The Role of Item Distributions on Reliability Estimation: The Case of Cronbach's Coefficient Alpha

Educational and Psychological

Measurement

2020, Vol. 80(5) 825–846

© The Author(s) 2020

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0013164420903770

journals.sagepub.com/home/epm

Oscar Lorenzo Olvera Astivia¹, Edward Kroc²
and Bruno D. Zumbo²

Abstract

Simulations concerning the distributional assumptions of coefficient alpha are contradictory. To provide a more principled theoretical framework, this article relies on the Fréchet–Hoeffding bounds, in order to showcase that the distribution of the items play a role on the estimation of correlations and covariances. More specifically, these bounds restrict the theoretical correlation range $[-1, 1]$ such that certain correlation structures may be unfeasible. The direct implication of this result is that coefficient alpha is bounded above depending on the shape of the distributions. A general form of the Fréchet–Hoeffding bounds is derived for discrete random variables. R code and a user-friendly shiny web application are also provided so that researchers can calculate the bounds on their data.

Keywords

reliability theory, coefficient alpha, item distribution, classical test theory

Introduction

Cronbach's coefficient alpha is, to this day, one of the most widely used statistics within the psychological and social sciences (Hogan et al., 2000; Peterson, 1994). With almost 38,000 citations and counting in Google Scholar, Cronbach's original

¹University Of South Florida, Tampa, Florida

²University of British Columbia, Vancouver, British Columbia, Canada

Corresponding Author:

Bruno D. Zumbo, University of British Columbia, 2125 Main Mall, Scarfe Building, Vancouver, British Columbia, V6T1Z4, Canada.

Email: bruno.zumbo@ubc.ca

1951 paper introducing this measure of reliability has become the second most cited article within the methodological literature in the behavioural sciences, only surpassed by Baron and Kenny's (1986) introduction to mediation models. Given its influence in the theory of measurement and classical and modern psychometrics, it is not surprising that studying its small sample properties and sensitivity to assumptions have been the subject of much research, from both an analytical perspective and a simulation-based perspectives (for a comprehensive list of these studies, see McNeish, 2018). An assumption that, until relatively recently, had not received as much attention (and is still the subject of some controversy) is the issue of distributional assumptions and whether or not continuous, normally distributed items are needed for the estimation of coefficient alpha. Although distributional assumptions were never invoked by Cronbach himself in the derivation of this statistic and Raykov and Marcoulides (2019) lay the argument of why these assumptions are, indeed, unnecessary, research around this area has continued with what could be construed as paradoxical results.

Distributional Assumptions of Cronbach's Coefficient Alpha

Bay (1973), Shultz (1993), and Zimmerman et al. (1993) were among the first to investigate what kind of distributional shapes influence the estimation and variability of coefficient alpha. The results from their simulation studies are mixed. Bay (1973) argues that true scores with positive skew and positive excess kurtosis both bias its small sample estimate and negatively influence its standard error and F statistic as derived by Kristof (1963) and Feldt (1965), making the shape of the test item distributions relevant to performing inference. Shultz (1993) and Zimmerman et al. (1993), on the other hand, state that the distribution of the error scores bears little influence on the estimation of coefficient alpha and that violations of other assumptions (e.g., correlated errors or the congeneric test model) exert a stronger influence on the correct estimation of this reliability coefficient. It is important to point out that the simulation designs and conditions that each article investigated are not directly comparable. For instance, Zimmerman et al. (1993) and Shultz (1993) included only the shape of the error distribution as a simulation condition, whereas Bay (1973) investigated both true score and error distribution. For number of respondents (i.e., sample size) and number of items investigated, Bay (1973) looked only at the condition of 30 respondents with eight-item tests, whereas Zimmerman et al. (1993) and Shultz (1993) looked at several distribution shapes (e.g., uniform, exponential, mixed-normal, and negative exponential) and various combinations of number of items and sample sizes.

