.44$).

The relation between β and ω_h is also complex in this case. Unfortunately, it is not amenable to formal analysis as there is no general rule for producing the worst split half that holds across the different factor patterns that are possible when group factors are present and the general factor loadings are not all equal. Thus, β can be greater than, equal to or less than ω_h in this case depending on the specifics of the factor pattern. For example, consider cases in which there is variance in the general factor loadings but the group factors are all loaded on by the same number of items without any variance in the size of the non-zero loadings on the group factors. Further imagine that the variance in general factor loadings is also perfectly confounded with group factor membership. That is, all the items with the weakest loadings on the general factor load on the same group factor whereas all the items with the next weakest loadings on the general factor load on a second group factor and so on. In such cases, the items with the weakest loadings on the general factor will be assigned to the same half when the worst split is created causing the average covariance between items in different halves for the worst split to underestimate the strength of the common construct in the remaining items. Both scales A and B discussed above provide examples of such structures in which β is smaller than ω_h with the difference between β and ω_h being quite large in the case of scale A (0.44 vs. 0.73).

It is also possible to imagine structures, again involving heterogeneity in general factor loadings, in which beta will be greater than ω_h . For example, consider a final example of a six-item scale — scale C. For scale C, the loadings on the general factor are 0.9, 0.8, 0.5, 0.5, 0.5 and 0.1. In addition, the first two items each load 0.2 on one group factor whereas the last four items all load 0.8 on a second group factor. In this case, the worst split will be produced by splitting off Item 6 from the first five items. If all the items have unit variance then $\omega_h = 0.53$ whereas $\beta = 0.62$ for the scale score obtained by summing all six items (and $\alpha = 0.85$). Though the differences among α , β , and ω_h are not large in some of our examples, given that the examples do not incorporate sampling error, any difference illustrates that neither α nor β are the same parameter as ω_h under these conditions.

5.2. Case II: Multidimensional, Equal General Factor Loadings

For multidimensional scales with equal general factor loadings, (10) reduces to $\alpha = [k/(k - 1)][(\mathbf{1}'\mathbf{c}\mathbf{c}'\mathbf{1} + \mathbf{1}'\mathbf{A}\mathbf{A}'\mathbf{1} - \text{Tr}(\mathbf{A}\mathbf{A}'))/\text{Var}(X)]$. Thus, (11) reduces to

$$\rho - \alpha = \frac{\mathbf{1}'\mathbf{D}\mathbf{D}'\mathbf{1}}{\text{Var}(X)} + \frac{k^2}{(k - 1)\text{Var}(X)} \left\{ \frac{\text{Tr}(\mathbf{A}\mathbf{A}')}{k} - \frac{\mathbf{1}'\mathbf{A}\mathbf{A}'\mathbf{1}}{k^2} \right\}, \tag{14}$$

and (12) reduces to

$$\omega - \alpha = \frac{k^2}{(k - 1)\text{Var}(X)} \left\{ \frac{\text{Tr}(\mathbf{A}\mathbf{A}')}{k} - \frac{\mathbf{1}'\mathbf{A}\mathbf{A}'\mathbf{1}}{k^2} \right\} \tag{15}$$

(14) and (15) reveal that the extent to which α is less than reliability in this case increases as variability in the group factor loadings increases. Equation (14) also shows in this case that if one defines true scores as including specific factors, then the proportion of variance in scale scores due to item specific factor, also influences the extent to which α is less than reliability.

Similarly, (13) reduces to

$$\omega_h - \alpha = \frac{k}{(k-1)\text{Var}(X)} (\text{Tr}(\mathbf{A}\mathbf{A}') - \mathbf{1}'\mathbf{A}\mathbf{A}'\mathbf{1}). \quad (16)$$

Equation (16) reveals that α will always be greater than ω_h in this case as $\text{Tr}(\mathbf{A}\mathbf{A}') - \mathbf{1}'\mathbf{A}\mathbf{A}'\mathbf{1} < 0$.

β can be easily shown to be algebraically equivalent to ω_h when the general factor loadings are all equal. First, consider the relation between the covariance between any two items and their factor pattern loadings. When group factors are present and the general factor loadings are all equal ($\lambda_{gi} = \lambda_{gj} = \dots = \lambda_{gk} = \bar{\lambda}_g$), the covariance of any two items, i and j , loading on different group factors equals

$$\sigma_{ij} = \bar{\lambda}_g^2 \quad (17)$$

whereas the covariance of any two items, l and m , loading on the same group factor, f_r , equals

$$\sigma_{lm} = \bar{\lambda}_g^2 + \lambda_{f_r l} \lambda_{f_r m} \quad (18)$$

From (17) and (18) it is obvious that $\sigma_{ij} < \sigma_{lm}$. Thus, when group factors are present and all of the item loadings on the general factor are equal, the worst split half will result when all items that load on the same group factor are put into the same half. Therefore, each covariance between items in different halves for the worst split half will be given by (17) and $\beta = k^2 \bar{\lambda}_g^2 / \text{Var}(X)$ which is algebraically equivalent to ω_h .

