The world should not revolve around Cronbach's alpha ≥ .70

El mundo no debería girar alrededor del alfa de Cronbach ≥ ,70

José Ventura-León*, Brian Norman Peña-Calero**.

* Universidad Privada del Norte, Lima, Perú.

** Universidad Nacional Mayor de San Marcos, Grupo de Estudios Avances en Medición Psicológica, Lima, Perú.

The alpha coefficient (α) is one of the most widely used for estimating reliability and is normally understood as a measure of internal consistency; that is, the degree of interrelation between items (Cortina, 1993). However, it was also considered to be a measure of homogeneity, a concept reflecting the one-dimensionality of items; this confusion was actually generated by Cronbach himself (1951) by using the terms interchangeably in his seminal article. Nevertheless, Cho and Kim (2015) indicate that α is a function of the interrelation of items with the number of items (see equation 1), so the number of items is a factor which affects the α coefficient (Murphy & Davidshofer, 2004). Thus, an article recently published in the journal *Adicciones* used the α coefficient to create a scale with 38 items, and noted that the limited reliability obtained in the study was due to the low number of items (Benito et al., 2019). It therefore appears necessary to reflect on the use of α ≥ .70 in studies on addictions.

$$\alpha = \frac{K}{K-1}\left[1-\frac{\sum S_i^{\,2}}{S_T^{\,2}}\right] \quad \text{Equation (1)}$$

The α coefficient ranges from 0 to 1, and values ≥ .70 are considered acceptable (Cicchetti, 1994). This recommendation stems from Nunnally's proposal (1978) and is used as the cut-off criterion in 44% of articles (Lance, Butts & Michels, 2006). However, a careful review of Nunnally (1978) reveals that the recommendation applied to preliminary research and that the value was not attributed to the α coefficient itself, but to a measure of general reliability. In addition, a meta-analytic study showed that the average value of α is .77 and that this changes depending on the subject area in which it is applied (Peterson, 1994). It is therefore rather simplistic to make decisions based on a single value, and it is necessary to incorporate the inter-item correlation matrix, its mean and standard error (Cortina, 1993; see Equation 2).

$$\frac{DE_r}{[(\frac{1}{2} * k * k[k-1]) - 1]^{1/2}} \quad \text{Equation (2)}$$

To demonstrate the sensitivity of α to the number of items, the R program was used to simulate data, generating scales from 3 to 12 items in a one-dimensional structure, following a normal distribution (Mean = 0, *SD* = 1) with mean inter-item correlations of 0.10, 0.15, 0.20 and 0.25, reflecting little variability between the items, and with sample sizes of 50, 100, 250, 500 and 1000. Thus, 200 (10 x 4 x 5) simulation conditions were generated, with 1000 repetitions for each condition, giving a total of 200,000 simulated data sets (see https://osf.io/fngte/).

The results showed that α increases with an increasing number of items, despite the fact that the inter-item correlations had little variability; α ≥ 0.70 was found on scales with 7 or more items in samples of 50 or higher and with inter-item correlations of 0.15 (see Figure 1). These findings confirm the hypothesis that the number of items has an impact on the α coefficient, a factor which must be taken into account in interpreting the data to ensure that internal consistency is actually measured as a product of the variability of item scores rather than seen as a function of the increase in their number.

Received: June 2020; Accepted: July 2020.

Send correspondence to:
Jose Ventura-León. Av. Tingo María 1122, Breña, Lima. Telf. (01)6044700. Anexo: 3462
E-mail: jose.ventura@upn.pe

