Bilgisayar ve Eğitim Araştırmaları Dergisi
www.joucer.com

Journal of Computer and Education Research
www.joucer.com

Cronbach's Alpha is *not* a Measure of Unidimensionality or Homogeneity[1]

Cronbach'ın Alfası, Tek Boyutluluğun veya Benzeşikliğin (Homojenliğin) Bir Ölçüsü Değildir

Vahit BADEMCİ [2]

Abstract

Reliability is a characteristic of scores, not tests or measurement instruments. One of the methods to estimate the reliability of the scores on a test or measurement instrument is Cronbach's coefficient alpha method. In other words, Cronbach's alpha is an estimate of the score reliability based on the internal consistency among the [item] scores. In addition to this, Cronbach's coefficient alpha should not be interpreted as a measure of the test's or the measurement instrument's unidimensionality or taken as an index of the homogeneity of a measurement instrument. Briefly, Cronbach's alpha is not a measure of unidimensionality or homogeneity.

Keywords: Cronbach's alpha, score reliability, misconception, unidimensionality, factor analysis

Öz

Güvenirlik, testler veya ölçme araçlarının *değil*, ölçümlerin bir özelliğidir. Cronbach'ın alfası katsayısı yöntemi, bir test ya da ölçme aracından elde edilen ölçümlerin güvenirliğini kestirme yöntemlerinden biridir. Başka bir ifadeyle, Cronbach'ın alfası, [madde] ölçümler[i] arasındaki iç tutarlılık üzerine temellenmiş bir ölçüm güvenirliğinin kestirilmesidir. Bunun yanı sıra, Cronbach'ın alfası katsayısı, ölçme aracının benzeşikliğinin (homojenliğinin) bir göstergesi olarak *alınmamalı* veya testin ya da ölçme aracının tek boyutluluğunun (bir boyutluluğunun) bir ölçüsü olarak *yorumlanmamalıdır*. Kısaca, Cronbach'ın alfası, benzeşikliğin veya tek boyutluluğun bir ölçüsü *değildir*.

Anahtar kelimeler: Cronbach'ın alfası, ölçüm güvenirliği, yanıltmaç (kavram yanılgısı), tek boyutluluk (bir boyutluluk), faktör analizi (etken çözümlemesi)

Misconceptions and their overcome within the context of this article has been brought to agenda in Turkey by Bademci (2001; 2006a; 2006b; 2006c).

[1] *This article is a part of the declaration presented by the author during the conference "Paradigm shift: Tests are not reliable" given at Gazi University on 28 April 2006 [Conference, in Turkish]. [For your information; Gazi University. (2006). Gazi Haber, 66, 64. - in Turkish].*

[2] Gazi University, bademci@gazi.edu.tr

On Cronbach's Alpha

One of the methods to estimate the reliability of test scores is Cronbach's (1951) coefficient alpha (α) method (Bademci, 2006a; 2006b; Linn & Miller, 2005; Mehrens & Lehmann, 1991; Worthen, White, Fan & Sudweeks, 1999). In other words, Cronbach's alpha is an estimate of the score reliability based on the internal consistency among the [item] scores. Coefficient alpha is calculated by the following formula (Crocker & Algina, 1986; Henson, 2000; Reinhardt, 1996).

$$\alpha = k / (k-1) \, [1 - (\sum \sigma_k^2 / \sigma_T^2)] \qquad \text{[Formula 1]}$$

k = the number of items on the test

$\sum \sigma_k^2$ = the sum of the k item score variances

σ_T^2 = the variance of the total test scores

Coefficient alpha is affected by item difficulty, sum of item score variances, and total test score variance (Helms, 1999). Reinhardt (1996) demonstrated that total test score variance has the greatest effect on Cronbach's coefficient alpha. Thompson (2003) also noted the variance of total test scores is not only a function of the item score variances, but also of the covariances of the item scores with each other. Smaller total test score variance leads to smaller coefficient alpha and bigger total test score variance leads the coefficient alpha to increase (Arnold, 1996; Helms, 1999).

Because, classical true score theory reliability estimates are affected by the total test score variance (Capraro, Capraro & Henson, 2001), also the total test score variance is greatly affected by how homogeneous or heterogeneous are the group of examinees (Helms, 1999).

