

Ivan Šerbetar
Iva Sedlar

Assessing Reliability of a Multi-Dimensional Scale by Coefficient Alpha

Short scientific article

UDK: 796.012

ABSTRACT

The purpose of the study was to assess internal consistency by calculating coefficient alpha. It presents the variation in coefficient alpha, depending on questionnaire length and the homogeneity or heterogeneity of the questionnaire. The maximum possible value for coefficient alpha was also calculated by the item elimination method.

The study included 99 children aged 10. The children completed *The Athletic Coping Skills Inventory – 28* (ACSI-28; Smith et al., 1995), which contains seven constructs: *coping with adversity, coachability, concentration, confidence and achievement motivation, goal setting and mental preparation, peaking under pressure and freedom from worry*. The results confirmed that the values of the alpha coefficient vary depending on the number and composition of items and the sample size. In terms of item structure, homogeneous constructs yielded lower values for the alpha coefficient (in a range from .48 to .61) than the questionnaire with all the constructs (alpha = .79), despite higher inter-item correlations. In terms of the number of items, the longer test generated higher alpha coefficients (alpha = .79) than the shorter test (*half-sets of items* = .60, .73, .69, .70).

A higher overall value (alpha = .83) can be achieved by item elimination.

Key words: coefficient alpha, internal consistency, reliability

Ocenjevanje zanesljivosti večrazsežnostne lestvice s koeficientom alfa

Kratki znanstveni članek

UDK: 796.012

POVZETEK

Namen raziskave je bil oceniti notranjo skladnost z izračunavanjem koeficienta alfa. Predstavlja spreminjanje koeficienta alfa odvisno od dolžine vprašalnika ter njegove homogenosti ali heterogenosti. Najvišja možna vrednost za koeficient alfa je bila izračunana tudi po metodi izločanja postavk.

Raziskava je vključevala 99 otrok, starih 10 let. Otroci so izpolnjevali *The Athletic Coping Skills Inventory – 28* (ACSI-28; Smith idr., 1995), ki vsebuje sedem sestavin: *obvladovanje*

stiske, sposobnost biti voden, koncentracija, zaupanje in sprejemanje motivacije, zastavljanje ciljev in miselna priprava, doseganje vrhunca pod pritiskom in osvobojenost skrbi. Rezultati so potrdili, da se vrednosti koeficienta alfa spreminjajo odvisno od števila in sestave postavk in od velikosti vzorca. V smislu strukture postavk so homogene konstrukcije dajale nižje vrednosti za alfa koeficient (v razponu med 0,48 do 0,61) kot vprašalnik z vsemi sestavinami (alfa = 0,79), kljub višjim korelacijam med postavkami. V smislu števila postavk je daljši test generala višji alfa koeficient (alfa = 0,79) kot krajši test (*polovični nabori postavk* = 0,60, 0,73, 0,69, 0,70). Višjo skupno vrednost (alfa = 0,83) je mogoče doseči z izločanjem postavk.

Ključne besede: koeficient alfa, notranja skladnost, zanesljivost

Introduction

Coefficient alpha, usually known as *Cronbach alpha*, on account of Cronbach's seminal article in 1951, is probably the most widely used reliability coefficient. The Split-half and Kuder-Richardson KR-20 (1937) approaches are usually mentioned as predecessor methods, while somewhat less credit is given by the literature to the Guttman (1945) series of lower bounds for reliability $\lambda_1 \dots \lambda_6$, which formed the basis for most of the later estimates of reliability. Of those six, the third was λ_3 , the most prominent estimate, known as *coefficient alpha* or just α , as Cronbach (1951) later named it. Coefficient alpha is an estimate of reliability, or more precisely, an estimate of internal consistency, but perhaps it is best explained as an *index of internal consistency of the scale*. Consistency here means inter-relatedness among the items of the test (Cortina, 1993) or in other words, whether the items are consistent between themselves to a sufficient degree for them to be combined with one another.

Coefficient alpha is grounded in classic test theory (Nunnally & Bernstein, 1994), which assumes that the observed score is composed of the true score and the measurement error ($Y = T + E$). Consequently, reliability may be defined as the ratio of true score variance and observed score variance. If we square the correlation between the observed scores and the true scores, we get the coefficient alpha. Following the theory, it is also assumed that the measurement error is minimized in reliable tests, and the correlation between error and true score is low. A further assumption is that the mean of the error component should be zero, which means that the error scores are random and not correlated among themselves. If that assumption is not met, the coefficient alpha may be over-estimated.

Streiner and Norman (1995) defined reliability as the degree to which "measurement of individuals on different occasions, or by different observers or by similar or parallel tests, produce the same or similar results" (p. 6). That can add the third source of error, the one associated with the *homogeneity* of the items of the scale (Streiner, 2003). Tests are said to be homogeneous if they contain items that measure a single trait (Cohen, Swerdlik & Phillips, 1996).

