Cronbach's α for Reliability

From Carmines, E. G., & Zeller, R. A (1979). Reliability and validity assessment. Newbury Park, CA: Sage.

Internal Consistency Method

We noted above that an important limitation of the split-halves method of assessing reliability is that reliability coefficients obtained from different ways of subdividing the total set of items would not be the same. For example, it is quite possible that the correlation between the first and second halves of the test would be different from the correlation between odd and even items. However, there are methods for estimating reliability that do not require either the splitting or repeating of items. Instead, these techniques require only a single test administration and provide a unique estimate of reliability for the given test administration. As a group, these coefficients are referred to as measures of internal consistency. By far the most popular of these reliability estimates is given by Cronbach's alpha (Cronbach, 1951), which can be expressed as follows:

$$\alpha = N/(N - 1)\,[1 - \Sigma\sigma^2(Y_i)/\sigma_X^2] \qquad [19]$$

where N is equal to the number of items; $\Sigma\sigma^2(Y_i)$ is equal to the sum of item variances; and σ_X^2 is equal to the variance of the total composite. If one is working with the correlation matrix rather than the variance-covariance matrix, then alpha reduces to the following expression:

$$\alpha = N\bar{\rho}/[1 + \bar{\rho}(N - 1)] \qquad [20]$$

where N is again equal to the number of items and $\bar{\rho}$ is equal to the mean interitem correlation. To take a hypothetical example applying Equation 20, if the average intercorrelation of a six-item scale is .5, then the alpha for the scale would be:

$$\alpha = 6(.5)/[1 + .5(6 - 1)]$$

$$= 3 / 3.5$$

$$= .857.$$

To give an example of how alpha is calculated, consider the 10-item self-esteem scale developed by Rosenberg (1965). The intercorrelations among the items for a sample of adolescents are presented in Table 3 (for further discussion of these data see the appendix). To find the mean interitem correlation we first sum the

45 correlations in Table 3: .185 + .451 + .048 + . . . + .233 = 14.487. Then we divide this sum by 45: 14.487/45 = .32. Now we use this mean interitem correlation of .32 to calculate alpha as follows:

$$\alpha = 10(.32)/[1 + .32(10 - 1)]$$

$$= 3.20 / 3.88$$

$$= .802.$$

From Equation 20 it is not difficult to see that alpha varies between .00 and 1.00, taking on these limits when the average interitem correlations are zero and unity, respectively.

The interpretation of Cronbach's alpha is closely related to that given for reliability estimates based on the split-halves method. Specifically, coefficient alpha for a test having 2N items is equal to the average value of the alpha coefficients obtained for all possible combinations of items into two half-tests (Novick and Lewis, 1967). Alternatively, alpha can be considered a unique estimate of the expected correlation of one test with an alternative form containing the same number of items. Nunnally (1978) has demonstrated that coefficient alpha can also be derived as the expected correlation between an actual test and a *hypothetical* alternative form of the same length, one that may never be constructed.

Novick and Lewis (1967) have proven that, in general, alpha is a *lower bound* to the reliability of an unweighted scale of N items, that is, $\rho_x \geqslant \alpha$. It is equal to the reliability if the items are parallel. Thus, the reliability of a scale can never be lower than alpha even if the items depart substantially from being parallel measurements. In other words, in most situations, alpha provides a conservative estimate of a measure's reliability.

Equation 20 also makes clear that the value of alpha depends on the average interitem correlation and the number of items in the scale. Specifically, as the average correlation among items increases and as the number of items increases, the value of alpha increases. This can be seen by examining Table 1 which shows the value of alpha given a range in the number of items from 2 to 10 and a range in the average interitem correlation from .0 to 1.0. For example,

TABLE 1

Values of Cronbach's Alpha for Various Combinations of Different Number of Items and Different Average Interitem Correlations

Number of Items	Average Interitem Correlation .0	.2	.4	.6	.8	1.0
2	.000	.333	.572	.750	.889	1.000
4	.000	.500	.727	.857	.941	1.000
6	.000	.600	.800	.900	.960	1.000
8	.000	.666	.842	.924	.970	1.000
10	.000	.714	.870	.938	.976	1.000

a 2-item scale with an average iteritem correlation of .2 has an alpha of .333. However, a 10-item scale with the same average interitem correlation has an alpha of .714. Similarly, an 8-item scale with an average interitem correlation of .2 has an alpha of .666 whereas if the 8 items had an average intercorrelation of .8, then the scale's alpha would be .970. In sum, the addition of more items to a scale that do *not* result in a reduction in the average interitem correlation will increase the reliability of one's measuring instrument.