If a test is given to a homogeneous group, the variability in total test score will decrease and then the alpha coefficient will be smaller; if the same test is given to a heterogeneous group, the variability in the total test score will increase and then the alpha coefficient will also increase (Arnold, 1996; Helms, 1999). In this situation, when the same

measurement instrument is given to the more homogeneous or the more heterogeneous groups (or samples), differing score reliability will occur (Thompson, 1994).

Reliability is a Property of Scores, not Measurement Instruments

To say in other words, the alpha coefficients, calculated from the scores obtained from applying the same measurement instrument to different samples, will take changing values according to the sample characteristics (Guthrie, 2000). Cronbach's alpha can (and will) vary from sample to sample (Bademci, 2001; 2006b; Knapp, 1991). Bademci (2006b) emphasized, "reliability is greatly effected by the sample characteristics" (p. 7). Briefly, reliability is population or sample dependent (Bademci, 2001; 2006c; Mellenberg, 1996). From this point, it can be stated that reliability is not a characteristic of tests, but a property of the scores on a test or measurement instrument for a *particular* population or sample of examinees (Bademci, 2001; 2006c; Caruso, 2000; Crocker & Algina, 1986; Thompson, 1994; 1999; Wilkinson & APA Task Force on Statistical Inference, 1999).

Therefore, to accept reliability as a characteristic of tests or measurement instruments but not the scores, to act in the direction of this "unconscious paradigmatic belief" (Thompson, 1994, p. 839) in the researches and to speak of "the reliability of the test" or to say "the test is reliable" when referring to an instrument of measurement is *not convenient* and is *not correct* (Bademci, 2004; 2005; Guthrie, 2000). Reliability refers to the reproducibility or the consistency of the scores on a test or measurement instrument for a *particular* population or sample of examinees (Bademci, 2001; 2006c). In addition, the scores to be used in estimating reliability coefficients can be collected by a variety of methods (e.g., test-retest method, Cronbach's coefficient alpha method).

Briefly, reliability is a characteristic of scores, not an inherent or a stable property of tests or measurement instruments. J. E. Helms (2005) also emphasized that

> When researchers wrongfully assume that reliability is a stable property of tests, they tend to erroneously report the results of their reliability analyses as well. Some of these wrongful practices include (1) reliability induction, inferring reliability from previous studies rather than conducting one's own reliability analyses; (2)

> assuming that reliability coefficients are "inadequate" according to some specified or unspecified standard (e.g., .70) without conducting tests of significance or computing confidence intervals; (3) assuming homogeneity of sample characterististics such that all samples are supposed to respond in the same manner regardless of their characteristics; and (4) misconstruing the relationships between reliability and validity coefficients (p. 368).

Thompson (1994) noted that "Too few researchers act on a conscious recognition that *reliability is a characteristic of scores or the data in hand*" (p. 839). However, many people [still] do not understand score reliability (Mittag & Thompson, 2000; Thompson & Vacha-Haase, 2000; Vacha-Haase, Kogan, & Thompson, 2000; Vacha-Haase, Ness, Nilsson, & Reetz, 1999; Yin & Fan, 2000).

Cronbach's Coefficient Alpha is n*ot* a Measure of Unidimensionality or Homogeneity

Pedhazur and Schmelkin (1991) noted, "when homogeneity is understood to mean unidimensionality, it means that, for an instrument to be viewed of as consisting of homogeneous items, it has to be demonstrated that a general factor is sufficient to account for the relations among them." (p.102) Another settled idea about Cronbach's alpha is to suggest the relatively high alpha value as the evidence of scale unidimensionality or homogeneity or to take Cronbach's alpha as a measure of the homogeneity or the unidimensionality of a measurement instrument or test items (e.g., Bolton & Humphreys, 2002; Caci, Nadalet, Staccini, Myquel, & Boyer, 1999; Cramer & Barry, 1999; Küçükdeveci, Sahin, Ataman, Griffiths, & Tennant, 2004; Marx, Bombardier, Hogg-Johnson, & Wright, 1999; Pike & Hudson, 1998; Radovanovic & Alexandre, 2004; Tiesinga, Dassen, & Halfens, 1998). But, this is not also true. Cronbach's coefficient alpha is *not* a measure of unidimensionality or homogeneity (Bademci, 2001; 2006c; Cortina, 1993; Gerbing & Anderson, 1988; Hattie, 1985; Miller, 1995; Netemeyer, 2001; Netemeyer, Bearden, & Sharma, 2003; Rogers, Schmitt, & Mullins, 2002; Schmitt, 1996). Hattie (1985) pointed out, "Alpha is

dependent on the length of the test,...the unidimensionality of a test should be independent of its length" (p. 145). Streiner (2003) also remarked,

> "It is true that the higher the correlations among the items of a scale, the higher will be the value of α. But, the converse of this -that a high value of α implies a high degree of internal consistency- is not always true.... Because α is affected by the length of the scale, high values do not guarantee internal consistency or unidimensionality" (p. 101, 103).

