Understanding a Widely Misunderstood Statistic: Cronbach's α

Nicola L. Ritter

Texas A&M University

Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, February 18, 2010

Abstract

It is important to explore score reliability in virtually all studies, because tests are not reliable. The present paper explains the most frequently used reliability estimate, coefficient alpha, so that the coefficient's conceptual underpinnings will be understood. Researchers need to understand score reliability because of the possible impact reliability has on the interpretation of research results. There are several common misconceptions about the basic ideas of score reliability. Misconceptions are formed due to lack of understanding of the concept of reliability and through careless speech involving statistical jargon. This paper addresses common misconceptions so that later discussions over score reliability will not be hindered. Misconceptions have caused some authors to devalue the reporting of reliability estimates in published research, while others report reliability coefficients inappropriately. A better understanding of score reliability can resolve these misconceptions and enable authors to use reliability coefficients appropriately in literature and speech. A background of the basic ideas of score reliability is introduced and concludes with an explanation of the most frequently used reliability estimate, coefficient alpha, so that the coefficient's conceptual underpinnings will be understood.

Understanding a Widely Misunderstood Statistic: Cronbach's α

Researchers often want to evaluate the importance of a study's results by using at least one of the types of significance: statistical significance, practical significance, and clinical significance. As practical significance gains support in publications, researchers will begin to notice the influence that reliability has on effect sizes and statistical power against Type II error. Researchers need to understand score reliability because of the possible impact reliability has on the interpretation of research results. Thompson (1994) warns,

> The failure to consider score reliability in substantive research may exact a toll on the interpretations within research studies. For example, we may conduct studies that could not possibly yield noteworthy effect sizes given that score reliability inherently attenuates effects sizes. Or we may not accurately interpret the effect sizes in our studies if we do not consider the reliability of the scores we are actually analyzing. (p. 840)

There are several common misconceptions about the basic ideas of score reliability. Misconceptions are formed due to lack of understanding of the concept of reliability and through careless speech involving statistical jargon. One should address these misconceptions to prevent misinterpretations of research results. A common misconception is that reliability is a characteristic of a test or a measurement tool; however, reliability instead is a characteristic of scores. Spearman (1904) introduced this characteristic by utilizing a method that measures each individual multiple times. In this method, Spearman determined reliability based on the consistency of the individual's scores across equivalent measurement forms. If consistency is seen across measurement forms, then one can conclude that the scores are reliable. If there is no consistency across measurement forms, then one can conclude the scores are not reliable. The method in which Spearman (1904) applied shows that individual scores were tested, not the measurement tool. As Henson (2001) suggested, "Because scores may vary in degree of reliability, a given test may yield grossly divergent reliability estimates on different administrations" (p. 178).

Another common misconception is that reliability is equivalent to validity. Validity pertains to the extent to which scores measure the intended concept. Reliability determines if the scores measure anything, while validity determines to what extent the scores measure the intended something. The relationship between validity and reliability is analogous to the relationship between effect size and $p_{\text{calculated}}$. For example, if a person repeatedly measured the same two grams of seasoning for a given recipe, consistently producing the same estimate of the seasoning's weight, this may support that the scores are reliable. However, if one implies from the scores, "This recipe tastes great because two grams of seasoning can make anything taste good," then questions of score validity may surface from the individual's dinner guests. Scores must be reliable to even consider if the scores are valid, but reliability does not necessarily imply validity. If scores were not reliable, then one would merely be consistently measuring nothing. Reliability is not equivalent to validity because reliability and validity are two separate properties of scores. These misconceptions are demonstrated through the verbiage in journal articles (Thompson, 1992) and careless jargon used in informal speech (Thompson, 2003). Thompson (2003) exposed these misconceptions to researchers in the hope that through enlightenment researchers will better evaluate scores.

Misconceptions have caused some authors to devalue the reporting of reliability estimates in published research (Vacha-Haase, Henson & Caruso, 2002) while others report reliability coefficients inappropriately (Thompson, 1992, 2003; Wilkinson & APA Task Force on Statistical Inference, 1999). A better understanding of score reliability can resolve these misconceptions and enable authors to use reliability coefficients appropriately in literature and speech. An explanation of the basic idea of score reliability and a focus on the properties of one of the most commonly reported reliability estimate, Cronbach's (1951) alpha (α), will be discussed further.