To provide a more complete analysis, Sheng and Sheng (2012) conducted a more extensive simulation investigating six types of distributions with varying degrees of skewness and excess kurtosis, at four sample sizes (30, 50, 100, and 1,000) and three test lengths (5, 10, and 30 items) for population Cronbach's alpha of .3, .6, and .8. They concluded that the reliability coefficient is not robust to nonnormal

distributions, particularly for excess kurtosis, where they found an upward small sample bias when the items exhibit negative kurtosis and an underestimation if the items had positive kurtosis. Their conclusions fall in line with Bay's (1973) previous findings, although Bay (1973) hypothesized that, with increasing sample size, the distribution of the items would exert less influence on the estimation of coefficient alpha. The results presented in Sheng and Sheng (2012) only weakly support this claim given that even at the largest sample size condition of 1,000, bias was still present (albeit reduced). Overall, the simulation studies point toward a more complex relationship between the distribution of the test items and the proper estimation of coefficient alpha; no definitive conclusions can be reached.

The Range of the Pearson Correlation and the Fréchet–Hoeffding Bounds

To reconcile some of the seemingly contradictory findings in the previous simulation studies, one can recast this problem as one of the relationship between the marginal distributions of the test (i.e., the items) and their joint distribution so that the problem can be analyzed via copula distribution theory.

Consider the random variables (X, Y) with joint cumulative distribution function (CDF) $H(x, y)$ and marginal CDFs $F(x)$ and $G(y)$. Hoeffding (1940) proved that, for X and Y with finite variances

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(x, y) - F(x)G(y) dx dy. \quad (1)$$

Define $H(x, y)_{\min} = \max[F(x) + G(y) - 1, 0]$ and $H(x, y)_{\max} = \min[F(x), G(y)]$. Fréchet (1951) and Hoeffding (1940) independently showed that

$$H(x, y)_{\min} \leq H(x, y) \leq H(x, y)_{\max} \quad (2)$$

so that the inequality shown in Equation 2 became known as the Fréchet–Hoeffding bounds. A trivial (albeit crucial) corollary to this inequality is that if $H(x, y)_{\min}$, $H(x, y)$, and $H(x, y)_{\max}$ are standardized to become Pearson correlation coefficients ρ , it follows that $\rho_{\min} \leq \rho \leq \rho_{\max}$. This implies that, given the distributional forms of $F(x)$ and $G(y)$, the Pearson correlation coefficient need not span its full theoretical range $[-1, +1]$. These bounds apply to any probability distribution, discrete, continuous, or mixture, as long as it has finite second moments (Joe, 2014). Moreover, these bounds are sharp, as can be seen by letting X and Y be increasing functions of each other (to obtain the upper bound) or by letting X and Y be decreasing functions of each other (to obtain the lower bound).

In light of seemingly conflicting results within the published literature, where coefficient alpha is portrayed sometimes as nonrobust to the assumption of normally distributed items (Bay, 1973; Sheng & Sheng, 2012) and sometimes as robust (Shultz, 1993; Zimmerman et al., 1993), the purpose of this article is to rely on the Fréchet–Hoeffding bounds to investigate under which conditions and what type of

item distributions a bias may exist for coefficient alpha. The exact upper bound will be derived under the assumption of discrete random variables and illustrative examples will be presented for the case of continuous items. A modification to the generate–sort–correlate (GSC) method introduced by Demirtas and Hedeker (2011) will be used to explore what the maximum possible correlations (and, by extension, alpha) are in certain sections of the studies presented in Bay (1973). Finally, a freely available shiny web application and code in the *R* programming language will be presented to help applied researchers estimate the upper bounds of their Pearson correlations and coefficient alpha to more carefully evaluate the internal consistency of their scales.