Cortina (1993) explained, "*Internal consistency* refers to the degree of interrelatedness among the items..., whereas *homogeneity* refers to *unidimensionality...*" (p. 100). Consequently, Cronbach's alpha *should not* be interpreted as a measure of the test's or the measurement instrument's unidimensionality or taken as an index of the homogeneity of a measurement instrument (Crocker & Algina, 1986; Pedhazur & Schmelkin, 1991).

Examine the Unidimensionality

Nunnally and Bernstein (1994) suggested that "Factor analysis is at the heart of the measurement of psychological constructs" (p. 111). Unidimensionality may be examined using exploratory factor analysis or especially confirmatory factor analysis (see Pedhazur and Schmelkin, 1991; Stevens, 2002; Thompson, 2004). But, Cronbach's alpha should not be used as a measure of unidimensionality.

Netemeyer (2001) also noted "...alpha should be used to assess internal consistency only after unidimensionality is established" (p. 57). In other words, Cronbach's alpha should be used to estimate of the score reliability based on the internal consistency among the [item] scores *after* unidimensionality is examined.

References

Arnold, M. E. (1996). *Influences on and limitations of classical test theory reliability estimates* (ERIC Document Reproduction Service No. ED 395 950).

Bademci, V. (2001, November). *To teach thinking and the methods-techniques used in instruction*. Conference Hall of the SMMM Chamber, Bursa [Paper presented at the Conference, in Turkish].

Bademci, V. (2004). It is incorrect to speak of "the reliability of the test" or "the test is reliable". *The Journal of Turkish Educational Sciences*, *2* (3), 367- 372 [in Turkish].

Bademci, V. (2005). On the errors inside of the arbiters evaluations: Fisher's z transformation and sample size for reliability studies. *The Journal of the Industrial Arts Education Faculty of Gazi University*, *17*, 46- 75 [in Turkish].

Bademci, V. (2006a). To put an end to the discussion: Cronbach's alpha coefficient can be used with dichotomously scored items [0,1]. *Journal of Kâzım Karabekir Education Faculty*, *13*, 438- 446 [in Turkish].

Bademci, V. (2006b). To understand reliability properly and to overthrow some clichés: Contrary to the known facts, Cronbach's coefficient alpha can be negative and smaller than "–1". *İnönü University Journal of the Faculty of Education*, *7*(12), 3- 26 [in Turkish].

Bademci, V. (2006c, April). *Paradigm shift: Tests are not reliable*. Conference Hall of the Vocational Education Faculty of Gazi University, Ankara [Paper presented at the Conference, in Turkish].

Bolton, J. E., & Humphreys, B. K. (2002). The Bournemouth Questionnaire: A short-form comprehensive outcome measure. II. psychometric properties in neck pain patients. *Journal of Manipulative and Physiological Therapeutics*, *25*(3), 141- 148.

Caci, H., Nadalet, L., Staccini, P., Myquel, M., & Boyer, P. (1999). Psychometric properties of the French version of the Composite Scale of Morningness in adults. *European Psychiatry*, 14, 284- 290.

Capraro, M. M., Capraro, R. M., & Henson, R. K. (2001). Measurement error of scores on the mathematics anxiety rating scale across studies. *Educational and Psychological Measurement*, 61, 373- 386.

Caruso, J. C. (2000). Reliability generalization of the NEO personality scales. *Educational and Psychological Measurement*, 60, 236- 254.

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

Cramer, K. M., & Barry, J. E. (1999). Psychometric properties and confirmatory factor analysis of the Self-Concealment Scale. *Personality and Individual Differences*, *27*, 629- 637.

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

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

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research*, *25*, 186- 192.

Guthrie, A. C. (2000). *A review of coefficient alpha and some basic tenets of classical measurement theory* (ERIC Document Reproduction Service No. ED 438 307).

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

Helms, L. S. (1999). *Basic concepts in classical test theory: Tests aren't reliable, the nature of alpha, and reliability generalization as meta-analytic method* (ERIC Document Reproduction Service No. ED 427 083).