Background of Reliability

Consistency of Scores

Reliability pertains to the consistency of scores. The less consistency within a given measurement, the less useful the data may be in analysis. For example, a recipe calls for two grams of a seasoning. The package of seasoning states the contents of the package includes two grams of seasoning. To begin cooking, one measures two grams of seasoning that does not seem to use the entire package. Curiosity sets in and the person decides to measure the seasoning again. To the person's surprise, the second measure indicates a score more than two grams of seasoning. Stubbornly, the person measures the seasoning a third time and notices a score less than two grams of seasoning. Baffled at these results, one may begin to question all of the scores produced by the measuring tool. The person concludes the measurement tool does not measure anything. When measurement tools generate random scores, the scores are not reliable. On the other hand, suppose the person measured the seasoning multiple times and generated two grams each time; this set of scores would then be considered reliable.

Types of Reliability

There are several coefficients to estimate the reliability of scores, such as internal consistency, test-retest, and form equivalence coefficients. Each type of coefficient estimates consistency across different parameters. Internal consistency coefficients estimate the degree in which scores measure the same concept. To put this in context of the cooking example, the individual is testing the weight of the seasoning instead of the chemical composition or pH of the seasoning. Test-retest coefficients estimate stability of scores over a period of time. Form equivalence coefficients estimate consistency of scores between two test forms. Internal consistency coefficients are convenient to calculate because such coefficients require only a single measurement given at one time. Internal consistency coefficients are more practical than other reliability coefficients due to the lack of time and resources to perform the multiple tests seen in test-retest coefficients and the multiple formats seen in form equivalence coefficients. There is no preference for a single method. A method should be selected based on the context of the research being conducted.

Properties of Cronbach's Alpha (α)

There are various types of reliability coefficients. Cronbach's (1951) alpha is one of the most commonly used reliability coefficients (Hogan, Benjamin & Brezinksi, 2000) and for this reason the properties of this coefficient will be emphasized here.

Type of Reliability Coefficient

One property of alpha (Cronbach, 1951) is it is one type of internal consistency coefficient. Before alpha, researchers were limited to estimating internal consistency of only dichotomously scored items using the KR-20 formula. Cronbach's (1951) alpha was developed based on the necessity to evaluate items scored in multiple answer categories. Cronbach (1951) derived the alpha formula from the KR-20 formula:

$$\text{KR-20} = \text{K} / (\text{K}-1)\,[1 - \sum p_k q_k / \sigma_{\text{total}}^{2})], \tag{1}$$

where K is the number of items, p_k is the proportion of people with a score of 1 on the kth item, q_k is the proportion of people with a score of 0 on the kth item, and σ_{total}^2 is the variance of scores on the total measurement, to include both dichotomously and polychotomously scored items.

Calculating Alpha (α)

Alpha is calculated using the following formula:

$$\alpha = K / (K - 1) [1 - (\sum \sigma_k^2 / \sigma_{\text{total}}^2)], \quad (2)$$

where K is the number of items, $\sum \sigma_k^2$ is the sum of the k item score variances, and σ_{total}^2 is the variance of scores on the total measurement. By comparing both equations, one can see that the only difference between the two formulas is numerators, $\sum p_k q_k$ and $\sum \sigma_k^2$. The two numerators are computationally equivalent when items are dichotomously scored (Thompson, 2003). One way to calculate alpha, is to use a statistical software program such as SPSS. Select Analyze > Scale > Reliability Analysis. Next, select the scores you wish to analyze. Finally, select paste to paste the below syntax into your syntax file and run.

```
RELIABILITY
/VARIABLES=X1 X2 X3
/SCALE(ALL VARIABLES) ALL
/MODEL=ALPHA
```