Results

Bounds on Coefficient Alpha: The Discrete Case

As pointed out in Novick and Lewis (1967), Lord and Novick (1968), and Raykov (2012), the original derivation of coefficient alpha does not invoke any distributional assumptions on the data. Although this is true, working through the computation of alpha in terms of the correlation matrix will help highlight how the distribution of the items still plays a role in shrinking the theoretical range that alpha may exhibit.¹

In the correlation metric, an alternative definition of Cronbach's alpha can be found in Kelley (1927, equation 7):

$$\alpha = \frac{A\bar{\rho}}{1 + (A - 1)\bar{\rho}}, \quad (3)$$

where A is the number of items and $\bar{\rho}$ is the average population interitem correlation.

Consider two jointly distributed Bernoulli random variables with marginal probability distributions $X \sim \text{Bern}(p_x)$ and $Y \sim \text{Bern}(p_y)$. It is well-known that the Pearson correlation between X and Y (more commonly known as the phi coefficient) is bounded below by $\max \left[-\sqrt{\frac{p_x p_y}{(1-p_x)(1-p_y)}}, -\sqrt{\frac{(1-p_x)(1-p_y)}{p_x p_y}} \right]$ and bounded above by $\min \left[\sqrt{\frac{p_x(1-p_y)}{(1-p_x)p_y}}, \sqrt{\frac{(1-p_x)p_y}{p_x(1-p_y)}} \right]$ as shown in Ferguson (1941) and Guilford (1965). Only when $p_x = p_y = .5$ does the Pearson correlation fully span its theoretical $[-1, +1]$ range, which is why a commonly recommended strategy is to divide the phi coefficient by its theoretical maximum (or minimum) value to mitigate the shrinkage effect of the marginal distributions on this correlation metric (Davenport & El-Sanhuray, 1991).

The fact that the marginal proportions of dichotomous items can influence the estimation of the Kuder–Richardson Formula 20 (KR-20) is not a new idea. Brogden (1946) was among the first researchers to raise awareness of this issue and, based on his observations, Sun et al. (2007) developed SPSS and SAS macros that allow applied researchers to calculate the upper bound of this reliability metric to help them

interpret their results. In a similar vein, this section attempts to generalize the results from binary items to items with multiple categories by highlighting the fact that the bounds of the phi coefficient obtained by Ferguson (1941) and Guilford (1965) are, in fact, the Fréchet–Hoeffding bounds for two jointly distributed Bernoulli random variables (Demirtas, 2016; Demirtas & Hedeker, 2011).

Let (X, Y) be two jointly distributed discrete random variables where X has k categories, with associated probabilities p_i , and Y has m categories with associated probabilities q_j . Moreover, (X, Y) can be arranged as a two-way contingency table with $k \times m$ cells. Define π_{ij} as the probability of belonging to cell (i, j) such that $Pr(x=i, y=j) = \pi_{ij}$. Then it follows that the maximum, positive Pearson correlation between X and Y is given by

$$\rho_{max} = \frac{\mathbb{E}[XY]_{max} - \left(\sum_{i=1}^k p_i \cdot i \right) \left(\sum_{j=1}^m q_j \cdot j \right)}{\sqrt{\left(\sum_{i=1}^k p_i \cdot i^2 - \left(\sum_{i=1}^k p_i \cdot i \right)^2 \right) \left(\sum_{j=1}^m q_j \cdot j^2 - \left(\sum_{j=1}^m q_j \cdot j \right)^2 \right)}}, \quad (4)$$

where $\mathbb{E}[XY]_{max}$ is defined as

$$\mathbb{E}[XY]_{max} = \min \left(\sum_{i=1}^k \frac{i(i+1)}{2} p_i + \sum_{j=1}^m \frac{j(j-1)}{2} q_j, \sum_{j=1}^m \frac{j(j+1)}{2} q_j + \sum_{i=1}^k \frac{i(i-1)}{2} p_i \right). \quad (5)$$

By plugging Equations 4 and 5 in Equation 3, one can derive the maximum bound for Cronbach's alpha on the correlation metric as

$$\alpha_{max} = \frac{A}{\frac{1}{\frac{\rho_{max1,2} + \dots + \rho_{maxA,A-1}}{A(A-1)/2}} + A - 1}. \quad (6)$$

A proof for Equations 3 to 5 is presented in the appendix.

