EBSCO HOST Research Databases
Sign In to My EBSCOhost
New Search | Folder | Preferences | Help
Basic Search | Advanced Search | Visual Search | Choose Databases
Keyword | Thesaurus | Cited References | Indexes
PIONEER
Language

Back Print E-mail Save Add to folder
View: Citation HTML Full Text

Englisch/Deutsch Translate

Title: *Uses and abuses of coefficient alpha.*, By: Schmitt, N., Psychological Assessment, 1996, Vol. 8
Database: *PsycINFO References*

Uses and Abuses of Coefficient Alpha

Contents

Alpha Is Not a Measure of Unidimensionality
What Is an Adequate Level of Alpha?
References
Acknowledgment:
Corrections for Unreliability and Multidimensionality
Presenting Alpha Information Is Not Enough
Summary and Conclusions

By: Neal Schmitt
Department of Psychology, Michigan State University

Acknowledgment:

An earlier version of this article was presented at the 103rd Annual Convention of the American Psychological Association, New York, August 1995.

Correspondence concerning this article should be addressed to: Neal Schmitt, Department of Psychology, Michigan State University, East Lansing, Michigan, 48824.

Presentation of coefficient alpha (hereinafter *alpha;*Cronbach, 1951) as an index of the internal consistency or reliability of psychological measures has become routine practice in virtually all psychological and social science research in which multiple-item measures of a construct are used. In this article I describe four ways in which researchers' use of alpha to convey information about the operationalization of a construct or constructs can represent a lack of understanding or can convey less information than is actually required to evaluate the degree to which measurement problems are or are not a concern in the interpretation of the research results. In each instance, I will also indicate which additional or supplementary information is necessary to evaluate the measurements used in the research.

Alpha Is Not a Measure of Unidimensionality

One important confusion in the literature involves the use of homogeneity and internal consistency as though they were synonymous. *Internal consistency* refers to the interrelatedness of a set of items, whereas *homogeneity* refers to the unidimensionality of the set of items. Internal consistency is certainly necessary for homogeneity, but it is not sufficient. The most recent explication and discussion of this distinction is that of Cortina (1993). Hattie (1985) made a similar distinction in a comprehensive review of alternative ways in which researchers have indexed unidimensionality. Cronbach (1951) viewed reliability, including internal consistency measures, as the proportion of test variance that was attributable to group and general factors. Specific item variance, or uniqueness, was considered error. Clearly, Cronbach, Cortina, and Hattie would not treat alpha as a measure of unidimensionality. In fact, Cronbach stated that alpha is an underestimate of reliability (as he defined it) unless the interitem correlation matrix is of unit rank (i.e., unidimensional). Cronbach's early statements (1947 , 1951) about reliability suggest that the reliability of a multidimensional measure can only be estimated by correlating scores on parallel forms of a test that each represent the same factor structure.

It is also the case that alpha increases as a function of test length. The widely used Spearman-Brown correction formula expresses the relationship between test length and reliability. Lord and Novick (1968) , among others, have provided a discussion of the Spearman-Brown along with tabular illustrations of the relationship between test length and reliability. In fact, the Kuder-Richardson derivations of various reliability formulas that are specific forms of alpha involve the use of the Spearman-Brown correction of a single item's reliability. The single item reliability expressed as the average intercorrelation among items in a measure is extended to express the full-length test reliability in the Kuder-Richardson derivation. The Spearman-Brown formula in this instance would be equal to [*N* times the single item reliability]/[1 + (*N* – 1) times the single item reliability].

Given that alpha is a function of the interrelatedness of the items in a test and the test length rather than the homogeneity of the interitem correlations or their unidimensionality (as is often assumed), what are the measurement implications? Consider the two interitem matrices depicted in Table 1. In the case of both of these six-item matrices, coefficient alpha (actually standardized alpha) is .86, but it is clear that the interrelationships among the first set of items indicate that the responses to the items are a function of two factors. Removal of a single general factor from the second set of items would yield zero off-diagonal correlations, indicating no item-specific and no group factors were responsible for item responses; hence this second six-item measure is unidimensional. This would not be true of the first set of items; the intercorrelations of these items indicate the presence of two factors.