Ratio of Variances

Another property of alpha (Cronbach, 1951) is it is a ratio of variances that follows the general linear model (GLM). In the term, $\sum\sigma_k^{\ 2} / \sigma_{\text{total}}^{\ 2}$, there is division of the true score variance by the total score variance. Given that variance is a squared metric statistic, when we divide a squared metric statistic (i.e. $\sum\sigma_k^2$) by a squared metric statistic (i.e. $\sigma_{\text{total}}^{\ 2}$) the result will also be in a squared metric. One misconception about alpha is that alpha can only be positive because alpha is a squared metric statistic. However, computationally, alpha can be negative. When alpha is negative the integrity of the scores should be severely questioned (Thompson, 2003). A negative alpha is a symptom of two differential diagnoses: 1) an incorrect measurement model or 2) very bad scores. Alpha is a direct analog of effect size, r^2, due to the nature of variance-accounted-for effect sizes such as r^2, R^2, and η^2 (Thompson, 2003). Alpha takes into consideration the correlation between item scores. More directly, alpha is the square of the correlation between true score variance and total score variance. The degree of correlation and the direction of the relationship will help explain how a negative alpha can be generated. Consider three possible scenarios: alpha equal to zero, alpha equal to one, and alpha equal to a negative value. These heuristic examples have been adapted from Henson (2001) and Thompson (2003).

Scenario #1: all item score correlations are perfectly uncorrelated. According to the formula for alpha, alpha can be calculated if we have the number of items, the sum of the item variances, and the total score variance. Table 1 provides information on the number of items, $k = 4$ and the sum of the item variances. Using the information in Table 1, the sum of the item variances can be computed as:

$$\sum \sigma_k^2 = 0.24 + 0.22 + 0.21 + 0.15 = \mathbf{0.82}.$$

Crocker and Algina (1986, p. 95) provide a formula to calculate the total score variance using the information found in Table 1:

$$\sigma_{\text{total}}^{2} = \sum \sigma_k^{2} + \left[\sum COV_{ij} \text{ (for } i < j\text{) x } 2\right]. \tag{3}$$

Using the information in Table 2, the total variance can be computed using Equation 3:

$$\begin{aligned}\sigma_{\text{total}}^{2} &= \sum \sigma_{k_k}^{2} + \left[\sum COV_{ij} \text{ (for } i < j\text{) x } 2\right] \\ &= \mathbf{0.82} + [0 \text{ x } 2] \\ &= \mathbf{0.82} + 0 = 0.82\end{aligned}$$

Alpha can then be found using Equation 2:

$$\begin{aligned}\alpha &= \text{K} / (\text{K} - 1)\, [1 - (\sum \sigma_k^2 / \sigma_{total}^{2})] \\ &= 4 / (4 - 1)\, [1 - (0.82 / 0.82)] \\ &= 4 / 3\, [1 - 1] \\ &= 1.33\, [0] \\ &= 0\end{aligned}$$

When items are perfectly uncorrelated, the items share no variance; therefore there is no internal consistency between the item scores. Accordingly, alpha will equal zero when items are perfectly uncorrelated.

Table 1

Covariance and Correlation Matrices for Scenario #1

	Variance / Covariance				Correlation			
Var.	1	2	3	4	1	2	3	4
1	**0.24**				**1.00**			
2	.00	**0.22**			.00	**1.00**		
3	.00	.00	**0.21**		.00	.00	**1.00**	
4	.00	.00	.00	**0.15**	.00	.00	.00	**1.00**

Note. Adapted from *Score reliability: Contemporary thinking on reliability issues* by B. Thompson, 2003, p. 15 and from" Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha," by R. K. Henson, 2001, *Measurement and Evaluation in Counseling and Development*, *34*, 183.

Table 2

Total Score Variance as a Function of Item Variances and Covariances for Scenario #1

Pairs		Covariance / Variance			*r* / *SD*		
i	j	COV_{ij}	VAR_i	VAR_j	r_{ij}	SD_i	SD_j
1	2	**.00**	0.24	0.22	**.00**	0.49	0.47
1	3	**.00**	0.24	0.21	**.00**	0.49	0.46
1	4	**.00**	0.24	0.15	**.00**	0.49	0.39
2	3	**.00**	0.22	0.21	**.00**	0.47	0.46
2	4	**.00**	0.22	0.15	**.00**	0.47	0.39
3	4	**.00**	0.21	0.15	**.00**	0.46	0.39

Note. Adapted from *Score reliability: Contemporary thinking on reliability issues* by B. Thompson, 2003, p. 15 and from" Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha," by R. K. Henson, 2001, *Measurement and Evaluation in Counseling and Development*, *34*, 183.