Bounds on Coefficient Alpha: The Continuous Case

From the general form of the Fréchet–Hoeffding bounds shown in Equation 2, one can also obtain the maximum possible correlation (or covariance) between two continuously distributed random variables as long as their marginal CDFs $F(x)$, $G(y)$ can be evaluated. Notice that these bounds are not restricted to continuous-by-continuous or discrete-by-discrete associations. The Fréchet–Hoeffding bounds can also be obtained for continuous-by-discrete associations, such as those described by the biserial or polyserial correlations, but they will not be considered in this article.

Contrary to the discrete case, general analytic expressions for these bounds cannot always be obtained and need to be considered on a case-by-case basis. A good

candidate to exemplify the role that these bounds play in limiting the association between continuous random variables is the log-normal distribution, whose correlational bounds have closed forms. Consider $Z_1 \sim \mathcal{N}(0, 1)$ and $Z_2 \sim \mathcal{N}(0, 1)$. Define $X = e^{Z_1}$ and $Y = e^{Z_2}$, where e is the base of the natural logarithm (i.e., Euler's number). Then it follows that X and Y are log-normally distributed with population mean $\sqrt{e}$ and population variance $e^2 - e$. Astivia and Zumbo (2017) and Demirtas and Hedeker (2011) show that, in this case, the Pearson correlation coefficient between X and Y is bounded as

$$-1/e \leq \rho \leq 1, \quad (7)$$

if both variables are skewed in the same direction. If they are skewed in opposite directions (i.e., one is the mirror image of the other), the bounds switch and the feasible correlation range can only exist in the interval $[-1, 1/e]$. Just as with the discrete case, an upper bound on alpha can be calculated if the marginal distributions of the continuous items are known by obtaining each maximum correlation individually and plugging them into Equation 6.

Empirical Simulations

The results obtained in the Results section are now presented within the context of simulations to highlight the small-sample behavior of the upper bounds to Cronbach's alpha. Although it is well-known that alpha is usually defined within the framework of structural equation modeling, the following simulations will only consider population correlation matrices where compound symmetry holds so that the case can be made that Cronbach's alpha is, in fact, an estimate of the internal reliability of a test and not just the lower bound (Barchard & Hakstian, 1997; Jöreskog, 1971).

For the continuous case, consider a hypothetical test with four items, where each item follows a log-normal distribution as described in the "Bounds on Coefficient Alpha: The Continuous Case" section. To highlight the use of the bounds, Items 1 and 2 are skewed to the right and Items 3 and 4 are mirror images of them, so that they are skewed to the left. To allow control over the exact distributional shape of the lower dimensional marginals and the association among them, data are simulated from a Gaussian copula with parameter $\rho = .5$, which implies a population $\alpha = .8$. For a more comprehensive description on Gaussian copula modeling and simulation, please see chapters 3 and 4 in Nelsen (2010).

Figure 1 shows a marginal density plot between simulated Items 1 and 3 with a sample size of 10,000. The regions with highest density are concentrated toward the upper left corner; density then decreases radially. Notice that the axes are not on the same scale to highlight the region with highest concentration of this bivariate distribution.

To highlight the role that the Fréchet–Hoeffding bounds play on the Pearson correlation, Figure 2 presents a sample path of these calculations at sample sizes from

Figure 1. Gaussian copula contour plot with lower dimensional log-normal marginals.

$N = 5$ to $N = 10,000$ in incremental steps of 5. Notice that there are no multiple replications at each sample size. This simulated process merely takes a random sample from the Gaussian copula defined above, calculates the Pearson correlation, and plots it as a function of N . There are two horizontal lines, a solid one at 0.5 and a dotted one at $1/e$. It becomes readily apparent that, at increasing sample sizes, the estimated Pearson correlation stabilizes around its Fréchet–Hoeffding upper bound and, if the sample size could be increased arbitrarily, it would eventually converge to it. Figure 3 conveys a similar message, albeit this time through the calculation of Cronbach’s alpha. Although there is more variability at each sample size associated with this coefficient, there is an approximate difference of 0.1 units between the original $\alpha = .8$ and what is being empirically estimated: $\alpha = .69951 \approx .7$.