Table 1
Sample Interitem Matrices With Equal Cronbach Alpha

Variable	1	2	3	4	5	6
1.	—					
2.	.8	—				
3.	.8	.8	—			
4.	.3	.3	.3	—		
5.	.3	.3	.3	.8	—	
6.	.3	.3	.3	.8	.8	—
	(α = .86)					

Variable	1	2	3	4	5	6
1.	—					
2.	.5	—				
3.	.5	.5	—			
4.	.5	.5	.5	—		
5.	.5	.5	.5	.5	—	
6.	.5	.5	.5	.5	.5	—
	(α = .86)					

Note. All examples are written in correlational form as opposed to covariance form for convenience and ease of interpretation only.

Sample Interitem Matrices With Equal Cronbach Alpha

Consider the second example in Table 2. In this case, a 6-item measure and a 10-item measure have the same alpha, but the shorter measure clearly is a function of two factors. The 10-item measure in this example is a function of a single factor. Both of these comparisons clearly indicate that alpha is not a good indicator of the unidimensionality of a set of items. In summary, if alpha is used as "proof" that a set of items have an unambiguous or unidimensional interpretation, the conclusions drawn may or may not be correct.

Table 2
Tests of Different Length and Dimensionality With Equal Alpha

Variable	1	2	3	4	5	6
1.	—					
2.	.6	—				
3.	.6	.6	—			
4.	.3	.3	.3	—		
5.	.3	.3	.3	.6	—	
6.	.3	.3	.3	.6	.6	—
	(α = .81)					

Variable	1	2	3	4	5	6	7	8	9	10
1.	—									
2.	.3	—								
3.	.3	.3	—							
4.	.3	.3	.3	—						
5.	.3	.3	.3	.3	—					
6.	.3	.3	.3	.3	.3	—				
7.	.3	.3	.3	.3	.3	.3	—			
8.	.3	.3	.3	.3	.3	.3	.3	—		
9.	.3	.3	.3	.3	.3	.3	.3	.3	—	
10.	.3	.3	.3	.3	.3	.3	.3	.3	.3	—
	(α = .81)									

Tests of Different Length and Dimensionality With Equal Alpha

There are several alternatives to this use of alpha (Hattie, 1985 , actually discussed 30). Cortina (1993) suggested that in addition to reporting alpha, researchers also report what Cortina called the *precision of alpha* or what he called the *standard error of alpha.* This statistic reflects the spread of interitem correlations. This index will yield a value of 0 when all interitem correlations are zero and relatively high values when the spread of interitem correlations is great. A large spread in interitem correlations indicates *either* some form of multidimensionality or a great deal of sampling error in the estimation of the interitem correlations. Cortina's index is not the standard error of alpha; the absence of sample size in his formula means sampling error does not necessarily influence this index. Given certain distributional assumptions, Feldt (1980) and Feldt, Woodruff, and Salih (1987) presented a formula for the computation of the standard error of alpha.

If concerned with sampling error, researchers should use the Feldt (1980) index when they want to assess the accuracy of their estimate of alpha. By contrast, when assessing the degree to which a measure is actually unidimensional, an increasingly popular approach in determining the extent of unidimensionality is to test whether the interitem correlation matrix fits a single-factor model (Jöreskog & Sörbom, 1979). For example, the second examples in both Tables 2 and 3 are perfectly fit by a single-factor model. In Table 2

, a single factor model fit the first matrix of interitem correlations poorly as indexed by a significant chi-square, but more important, by uniformly poor fit statistics as computed in *LISREL8* (Jöreskog & Sörbom, 1993). In this instance, a two-factor model fit the data perfectly. The same was true for the first example in Table 3 , but in this instance, the fit of a single-factor model was not as bad: normed nonfit index (NNFI) = .54; adjusted goodness-of-fit index (AGFI) = .45; and root mean square residual (RMSR) = .13. For readers interested in the assessment of unidimensionality, the relationship between classical test theory perspectives and structural equation modeling of measurement models has been very effectively and clearly illustrated and explained by Miller (1995) .