5.3. Case III: Single Factor, Unequal Loadings

For a unidimensional scale with unequal factor loadings, (6) reduces to

$$\rho = \frac{\mathbf{1}'\mathbf{c}\mathbf{c}'\mathbf{1} + \mathbf{1}'\mathbf{D}\mathbf{D}'\mathbf{1}}{\text{Var}(X)}, \quad (19)$$

(8) reduces to

$$\omega = \frac{\mathbf{1}'\mathbf{c}\mathbf{c}'\mathbf{1}}{\text{Var}(X)} = \omega_h, \quad (20)$$

(9) reduces to

$$\rho - \omega_h = \frac{\mathbf{1}'\mathbf{D}\mathbf{D}'\mathbf{1}}{\text{Var}(X)}, \quad (21)$$

(10) reduces to

$$\alpha = \frac{k}{k-1} \left(\frac{\mathbf{1}'\mathbf{c}\mathbf{c}'\mathbf{1} - \mathbf{c}'\mathbf{c}}{\text{Var}(X)} \right), \quad (22)$$

(11) reduces to

$$\rho - \alpha = \frac{\mathbf{1}'\mathbf{D}\mathbf{D}'\mathbf{1}}{\text{Var}(X)} + \frac{k^2}{(k-1)\text{Var}(X)} \left\{ \frac{\mathbf{c}'\mathbf{c}}{k} - \frac{(\mathbf{1}'\mathbf{c})^2}{k^2} \right\}, \quad (23)$$

(12) reduces to

$$\omega - \alpha = \frac{k^2}{(k-1)\text{Var}(X)} \left\{ \frac{\mathbf{c}'\mathbf{c}}{k} - \frac{(\mathbf{1}'\mathbf{c})^2}{k^2} \right\}, \quad (24)$$

and (13) reduces to

$$\omega_h - \alpha = \frac{k^2}{(k - 1)\text{Var}(X)} \left\{ \frac{\mathbf{c}'\mathbf{c}}{k} - \frac{(\mathbf{1}'\mathbf{c})^2}{k^2} \right\}. \tag{25}$$

That is, (20) shows that if one excludes specific factors from one’s definition of true scores as in (7) then ω_h gives the reliability for a unidimensional scale with unequal factor loadings. Equation (21) shows that ω_h is less than or equal to reliability if one includes item specific factors in one’s definition of true score as in (5) with equality holding if and only if there are no item specific factors. On the other hand, (23) and (24) reveal that α is less than reliability for a unidimensional scale with unequal factor loadings regardless of whether one includes item specific factors in one’s definition of true score or not. Equations (23) and (24) also show that the extent to which α is less than reliability increases as the variability in the factor loadings increases. Finally, (25) shows that $\omega_h > \alpha$ in the case of a unidimensional scale with unequal factor loadings and that the extent to which α is less than ω_h increases as the variability in the factor loadings increases.⁶

$\beta < \alpha$ in the case of a unidimensional scale with unequal loadings. In this case, the worst split will result when the item (or items) that has the smallest loading is put into one half and the remaining items are put in the other. Thus, the average covariance between items in different halves for the worst split will simply be the average of the covariances between the worst item(s) and all the remaining items and will be lower than the average covariance among all the items. Obviously, $\beta < \omega_h$ under these conditions as well.

5.4. Case IV: Single Factor, Equal Loadings (Essential Tau Equivalence)

In the most restrictive case of an essentially tau equivalent scale, that is a unidimensional scale with equal factor loadings, (6) reduces to (19), (8) reduces to (20) and (9) reduces to (21) as in Case III (single factor, unequal loadings). In addition, (10) reduces to $\alpha = \mathbf{1}'\mathbf{c}\mathbf{c}'\mathbf{1}/\text{Var}(X) = \omega_h$ as originally proved by McDonald (1970, 1999) for the case of a unidimensional scale with equal factor loadings. It is also easy to show that $\beta = \omega_h$ in this case as well. When all of the item loadings on the single, general factor are equal ($\lambda_{gi} = \lambda_{gj} = \dots = \lambda_{gk} = \bar{\lambda}_g$), the covariance of any two items, i and j ,

$$\sigma_{ij} = \lambda_{gi}\lambda_{gj} = \bar{\lambda}_g^2. \tag{26}$$

Note that regardless of how the items are split, including for the worst split, each covariance between items in different halves will be given by (26). That is, $\beta = k^2\bar{\lambda}_g^2/\text{Var}(X)$ which is equivalent to ω_h as noted above.