A simulated example is provided in Figure 4 for the case of four binary items. For this scenario, data were generated according to the latent trait model where an unobserved continuous variable becomes discretized by the act of measuring it (Lord & Novick, 1968). The latent variable is assumed to be multivariate normal with population tetrachoric correlations of $\rho_{tet} = .6$, $\rho_{tet} = .8$, and $\rho_{tet} = .99$. For the items, the probabilities of endorsement were fixed at $p_1 = .62$, $p_2 = .21$, $p_3 = .84$, and $p_4 = .10$. Substituting these values in Equation 6 yields a maximum possible Cronbach’s alpha (KR-20 in this case) of $\alpha = .7042674 \approx .7$.

Figure 2. Sample path estimates of upper bound to correlation for the case of log-normal marginals. Solid line marks the theoretical population $\rho = .5$ and dotted line the maximum upper bound at $\rho = 1/e$.

Three (large, positive) latent correlations were chosen to highlight the fact that, irrespective of how high or low the latent correlation may be, the Fréchet–Hoeffding bounds (which operate exclusively on the marginals) still apply and place a theoretical maximum that is independent of the actual internal structure underlying the test. Although the tetrachoric correlations were large, the process of discretization tends to reduce its size in absolute value (Bollen, 1989; Greer et al., 2003). This further reduces the estimated Cronbach's alpha, which is calculated on the observed binary variables and a direct function of the phi coefficients. For the cases of $\rho_{tet} = .6$ and $\rho_{tet} = .8$, the calculated alphas do not cross the theoretical upper bound, although the maximum possible reliability is still somewhat far from the theoretical one. This emphasizes why the largest tetrachoric correlation of $\rho_{tet} = .99$ is such a particular special case. This extreme scenario helps emphasize the fact that, irrespective of the strength of associations in the internal structure of the test, the maximum (or minimum) allowable correlations of the observed binary variables still hold as prescribed by the Fréchet–Hoeffding upper bound. If the latent correlation were set at unity, the resulting phi correlation would reach its theoretical upper bound. Generally speaking, the more asymmetrical the probabilities of each item category are, the narrower the

Figure 3. Stochastic process plot of upper bound to Cronbach’s alpha for the case of log-normal marginals. Solid line marks the theoretical population $\alpha = .8$ and dotted line the maximum upper bound at $\alpha = .7$.

valid correlation range becomes. For instance, consider two binary items, X and Y , where X has an endorsement probability of $p_x = 0.9$ and Y has an endorsement probability of $p_y = .1$. Through Equation 4 we can deduce the upper Fréchet–Hoeffding bound so that all positive phi coefficients can only span the range $[0, 0.111 \dots]$, given those marginal distributions for X and Y .

Distributional Assumptions and Cronbach’s Alpha: Revisiting Bay (1973)

Bay (1973) argues that the distribution of what he calls “part tests” (i.e., the items) has an influence both in the proper estimation of coefficient alpha and in the inferential procedures surrounding it, particularly the width of its confidence intervals. Because his computer simulations rely on the use of well-known distributions, Table 1 presents a sample of the combination of distributions he used, the Fréchet–Hoeffding bounds on the correlations implied by them, and the maximum possible Cronbach’s alpha. The approximated Fréchet–Hoeffding bounds were obtained using a modified version of the GSC algorithm introduced in Demirtas and Hedeker

Figure 4. Maximum observed Cronbach's alpha (KR-20) as a function of the tetrachoric correlation and marginal binary distributions. Population latent correlations of $\rho_{tet} = .6$, $\rho_{tet} = .8$, and $\rho_{tet} = .99$ are used. The solid, black line represents the upper bound on observed alpha at $\alpha \approx .7$.