While increasing the number of items in a scale can thus improve the scale's reliability, there are significant limitations to this procedure. First, the adding of items indefinitely makes progressively less impact on the reliability. Thus, given an average interitem correlation of .4, increasing the number of items from 2 to 4 increases the alpha for the scale by .155 (i.e., .727 - .572 = .155). However, increasing the number of items from 8 to 10 with the same average interitem correlation only increases the alpha by .028 (i.e., .870 - .842 = .028). Second, the greater the number of items in a scale, the more time and resources are spent constructing the instrument. It should be noted, finally, that adding items to a scale can, in some instances, *reduce* the lengthened scale's reliability if the additional items substantially lower the average interitem correlation.

Alpha is more difficult to compute than coefficients based on other methods of assessing reliability. In the retest, alternative-form, and split-halves methods, it is only necessary to calculate a single

correlation to obtain the desired reliability estimate. Specifically, in the retest method, scores for the same group of people on the same test administered on two occasions are correlated; in the alternative-forms approach, scores on different versions of the same test are correlated; and in the split-halves method, the items are divided into arbitrary halves and scores between the half-tests are correlated. In contrast, as we have seen, alpha depends on the average intercorrelation among all of the items. Yet, it is important to realize that although more complex computationally, alpha has the same logical status as coefficients arising from the other methods of assessing reliability. This is easy to see once we consider some additional properties of parallel measurements. In addition to having equal true scores and equal error variances, parallel measurements are assumed to have the following useful properties:

(1) The expected (mean) values of parallel measures are equal: $E(X) = E(X')$.
(2) The observed score variance of parallel measures is equal: $\sigma_x^2 = \sigma_{x'}^2$.
(3) The intercorrelations among parallel measurements are equal from pair to pair: $\rho_{xx'} = \rho_{xx''} = \rho_{x'x''}$.
(4) The correlations of parallel measures with other variables are equal: $\rho_{xy} = \rho_{x'y} = \rho_{x''y}$.

These properties imply that there are no systematic differences between parallel measurements; instead, they only differ from another because of strictly random error, and thus, for essential purposes, are completely interchangeable. Moreover, since parallel measurements have equal intercorrelations, the average interitem correlation is simply equal to the correlation between any arbitrary pair of items. In other words, *if the items are truly parallel*, the average interitem correlation accurately estimates all of the correlations in the item matrix. Thus, *logically*, using the average correlation in the calculation of alpha amounts to exactly the same thing as calculating a simple correlation between parallel measurements.

KR20

Cronbach's alpha is a generalization of a coefficient introduced by Kuder and Richardson (1937) to estimate the reliability of scales composed of dichotomously—scored items. Dichotomous items are scored one or zero depending on whether the respondent does or does not possess the particular characteristic under investigation. Thus, for the items making up a spelling test, a score of 1 would be given when the students spelled a particular word correctly but zero if the word is spelled incorrectly. To determine the reliability of scales composed of dichotomously scored items, one uses the following Kuder-Richárdson formula number 20 (symbolized KR20):

$$KR20 = N/(N - 1)\,[1 - \Sigma p_i q_i/\sigma_x^2] \qquad [21]$$

where N is the number of dichotomous items; p_i is the proportion responding "positively" to the i^{th} item; q_i is equal to $1 - p_i$; and σ_x^2 is equal to the variance of the total composite. Since KR20 is simply a special case of alpha, it has the same interpretation as alpha: that is, it is an estimate of the expected correlation between one test and a hypothetical alternative form containing the same number of items.

Correction for Attenuation

Whatever particular method is used to obtain an estimate of reliability, one of its important uses is to "correct" correlations for unreliability due to random measurement error. That is, if we can estimate the reliability of each variable, then we can use these estimates to determine what the correlation between the two variables would be if they were made perfectly reliable. The appropriate formula is as follows:

$$\rho_{x_t y_t} = \rho_{x_i y_j} / \sqrt{\rho_{xx'}\,\rho_{yy'}} \qquad [22]$$

where $\rho_{x_t y_t}$ is the correlation corrected for attenuation; $\rho_{x_i y_j}$ is the observed correlation; $\rho_{xx'}$ is the reliability of X; and $\rho_{yy'}$ is the reliability of Y. For example, if the observed correlation between two

variables was .2 and the reliability of each variable was .5, then the correlation corrected for attentuation would be:

$$\rho_{x_t y_t} = .2/\sqrt{(.5)(.5)} = .4.$$

This means that the correlation between these two variables would be .4 if both were perfectly reliable (measured without random error).