Scenario #2: all item score correlations are perfectly correlated. Using the information in Table 3, the sum of the item variances can be computed as:

$$\sum \sigma_k^2 = 0.24 + 0.22 + 0.21 + 0.15 = \mathbf{0.82}.$$

Using the information in Table 4, the total variance can be computed using Equation 3:

$$\begin{aligned}\sigma_{\text{total}}^{2} &= \sum \sigma_k^2 + \left[\sum COV_{ij} \text{ (for } i < j) \text{ x } 2\right] \\ &= \mathbf{0.82} \quad + \left[(0.23 + 0.22 + 0.19 + 0.21 + 0.18 + 0.18) \text{ x } 2\right] \\ &= \mathbf{0.82} \quad + \left[1.22 \text{ x } 2\right] = 3.26\end{aligned}$$

Alpha can then be found using Equation 2:

$$\begin{aligned}\alpha &= \text{K} / (\text{K} - 1)\,[1 - (\sum \sigma_k^2 / \sigma_{total}^{2})] \\ &= 4 / (4 - 1)\left[1 - (.82 / 3.26)\right] \\ &= 4 / 3\left[1 - 0.2515\right] \\ &= 1.33\left[0.7485\right] \\ &= .9955\end{aligned}$$

When items are perfectly correlated, there is perfect internal consistency between the item scores.

Accordingly, $\alpha = 1$ (within rounding error) when items are perfectly correlated.

Table 3

Covariance and Correlation Matrices for Scenario #2

	Variance / Covariance				Correlation			
Var.	1	2	3	4	1	2	3	4
1	0.24				1.00			
2	**0.23**	0.22			**1.00**	1.00		
3	**0.22**	**0.21**	0.21		**1.00**	**1.00**	1.00	
4	**0.19**	**0.18**	**0.18**	0.15	**1.00**	**1.00**	**1.00**	1.00

Note. Adapted from *Score reliability: Contemporary thinking on reliability issues* by B. Thompson, 2003, p. 16 and from" Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha," by R. K. Henson, 2001, *Measurement and Evaluation in Counseling and Development*, *34*, 185.

Table 4

Total Score Variance as a Function of Item Variances and Covariances for Scenario #2

Pairs		Covariance / Variance			*r* / *SD*		
i	j	COV_{ij}	VAR_i	VAR_j	r_{ij}	SD_i	SD_j
1	2	**0.23**	0.24	0.22	**1.00**	0.49	0.47
1	3	**0.22**	0.24	0.21	**1.00**	0.49	0.46
1	4	**0.19**	0.24	0.15	**1.00**	0.49	0.39
2	3	**0.21**	0.22	0.21	**1.00**	0.47	0.46
2	4	**0.18**	0.22	0.15	**1.00**	0.47	0.39
3	4	**0.18**	0.21	0.15	**1.00**	0.46	0.39

Note. Adapted from *Score reliability: Contemporary thinking on reliability issues* by B. Thompson, 2003, p. 16 and from" Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha," by R. K. Henson, 2001, *Measurement and Evaluation in Counseling and Development*, *34*, 185.

Scenario #3: all item score correlations are perfectly correlated and have mixed signs.