Table 3
Intercorrelations, Alpha Coefficients, and Corrected Correlations for Two Multidimensional Tests That Are Perfectly Parallel

	Test A						Test B					
Variable	1	2	3	4	5	6	7	8	9	10	11	12
1.	—											
2.	.8	—										
3.	.8	.8	—									
4.	.3	.3	.3	—								
5.	.3	.3	.3	.8	—							
6.	.3	.3	.3	.8	.8	—						
7.	.8	.8	.8	.3	.3	.3	—					
8.	.8	.8	.8	.3	.3	.3	.8	—				
9.	.8	.8	.8	.3	.3	.3	.8	.8	—			
10.	.3	.3	.3	.8	.8	.8	.3	.3	.3	—		
11.	.3	.3	.3	.8	.8	.8	.3	.3	.3	.8	—	
12.	.3	.3	.3	.8	.8	.8	.3	.3	.3	.8	.8	—

Note. Observed Correlation$_{AB}$ = 19.8/(4.58 × 4.58) = .94. Corrected Correlation$_{AB}$ = .94/($\sqrt{.86}$ × $\sqrt{.86}$) = 1.09.

Intercorrelations, Alpha Coefficients, and Corrected Correlations for Two Multidimensional Tests That Are Perfectly Parallel

In examining the types of matrices computed from actual assessee responses, there are rarely instances in which the factorial nature of the interitem correlations is as clear as in these two examples. This also implies that unidimensionality is not unambiguously present or absent. The question can be reframed as Hattie (1985) suggested: “Are there decision criteria that determine how close a set of items is to being a unidimensional set?” (p. 159). It is clear that alpha is *not* an adequate index of unidimensionality, and to interpret or use it for this purpose is wrong. It is also an underestimate of reliability (as defined by Cronbach, i.e., as a measure of the communalities of the items) in the presence of multidimensionality. The latter statement has implications for the use of alpha in corrections for attenuation, which are elaborated on next.

What Is an Adequate Level of Alpha?

A second problem in the use of alpha arises from researchers' common presumption that a particular level of alpha (usually .70) is desired or adequate. Having obtained that level, they then proceed to use the measure without further consideration of its dimensionality or construct validity. This use of the statistic clearly represents a lack of appreciation of the meaning of alpha as discussed earlier and of the relationship between alpha and test length. There are two reasons why the use of any cutoff value (including .70) is shortsighted.

First, alpha is often used to make corrections for unreliability between two measures in an attempt to ascertain the relationship between the latent or true variables underlying the measures. This correction involves dividing the observed correlation between the two variables by the product of the square root of their reliabilities (Lord & Novick, 1968). Classic reliability theory also holds that the upper limit of validity (the relationship between a predictor and criterion) is the square root of the reliability of the criterion or

outcome variables rather than 1.00, which is the upper limit of a Pearson correlation. The concern then is that the true correlations involving a predictor and an unreliable outcome variable will be seriously attenuated (i.e., underestimated) because of inadequate criterion reliability rather than any lack of real or true relationship. In considering the implications of these findings for expected validity, it can be seen that with reliability equal to .70, validity has an upper limit of .84 (i.e., the square root of .70) as opposed to 1.00. Even with reliability as low as .49, the upper limit of validity is .70. When a measure has other desirable properties, such as meaningful content coverage of some domain and reasonable unidimensionality, this low reliability may not be a major impediment to its use. Of course, the usual correction for attenuation would allow the size of the relationship between the underlying constructs to be determined, and it would also allow for clearer correlations between this variable and other potential target variables of interest.