Table 2 illustrates the behavior of the correlation coefficient under varying conditions of correction for attenuation. Table 2A shows the value of the correlation corrected for attenuation given that the observed correlation is .3 with varying reliabilities of X and Y. As an example, when the reliabilities of X and Y are .4, respectively, the corrected correlation is .75. When the reliabilities of X and Y are 1.0, respectively, the corrected correlation is equal to the observed correlation of .3. Table 2B presents similar calculations when the observed correlation is .5. Examining sections A and B of Table 2 it is clear that the higher the reliabilities of the variables, the less the corrected correlation differs from the observed correlation.

Table 2C presents the value of the correlation that one will observe when the correlation between X_t and Y_t is .5 under varying conditions of reliability. If the reliabilities of X and Y are .8, respectively, the observed value of a theoretical .5 correlation is .4. Table 2D presents similar calculations when the correlation between X_t and Y_t is .7. For example, even if the theoretical correlation between X_t and Y_t is .7, the observed correlation will be only .14 if the reliabilities are quite low (.2). Thus, one must be careful not to conclude that the theoretical correlations are low simply because their observed counterparts are low; it may instead be the case that the measures are quite unreliable.

Conclusion

This chapter has discussed four methods for assessing the reliability of empirical measurements. For reasons mentioned in the chapter, neither the retest method nor the split-halves approach is

TABLE 2
Examples of Correction for Attenuation

A: $\rho_{x_t y_t} = .3/\sqrt{\rho_{xx'}\rho_{yy'}}$

$\rho_{yy'}$ \ $\rho_{xx'}$	.2	.4	.6	.8	1.0
.2	—	—	.87	.75	.67
.4	—	.75	.61	.53	.47
.6	.87	.61	.50	.43	.39
.8	.75	.53	.43	.38	.33
1.0	.67	.47	.39	.33	.30

B: $\rho_{x_t y_t} = .5/\sqrt{\rho_{xx'}\rho_{yy'}}$

$\rho_{yy'}$ \ $\rho_{xx'}$	.2	.4	.6	.8	1.0
.2	—	—	—	—	—
.4	—	—	—	.88	.79
.6	—	—	.83	.72	.65
.8	—	.88	.72	.63	.56
1.0	—	.79	.65	.56	.50

C: $.5 = \rho_{xy}/\sqrt{\rho_{xx'}\rho_{yy'}}$

$\rho_{yy'}$ \ $\rho_{xx'}$	.2	.4	.6	.8	1.0
.2	.10	.14	.17	.20	.22
.4	.14	.20	.24	.28	.32
.6	.17	.24	.30	.35	.39
.8	.20	.28	.35	.40	.45
1.0	.22	.32	.39	.45	.50

D: $.7 = \rho_{xy}/\sqrt{\rho_{xx'}\rho_{yy'}}$

$\rho_{yy'}$ \ $\rho_{xx'}$	.2	.4	.6	.8	1.0
.2	.14	.20	.24	.28	.31
.4	.20	.28	.34	.40	.44
.6	.24	.34	.42	.48	.54
.8	.28	.40	.48	.56	.63
1.0	.31	.44	.54	.63	.70

recommended for estimating reliability. The major defect of the retest method is that experience in the first testing usually will influence responses in the second testing. The major problem with the split-halves approach is that the correlation between the halves will differ somewhat depending on how the total number of items is divided into halves. As Nunnally argues, "it is best to think of the corrected correlation between any two halves of a test as being an *estimate* of coefficient alpha. Then it is much more sensible to employ coefficient alpha than any split-half method" (1978: 233).

In contrast, the alternative-form method and coefficient alpha provide excellent techniques for assessing reliability. The practical limitation of using the alternative-form method is that it can be

quite difficult to construct alternative forms of a test that are parallel. One recommended way of overcoming this limitation is by randomly dividing a large collection of items in half to form two randomly parallel tests. In sum, if it is possible to have two test administrations, then the correlation between alternative forms of the same test provides a very useful way to assess reliability.

Coefficient alpha should be computed for any multiple-item scale. It is particularly easy to use because it requires only a single test administration. Moreover, it is a very general reliability coefficient, encompassing both the Spearman-Brown prophecy formula as well as the Kuder-Richardson 20. Finally, as we have seen, alpha is easy to compute, especially if one is working with a correlation matrix (for further details on the computation of alpha see Bohrnstedt, 1969). The minimal effort that is required to compute alpha is more than repaid by the substantial information that it conveys about the reliability of a scale. What is a satisfactory level of reliability? Unfortunately, it is difficult to specify a single level that should apply in all situations. As a general rule, we believe that reliabilities should not be below .80 for widely used scales. At that level, correlations are attenuated very little by random measurement error. At the same time, it is often too costly in terms of time and money to try to obtain a higher reliability coefficient. But the most important thing to remember is to report the reliability of the scale and how it was calculated. Then other researchers can determine for themselves whether it is adequate for any particular purpose.