6. Limitations

The primary limitations of the approach taken in this article is that all of its developments are at the population level. There are several methods for estimating general factor loadings based on sample data and questions such as which of these methods produce the least biased statistical estimates of ω_h or the extent to which estimates of ω_h are affected by model misspecification such as over-factoring and under-factoring. These are important ones that we are currently working on. Relatedly, as we limited ourselves to considerations at the population level, we do not include any consideration of the important topic of using sample statistics to derive interval estimates for

⁶McDonald (1999, p. 93) published a proof that also concluded that alpha will be less than ω_h in the single factor, unequal loadings case and that the difference between alpha and ω_h under these conditions is directly related to the variance in the factor loadings. Unfortunately, there was a minor typographical error in that proof.

the population value of ω_h or ω . Though further discussion of this topic is beyond the scope of this article, we would be remiss if we did not at least state that when estimating ω_h or ω , good statistical practice would dictate that point estimation should be supplemented by interval estimation. Bootstrapping and analytic techniques for interval estimation of ω have been described by Raykov (1998a,b, 2002) and Raykov and Shrout (2002) and we suspect these methods should also be useful for obtaining a standard error and confidence interval when estimating ω_h . Certainly, interval estimation of ω_h is a topic that should be studied in detail in future investigations.

7. Conclusions

Equation (10) and its reduced forms in Cases I, II and III show that α is multiply determined under all but the most highly restrictive of conditions — it often reflects not only general factor saturation but also group factor saturation and even variability in factor loadings. Given such complexity, one could question whether the meaning of α is terribly clear in many conditions. Certainly, the nonequivalence of α , β , ω_h , ω and ρ demonstrated here suggests that information about an important psychometric property of a scale — the proportion of variance in scale scores associated with a general factor — may be missing when scale developers and users only report α as is almost always the case. Though ω will always be at least as large as ω_h , and α will often be at least as large as ω_h , choosing among these indices should be based on one's research question and considerations of dimensionality and equality of general factor loadings rather than which index is largest. When interested in the proportion of scale variance due to all common factors, ω is clearly the index of choice and there are conditions (i.e., Case II structures) in which α will be superior to ω_h . In contrast, when interested in the proportion of scale variance due to a general factor, ω_h is clearly the index of choice and is always at least as appropriate as α with equality holding under only the most highly restrictive conditions (i.e., Case IV structures/essential tau equivalence).

References

- Blalock, H. M. (1970). Estimating measurement error using multiple indicators and several points in time. *American Sociological Review*, 35, 101–111.
- Bollen, K. (1989). *Structural Equations With Latent Variables*. New York: Wiley & Sons.
- Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110, 305–314.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- 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.
- Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, 37, 827–838.
- Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review*, 34, 93–101.
- Joreskog, 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.
- Komaroff, E. (1997). Effect of simultaneous violations of essential tau—equivalence and uncorrelated error on coefficient alpha. *Applied Psychological Measurement*, 21, 337–348.
- Lord, F. M. (1955). Sampling fluctuations resulting from the sampling of test items. *Psychometrika*, 20, 1–22.
- Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores* (with contributions by A. Birnbaum). Reading, MA: Addison-Wesley.
- McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis and alpha factor analysis. *British Journal of Mathematical Psychology*, 23, 1–21.
- McDonald, R. P. (1978). Generalizability in factorable domains: "Domain validity and generalizability." *Educational and Psychological Measurement*, 38, 75–79.
- McDonald, R. P. (1985). *Factor Analysis and Related Methods*. Hillsdale, NJ: Erlbaum.
- McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. Mahwah, NJ: Erlbaum.
- Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32, 1–13.

- 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. (2001a). Bias of coefficient alpha for fixed congeneric measures with correlated errors. *Applied Psychological Measurement*, *25*, 69–76.
- Raykov, T. (2001b). Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology*, *54*, 315–323.
- Raykov, T. (2002). Analytic estimation of standard error and confidence interval for scale reliability. *Multivariate Behavioral Research*, *37*, 89–103.
- 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.
- 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.
- Wiley, D. E., & Wiley, J. A. (1970). The estimation of measurement error in panel data. *American Sociological Review*, *35*, 112–117.
- Tucker, L. (1940). The role of correlated factors in factor analysis. *Psychometrika*, *5*, 141–152.
- Wherry, R. J. (1959). Hierarchical factor solutions without rotation. *Psychometrika*, *24*, 45–51.

Manuscript received 18 FEB 2002

Final version received 12 NOV 2003