Researchers who do appreciate the relationship between test length and reliability sometimes attempt to excuse the low reliability of their measures by referencing the short length of the measure. The gist of this argument is typically that because the test is short, a low level of alpha would be expected and therefore the researchers should be allowed to use and interpret the findings of research using this measure of low reliability. In these instances, the researchers may or may not be correct in concluding that the low reliability of the measure is a function of test length. However, it remains true that the measures have low reliability, and estimates of the relationships between the variables and other variables will be correspondingly attenuated. Further, interpretations of these relationships should include caveats about low reliability and the potential for underestimating any relationships between the measured variable and other variables of interest. In this instance, if lack of reliability is deemed to be a significant problem in estimating effect sizes or evaluating hypotheses, the researcher should develop a longer measure with adequate reliability. Short length does not alleviate the problems of reliability.

Corrections for Unreliability and Multidimensionality

As previously demonstrated, a relatively high level of alpha can be obtained when the item responses are in fact the function of more than one construct; in these instances alpha is likely to be an underestimate of the measure's reliability as defined by Cronbach. What are the implications of these findings for the appropriateness of the correction for attenuation for unreliability when the correlation being corrected is an estimate of the relationship between two multidimensional measures? The correction for attenuation due to unreliability is computed to provide accurate estimates of the “true” relationship between constructs. Observed correlations are always distorted by any random measurement error in either of two measures correlated. The correction for attenuation serves to provide estimates of the relationships between the underlying constructs measured. The importance of this correction and the implications for research in many different areas of psychology have recently been discussed by Schmidt and Hunter (1996) .

The short answer to this question seems to be that when the factor structure of two multidimensional measures is the same, the correction for attenuation will be an overcorrection. Applying the classic correction for attenuation using alpha as an estimate of reliability in such cases will result in an overestimate of the true correlation between these two variables.

An illustration of this phenomenon is shown in Table 3. This is a case in which two tests (A and B) are composed of identical factors and the observed intercorrelation between the two measures is .94, as calculated from the matrix of correlations presented in Table 3. (Correlation equals the obtained cross scale correlations divided by the product of the standard deviations of the two measures.) This correlation corrected for attenuation (i.e., .94 divided by the product of the square roots of the two reliabilities) is 1.09. Obviously this is an overestimate of the true correlation of 1.00. This demonstration implies that one should not correct for attenuation using an alpha coefficient as the reliability estimate unless there is also evidence that the measures involved are unidimensional.

The practical implications of this demonstration (i.e., whether the correction as it is often applied in research is affected) can only be speculated. In some applied situations such as academic and job performance prediction situations, the practice may make a practical difference in results and interpretations. In most such instances an effort is made to construct measures of outcome variables that

reflect the dimensionality of the job or academic pursuit. These instruments should be appropriately multidimensional. If researchers were to compute a combined score across items or dimensions of this outcome measure, and then compute the validity of the predictor and use a composite alpha to correct this validity for attenuation, the resultant correction would be an overestimate of the "true" validity. In this instance, the researcher would be better advised to develop unidimensional measures of each predictor and criterion construct and then correct observed correlations using estimates of the reliability of these unidimensional measures.

Presenting Alpha Information Is Not Enough

Researchers fairly routinely report the level of alpha associated with the various measures they use in operationalizing key constructs. However, the intercorrelations among the measures are often not presented. This is particularly troublesome if it is important to the researchers' objectives that the measures possess some degree of discriminant validity. Perhaps the worst form of this problematic reporting occurs when a researcher derives measures of several constructs from a single paper-and-pencil measure or interview instrument and reports that the alphas of all measures were relatively high (e.g., above .85). The researcher then proceeds to make interpretations based on the profile of respondents' scores on these dimensions without presenting the intercorrelations among the scales. Or, these measures may be used in some multivariate analysis and the researcher then reports surprise at finding that multicolinearity renders any interpretation regarding the relative efficacy of the variables ambiguous. The minimum information that should be provided in these instances includes the alpha coefficients, the observed correlations, and the correlations corrected for attenuation due to unreliability. This can all be done efficiently with no additional use of space (for those who have been pressured by editors to use space sparingly). An example is presented in Table 4 .