Homogeneity is related to the *unidimensionality* of the items in the scale, which is a prerequisite for internal consistency. Basically, the concept of reliability assumes unidimensionality of the sample of test items, and if the assumption of unidimensionality is violated, it causes a major underestimate of reliability (Tavakol & Dennick, 2011). In contrast to *homogeneity*, there is the *heterogeneity* concept, which refers to the degree to which a test measures different factors.

Heterogeneity is closely related to the concept of *tau-equivalency*, which postulates that items in a scale are linearly related and differ only by a constant (Cortina, 1993); in other words, each item on the scale is supposed to measure the same construct. Hence, in the case of multidimensional tests, which contain more than one construct (heterogeneity), this assumption is violated and alpha underestimates the reliability of the test (Tavakol & Dennick, 2011). However, in a multidimensional test it is not necessary for alpha to have a lower value than in a unidimensional test. This is simply because of the inflation of variance in multidimensional tests as a result of the large number of items.

The computation of alpha is based on item variances or inter-item correlations (standardized alpha).

A common reference for the acceptability level is Nunnally and Bernstein's (1994) postulated value of .70 mentioned as a low acceptability value for "exploratory purposes". To put it simply, in terms of variance, this means that 70 % of the variance in the scores is reliable and 30 % of variance in the scores belongs to error variance. The value of coefficient alpha usually ranges from 0 to 1, but the value could also be negative when the covariance of the items is very low.

The purpose of this article is to demonstrate how coefficient alpha is affected by the dimensionality of the scale, and how the value of the alpha coefficient may be increased by item trimming. The article is also intended to further promote the proper application of the alpha coefficient, especially in kinesiology related research.

Methods

The above-described approach will be applied by analyzing Smith, Schutz, Smoll, and Ptacek (1995) *The Athletic Coping Skills Inventory-28* (ACSI-28), a multidimensional measure of sport-specific psychological skills. The inventory was designed with the purpose of assessing competitive stress in athletes.

ACSI-28 contains seven sport-specific scales (*Coping with Adversity, Peaking under Pressure, Goal Setting/Mental Preparation, Concentration, Freedom from Worry, Confidence and Achievement Motivation, and Coachability*), but also provides an overall ranking and composite score.

The data used in this study were obtained from 99 children aged 10 who were participants in sport; the children completed the inventory as part of the graduate thesis work of the second author of this study. The psychometric properties of the

ACSI-28 scale per se were not of primary interest in the present article. Instead, it was meant to test some classic issues concerning coefficient alpha, particularly the issue of dimensionality. Therefore, computation of overall subscales and split scale alpha values was performed, as well as stepwise procedures for item elimination with the goal of increasing coefficient alpha.

Results

An alpha coefficient of .79 was initially obtained. The values of alpha in the subscales ranged from .48 to .61. Two salient facts may be observed from Table 1: first, the higher the intercorrelations of the items, the higher the alpha value for that subscale; and second, although the subscale's alphas are low, the value of the alpha coefficient for the whole questionnaire is acceptably high, despite the very low average intercorrelation of .12. The explanation lies in the long-established fact that coefficient alpha is affected by the length of the scale (Nunnally & Bernstein, 1994; Cortina, 1993; Streiner, 2003).

Table 1: Initial scales and overall statistical values ($N = 99$)

Scale	Mean (SD)	Intercorrelations	σ^2 item	σ^2 subscale	Alpha
Coping with adversity	11.51 (2.52)	.25	.89	6.34	.579
Coachability	13.31 (2.37)	.26	.78	5.61	.587
Concentration	11.31 (2.28)	.19	.84	5.22	.478
Confidence and achievement motivation	12.34 (2.10)	.19	.69	4.41	.492
Goal setting and mental preparation	10.17 (2.64)	.25	.99	6.98	.573
Peaking under pressure	10.19 (2.87)	.29	1.12	8.22	.614
Freedom from worry	11.72 (2.55)	.23	.97	6.51	.542
Overall alpha	80.56 (10.29)	.12	.89	105.88	.791

Cortina (1993) has shown that a 6-item scale with an average intercorrelation of .30 may yield an alpha value of .72; moreover, by increasing the number of items by 6 and 12, while keeping the correlation constant, the alpha coefficient values are increased to .84 and .88, respectively. Cortina (1993) further showed that even if the two dimensions of the scale are orthogonal, the alpha could reach the value of .45 for a 6-item subscale if the inter-item correlations are constant.

In the shift to a practical context, it is debatable whether it makes sense to report coefficient alpha for the whole heterogeneous instrument because of the inflated alpha caused by a large number of items. On the contrary, reporting alpha values for the subscales, which display homogeneity for each construct, is supposed to be obligatory.