Using the information in Table 5, the sum of the item variances can be computed as:

$$\sum \sigma_k^2 = 0.24 + 0.22 + 0.21 + 0.15 = \mathbf{0.82}.$$

Using the information in Table 6, the total variance can be computed using Equation 3:

$$\begin{aligned}\sigma_{\text{total}}^{2} &= \sum \sigma_k^2 + \left[\sum COV_{ij} \text{ (for } i < j) \text{ x } 2\right] \\ &= \mathbf{0.82} \quad + \left[(\text{-}0.23 + -0.22 + -0.19 + 0.21 + 0.18 + 0.18) \text{ x } 2\right] \\ &= \mathbf{0.82} \quad + \left[\text{-}.07 \text{ x } 2\right] \\ &= \mathbf{0.82} \quad + \left[\text{-}.14\right] = .68\end{aligned}$$

Alpha can then be found using Equation 2:

$$\begin{aligned}\alpha &= \text{K} / (\text{K} - 1)\,[1 - (\sum \sigma_k^2 / \sigma_{total}{}^2)] \\ &= 4 / (4 - 1)\,[1 - (.82 / .68)] \\ &= 4 / 3\,[1 - 1.2059] \\ &= 1.33\,[-.2059] \\ &= -.2738\end{aligned}$$

When items are perfectly correlated and have mixed signs, the sum of item variances will be greater than the total score variance. When the individual score variance is greater than total score, internal consistency is non-existent between the item scores; therefore the items are measuring different concepts. In general, as items are more correlated, shared variance increases, increasing internal consistency; therefore increasing the magnitude of the alpha coefficient.

Table 5

Covariance and Correlation Matrices for Scenario #3

	Variance / Covariance				Correlation			
Var.	1	2	3	4	1	2	3	4
1	0.24				1.00			
2	**-0.23**	0.22			**-1.00**	1.00		
3	**-0.22**	**0.21**	0.21		**-1.00**	**1.00**	1.00	
4	**-0.19**	**0.18**	**0.18**	0.15	**-1.00**	**1.00**	**1.00**	1.00

Note. Adapted from *Score reliability: Contemporary thinking on reliability issues* by B. Thompson, 2003, p. 17 and from" Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha," by R. K. Henson, 2001, *Measurement and Evaluation in Counseling and Development*, *34*, 185.

Table 6

Total Score Variance as a Function of Item Variances and Covariances for Scenario #3

Pairs		Covariance / Variance			*r* / *SD*		
i	j	COV_{ij}	VAR_i	VAR_j	r_{ij}	SD_i	SD_j
1	2	**-0.23**	0.24	0.22	**-1.00**	0.49	0.47
1	3	**-0.22**	0.24	0.21	**-1.00**	0.49	0.46
1	4	**-0.19**	0.24	0.15	**-1.00**	0.49	0.39
2	3	**0.21**	0.22	0.21	**1.00**	0.47	0.46
2	4	**0.18**	0.22	0.15	**1.00**	0.47	0.39
3	4	**0.18**	0.21	0.15	**1.00**	0.46	0.39

Note. Adapted from *Score reliability: Contemporary thinking on reliability issues* by B. Thompson, 2003, p. 17 and from" Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha," by R. K. Henson, 2001, *Measurement and Evaluation in Counseling and Development*, *34*, 185.

Discussion

Because tests are not reliable, it is important to explore score reliability in virtually all studies. Reliability coefficients have the ability to impact how researchers interpret study results. Researchers should be aware of alpha’s properties to accurately gather data and interpret results. A better understanding of score reliability can resolve common misconceptions and employ authors to write and speak cautiously when referring to reliability estimates. The present paper explains the most frequently used reliability estimate, coefficient alpha, so that the coefficient's conceptual underpinnings will be understood.

References

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Chicago: Holt, Rinehart & Winston.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 197-334.

Henson, R.K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development*, *34*, 177-189.

Hogan, T.P., Benjamin, A., & Brezinksi, K.L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement*, *60*(4), 523-531.

Spearman, C.E. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*, 72-101.

Spearman, C.E. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, *3*, 271-295.

Thompson, B. (1992). Two and one-half decades of leadership in measurement and evaluation. *Journal of Counseling and Development*, *70*, 434-438.

Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement*, *54*, 837-847.

Thompson, B. (Ed.) (2003). *Score reliability: Contemporary thinking on reliability issues*. Newbury Park, CA: Sage.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, *58*(1), 6-20.

Vacha-Haase, T., Henson, R.K., & Caruso, J.C. (2002). Reliability Generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement*, *62*(4), 562-569.

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594-604.