Table 4
Observed Correlations, Alpha Coefficients, and Corrected Correlations Among Measures of Several Constructs

Variable	1	2	3	4	5
1.	(.81)	.80	.71	.83	.59
2.	.70	(.95)	.62	.81	.60
3.	.51	.47	(.64)	.86	.65
4.	.52	.55	.45	(.49)	.67
5.	.32	.35	.31	.28	(.36)

Note. Alpha coefficients are presented on the diagonal, observed correlations below the diagonal, and correlations corrected for attenuation above the diagonal.

Observed Correlations, Alpha Coefficients, and Corrected Correlations Among Measures of Several Constructs

The example in Table 4 can also be used to demonstrate why both corrected and uncorrected coefficients (or the information allowing their calculation) should be presented. First, without the intercorrelations of these variables, the reader does not have the information to evaluate whether the levels of reported alpha are good or bad. Second, the correlation between any two variables might suggest that they are so highly correlated that any differentiation between these two measures is not practically or theoretically useful. In Table 4 , the observed correlation between Variables 1 and 2 indicates they are less discriminable than are measures of the other constructs. However, when both the intercorrelations and the reliabilities of the measures are taken into account (or the corrected correlations are examined), it is clear that these conclusions about Variables 1 and 2 are incorrect. They are no more or less discriminable than Variables 2, 3, or 4. One might also conclude by examining observed correlations that Variable 5 shares little in common with the other four variables, but the corrected coefficients clearly contradict this view.

Other examples could be constructed to show other combinations of reliability and intercorrelations that would be very differently interpreted when relying on observed correlations rather than corrected

correlations. Clearly, both intercorrelations and alpha must be reported if the reader is to be adequately informed about the obtained results. Of course, it is also incumbent on the researcher to consider both sources of information when drawing conclusions about the adequacy of measures.

Summary and Conclusions

Four caveats are implied by this article regarding the proper use of the alpha coefficient.

1. Alpha is not an appropriate index of unidimensionality to assess homogeneity.
2. In correcting for attenuation due to unreliability, use of alpha as an estimate of reliability is based on the notion that the measures involved are unidimensional. When this is not the case, the corrected coefficients will be overcorrected.
3. There is no sacred level of acceptable or unacceptable level of alpha. In some cases, measures with (by conventional standards) low levels of alpha may still be quite useful.
4. Presenting only alpha when discussing the relationships of multiple measures is not sufficient. Intercorrelations and corrected intercorrelations must be presented as well.

References

1. Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98-104.

2. Cronbach, L. J. (1947). Test “reliability”: Its meaning and determination. *Psychometrika*, 12, 1-16.

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

4. Feldt, L. S. (1980). A test of the hypothesis that Cronbach's alpha coefficient is the same for two tests administered to the same sample. *Psychometrika*, 45, 99-105.

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

6. Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9, 139-164.

7. Jöreskog, K. G. & Sörbom, D. (1979). Advances in factor analysis and structural equation models. Cambridge, MA: Abt Books.

8. Jöreskog, K. G. & Sörbom, D. (1993). New features in LISREL8. Chicago, Illinois: Scientific Software International.

9. Lord, F. M. & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

10. Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modelling. *Structural Equation Modelling*, 2, 255-273.

11. Schmidt, F. L. & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, 1, 199-223.

Received: May 31, 1996 *Revised:* July 24, 1996 *Accepted:* July 24, 1996

Copyright 1996 American Psychological Association

This publication is protected by US and international copyright laws and its content may not be copied without the copyright holder's express written permission except for the print or download capabilities of the retrieval software used for access. This content is intended solely for the use of the individual user.
Source: Psychological Assessment. Vol. 8 (4) 1996, pp. 350-353
Accession Number: pas814350 **Digital Object Identifier:** 10.1037/1040-3590.8.4.350

EBSCO Support Site

Privacy Policy | Terms of Use | Copyright

© 2006 EBSCO Industries, Inc. All rights reserved.