Helms, J. E. (2005). Challenging some misuses of reliability as reflected in evaluations of the White Racial Identity Attitude Scale (WRIAS). In R. T. Carter (Ed.), *Handbook of racial-cultural psychology and counseling: Theory and research, volume 1* (pp. 360- 390). Hoboken, New Jersey: John Wiley & Sons.

Henson, R. K. (2000). *A primer on coefficient alpha* (ERIC Document Reproduction Service No. ED 447 210).

Knapp, T. R. (1991). Coefficient alpha: Conceptualizations and anomalies. *Research in Nursing and Health*, 14, 457- 460.

Küçükdeveci, A. A., Sahin, H., Ataman, S., Griffiths, B., & Tennant, A. (2004). Issues in cross-cultural validity: Example from the adaptation, reliability, and validity testing of a Turkish version of the Stanford Health Assessment Questionnaire. *Arthritis & Rheumatism (Arthritis Care & Research)*, *51*(1), 14- 19.

Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching* (9th ed.). Upper Saddle River, New Jersey: Pearson.

Marx, R. G., Bombardier, C., Hogg-Johnson, S., & Wright, J. G. (1999). Clinimetric and psychometric strategies for development of a health measurement scale. *Journal of Clinical Epidemiology*, *52*(2), 105- 111.

Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology* (4th ed.). Fort Worth : Harcourt Brace.

Mellenbergh, G. J. (1996). Measurement precision in test score and item response models. *Psychological Methods*, *1*(3), 293- 299.

Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling*, *2*(3), 255-273.

Mittag, K. C. & Thompson B. (2000). A national survey of AERA members' perceptions of statistical significance tests and other statistical issues. *Educational Researcher*, *29*(4), 14-20.

Netemeyer, R. (2001). [in] Measurement. *Journal of Consumer Psychology*, *10*(1&2), 55- 69.

Netemeyer, R. G., Bearden, W. O., & Sharma S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, California: Sage.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Pike, C. K., & Hudson, W. W. (1998). Reliability and measurement error in the presence of homogeneity. *Journal of Social Service Research*, *24*(1-2), 149- 163.

Radovanovic, C. A. T., & Alexandre, N. M. C. (2004). Validation of an instrument for patient handling assessment. *Applied Ergonomics*, *35*, 321- 328.

Reinhardt, B. (1996). Factors affecting coefficient alpha: A mini Monte Carlo study. In B. Thompson (Ed.), *Advances in social science methodology, volume 4* (pp. 3- 20). Greenwich, Connecticut: JAI.

Rogers, W. M., Schmitt, N., & Mullins, M. E. (2002). Correction for unreliability of multifactor measures: Comparison of alpha and parallel forms approaches. *Organizational Research Methods*, 5(2), 184- 199.

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

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80(1), 99- 103.

Tiesinga, L. J., Dassen, T. W. N., & Halfens, R. J. G. (1998). DUFS and DEFS: Development, reliability and validity of the Dutch Fatigue Scale and the Dutch Exertion Fatigue Scale. *International Journal of Nursing Studies*, *35*, 115- 123.

Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement*, 54, 834- 847.

Thompson, B. (1999). Five methodology errors in educational research: A pantheon of statistical significance and other faux pas. In B. Thompson (Ed.), *Advances in social science methodology, volume 5* (pp. 23- 86). Stamford, Connecticut: JAI.

Thompson, B. (2003). Understanding reliability and coefficient alpha, really. In B. Thompson (Ed.), *Score reliability* (pp. 3- 30). Thousand Oaks, California: Sage.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis*. Washington, DC: American Psychological Association.

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement*, 60, 174- 195.

Vacha-Haase, T., Kogan L. R., & Thompson, B. (2000). Sample compositions and variabilities in published studies versus those in test manuals: Validity of score reliability inductions. *Educational and Psychological Measurement*, 60, 509- 522.

Vacha-Haase, T., Ness, C., Nilsson, J., & Reetz, D. (1999). Practices regarding reporting of reliability coefficients: A review of three journals. *The Journal of Experimental Education*, 67(4), 335- 341.

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

Worthen, B. R., White, K. R., Fan, X., & Sudweeks, R. R. (1999). *Measurement and assessment in schools* (2nd ed.). New York: Addison Wesley Longman.

Yin, P., & Fan, X. (2000). Assessing the reliability of Beck Depression Inventory Scores: Reliability generalization across studies. *Educational and Psychological Measurement*, 60, 201- 223.