Table 2: Statistical values for split item sample

N = 99					
Items	Mean (SD)	Intercorrelations	$\sigma^2 \sigma^2$ items	$\sigma^2 \sigma^2$ total	Alpha
1. – 14.	40.74 (5.19)	.10	.85	27.03	.60
15. – 28.	39.82 (6.42)	.17	.94	41.17	.73
Even items	39.81 (5.96)	.14	.91	35.46	.69
Odd items	40.75 (5.97)	.15	.89	35.58	.70
All items	80.56 (10.29)	.12	.89	105.88	.79

Although scale reliability still falls short of the currently recommended value of .80 (Cortina, 1993), both methods of splitting the inventory improved reliability in comparison to reliability in the subscales (Table 2). At the same time, reliability among the halves of the scale is still lower than that for the total scale. Both findings again demonstrate that internal consistency is affected by the length of the scale. Streiner (2003) stressed that scales over 20 items or so will have acceptable values of alpha even if they consist of orthogonal dimensions. As seen before, inter-correlations are low. These are determined by the inter-relatedness of the items, which in turn, determines scale consistency. Therefore, if the sample of items is heterogeneous, variance of the total score will increase and the alpha will be higher.

Table 3: Increase in alpha values after the removal of less consistent items (initial alpha value = .79)

Step	Nr. of item removed	Inter-item correlation	Adjusted alpha
1.	23	.14	.806
2.	19	.15	.812
3.	3	.15	.815
4.	10	.16	.819
5.	12	.17	.823
6.	8	.18	.825
7.	7	.19	.828
8.	11	.20	.829

Removal of the items or the “alpha-if-deleted” approach (so designated in McDougal, 2011) is performed here only for demonstrative purposes, while otherwise the procedure is used in the early stages of test design. That method, supported by many software packages, involves the removal of the less consistent items in a stepwise fashion – one at a time, with the goal of improving reliability. However, what is shown in Table 3 is an increase in alpha from an initial value of .79 to one of .83, as a result of the eight-step removal process.

The utility of trimming the items is obvious: overall alpha has increased, but because of the trimming, the subscales’ alpha value may decrease. Therefore, adding new items in place of those that were removed may be the proper route to the development of the instrument.

Discussion

Coefficient alpha is a very economical statistic: it does not depend on multiple administration nor multiple examiners. Coefficient alpha is also easy to compute but, as seen from the current data, coefficient alpha yields only the extent to which all the items in a test measure the same concept or construct (Tavakol & Dennick, 2011), or provides an index of the inter-relatedness of the items. In other words, a high alpha does not guarantee homogeneity or unidimensionality of the scale. An alpha could be low for several reasons: an insufficient number of items, low inter-item correlations or heterogeneous constructs. An alpha could also be too high; i.e., a value higher than .90 often indicates redundancy (Streiner, 2003) and points to an excessive number of items in the scale. In many research studies, as noticed by Cortina (1993), a high alpha value is accepted as adequate with no further adjustment of the scale, which is not the proper way to use coefficient alpha.

Moreover, coefficient alpha is not a fixed property of a scale, which means that a scale may be sufficiently reliable for one group of subjects but unreliable for another. As has been stressed by many authors (quoted in Streiner, 2003), reliability is a characteristic of the test *scores*, not of the test itself; hence, reliability depends as much on the sample being tested as on the test. The practical implication is that each time the test is to be used, the consistency should be assessed on that particular sample.

It is also mandatory for instrument developers to report the psychometric properties of the instrument, including item statistics and dimensionality assessment (Rodriguez & Maeda, 2006). Additionally, researchers should bear in mind, as expressed by Cortina (1993), "that those who make decisions about the adequacy of a scale on the basis of nothing more than the level of alpha are missing the point of empirically estimating reliability" (p.102).

REFERENCES

- Cohen, R. J., Swerdlik, M. E., & Phillips, S. M. (1996). *Psychological Testing and Assessment: An introduction to tests and measurement* (3rd ed.). Mountain View, CA: Mayfield Publishing Company.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, Vol 78(1), 98-104.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of the tests. *Psychometrika*, 16, 297-334.
- Guttman, L. (1945). *A basis for analysing test-retest reliability*. *Psychometrika*, 10, 255–282.
- Kuder, G. F., Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, Vol. 2 (3), pp 151-160.
- MacDougall, M. (2011). Moving Beyond the Nuts and Bolts of Score Reliability in Medical Education: Some Valuable Lessons from Measurement Theory. *Advances and Applications in Statistical Sciences*, 6(7), 643-664.

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

Raubenheimer, J. E. (2004). An item selection procedure to maximize scale reliability and validity. *South African Journal of Industrial Psychology*, 30 (4), 59-64.

Rodriguez, M. C., Maeda, Y. (2006). Meta-analysis of coefficient alpha. *Psychological Methods*. Sep; 11(3)

Smith, R. E., Schutz, R. W., Smoll, F. L., Ptacek, J. T. (1995). Development and Validation of a Multidimensional Measure of Sport-Specific Psychological Skills: The Athletic Coping Skills Inventory-28. *Journal of Sport & Exercise Psychology*, 17: 379-398.

Streiner, D. L. (2003). Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment*, 80(1):99-103.

Streiner, D. L., & Norman, G. R. (1995). *Health Measurement Scales: A practical guide to their development and use* (2nd ed.) Oxford: Oxford University Press.

Tavakol, M., Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*. 2:53-55.