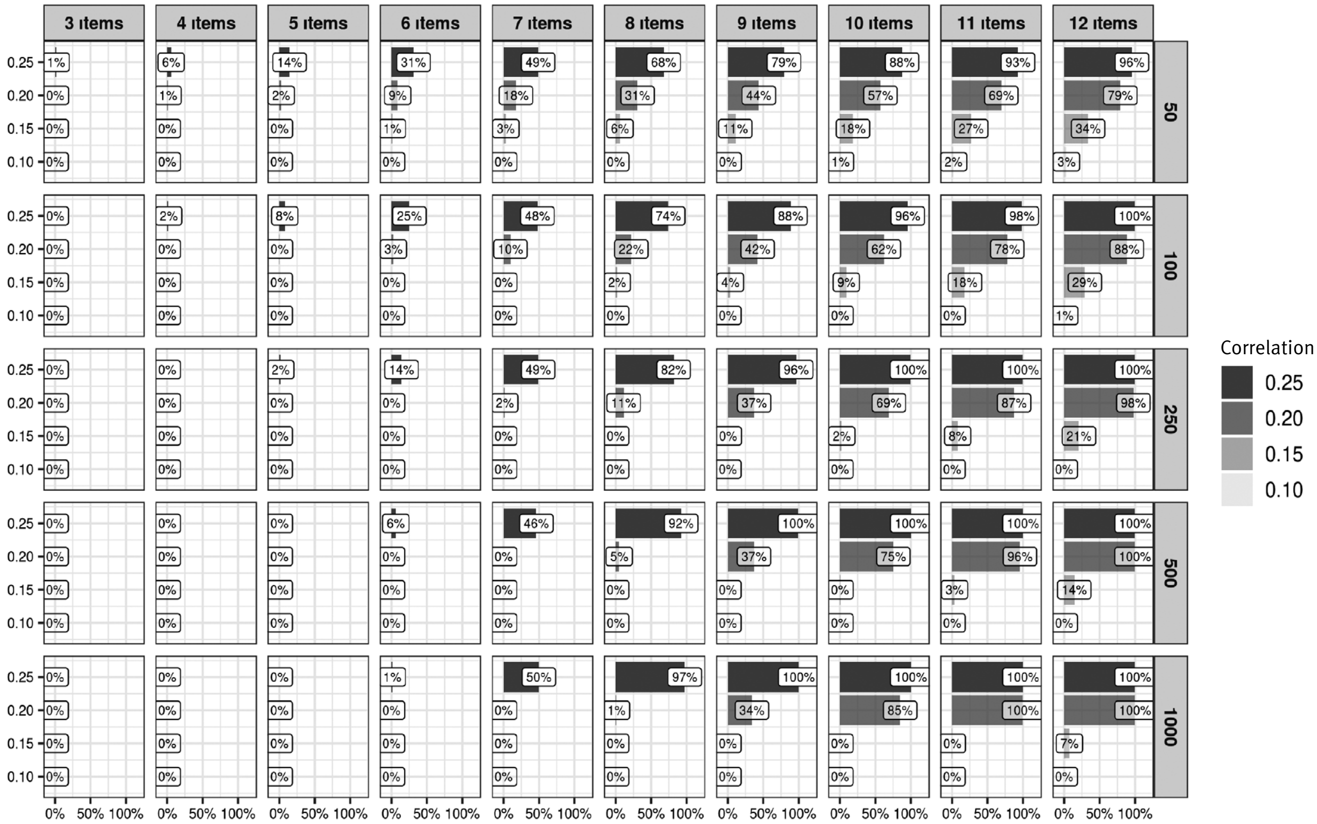

Figure 1. Percentage of cases with α coefficients ≥ .70 with different numbers of items, sample sizes and inter-item correlations.

Thus, α should be interpreted as the amount of common variability that can be attributed to the factor depending on the number of scale items. An α of .79 would therefore indicate that 79% of the variability of the items is due to the consistency (or coherence) of the responses of a group of people, in a measurement instrument comprising a certain number of items.

These results serve as an invitation to researchers to explore other ways of estimating reliability, such as the omega coefficient by means of factor models from classical test theory, or the information function from item response theory (Ventura-León, 2019), which can all be calculated with free access programs such as R, Factor, Jamovi and/or JASP. Furthermore, in conditions of non-tau-equivalence, multidimensionality, or the presence of correlated errors, reliability estimates of α are problematic (Raykov & Marcoulides, 2019), and since the α preconditions are difficult to find in real situations, this coefficient is falling out of use as an indicator of reliability (Peters, 2014). However, this does not indicate that the α is malfunctioning.

In conclusion, it does not appear to make sense to use a continuous measurement ranging from 0 to 1 as the product of α if in the end the interpretation is dichotomized as "Not Acceptable" or "Acceptable" on being below or above a value such as .70. Instead, indicating that a test has reliable scores requires reporting the inter-item correlation matrix, its mean, its standard deviation, the number of items, and previous or meta-analytical studies to account for the research context (Peterson, 1994) because relying solely on a value without the dataset is like looking at the tree without seeing the forest.

References

Benito, A., Calvo, G., Real-López, M., Gallego, M. J., Francés, S., Turbi, Á. & Haro, G. (2019). Creación y estudio de las propiedades psicométricas del cuestionario de socialización parental TXP. *Adicciones, 31*, 117-135. doi:10.20882/adicciones.983.

Cho, E. & Kim, S. (2015). Cronbach's coefficient alpha. *Organizational Research Methods, 18*, 207-230. doi:10.1177/1094428114555994.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284-290. doi:10.1037/1040-3590.6.4.284.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98-104. doi:10.1037/0021-9010.78.1.98.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334. doi:10.1007/BF02310555.

Lance, C. E., Butts, M. M. & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria. *Organizational Research Methods, 9*, 202-220. doi:10.1177/1094428105284919.

Murphy, K. R. & Davidshofer, C. O. (2004). *Psychological Testing: Principles and Applications* (6th ed.). New Jersey, USA: Pearson/Prentice Hall.

Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Peters, G. J. Y. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *The European Health Psychologist, 16*, 56-69.

Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research, 21*, 381. doi:10.1086/209405.

Raykov, T. & Marcoulides, G. A. (2019). Thanks coefficient alpha, we still need you! *Educational and Psychological Measurement, 79*, 200-210. doi:10.1177/0013164417725127.

Ventura-León, J. L. (2019). Is this the end for Cronbach's alpha? *Adicciones, 31,* 80-81. doi:10.20882/adicciones.1037.