(2011). In its original conceptualization, the GSC algorithm simulates a large sample ($N = 10^5$) of two distributions characterized by the third-order polynomial method (Fleishman, 1978). This method to generate nonnormal random variables allows the user to specify population values of skewness and excess kurtosis within certain limitations (Headrick, 2002). After the data are simulated, GSC proceeds by sorting the values of both variables from smallest to largest and calculates the Pearson correlation between them. Then it switches the ordering of one variable (so it is now sorted from largest to smallest whereas the other one remains sorted from smallest to largest) and calculates the Pearson correlation on this new pair. The first Pearson correlation becomes an approximation to the upper Fréchet–Hoeffding bound and the second correlation approximates the lower one. A similar approach can be taken if the marginal distributions can be simulated, and since Bay (1973) used three common ones (uniform, normal, and exponential), these distributions were used as opposed to the ones generated by the third-order polynomial method. Bay (1973) does not provide the parameters for these distributions, so $\mathcal{U}(0, 1)$, $\mathcal{N}(0, 1)$, and $\text{Exp}(1)$ were assumed.

Table 1. Fréchet–Hoeffding Bounds and Plausible Cronbach’s Alpha Range for Six Combinations of Distributions Investigated in Bay (1973).

Distribution	Fréchet–Hoeffding bounds	α range
Uniform–uniform	(−1, +1)	(0, 1)
Normal–normal	(−1, +1)	(0, 1)
Normal–uniform	(−1, +1)	(0, 1)
Exponential–exponential	(−0.645, +1)	(0, 1)
Exponential–normal	(−0.902, 0.902)	(0, 0.987)
Exponential–uniform	(−0.865, 0.865)	(0, 0.981)

Note. Bounds are rounded to three decimal points.

Table 1 presents a subsection of the distributions used by Bay (1973) in his simulation conditions reported on table 2, page 54 of his article. Only six combinations were used to highlight the role that the bound reduction has on the estimation of Cronbach’s alpha in simulated tests of eight items (the original study only considered this number of items). The top three distributions (combination of uniform marginals, normal marginals, or normal with uniform marginals) span the full theoretical correlation range (from −1 to +1) and theoretical alpha range (from 0 to +1). The bottom three distributions (exponential with exponential, exponential with normal, and exponential with uniform) show the most restricted ranges for both the correlation coefficient and alpha. Although there is no one-to-one correspondence, it is important to point out that the conditions where Bay (1973) found the strongest bias due to the distributional shape are also the conditions where the correlation range is most restricted.

To offer a more comprehensive overview of the interplay between marginal distributions, correlation, and coefficient alpha ranges, Table 2 presents approximated ranges for standard normal marginals and beta distributions with varying shape parameters to match population skewness and excess kurtosis values. Following the design in Bay (1973), simulated tests with eight items are considered. Both positive and negative values of higher order moments are showcased to obtain a more detailed view of how marginal distributions can influence range restriction. Overall, a similar pattern to the one observed in Table 1 presents itself here. Increasing values of higher order moments are associated with a more severe compression of the correlation range, which indirectly influences the plausible values that coefficient alpha can take.

Concluding Remarks and Future Directions

Although the analytical and computer simulations studies of Cronbach’s coefficient alpha permeate the psychometric literature, there still is a long way to go before one can fully understand the theoretical properties of this (and many other) statistics in

Table 2. Fréchet–Hoeffding Bounds and Plausible Cronbach’s Alpha Range for Six Combinations of Distributions.

Skewness/kurtosis	Fréchet–Hoeffding bounds	α range
0/–2	(–0.805, 0.805)	(0, 0.971)
0/–1	(–0.987, 0.987)	(0, 0.998)
–1/0	(–0.931, 0.931)	(0, 0.991)
–2/3	(–0.791, 0.791)	(0, 0.968)
3/8	(–0.649, 0.649)	(0, 0.903)
4/15	(–0.541, 0.541)	(0, 0.883)

Note. Correlation bounds have one marginal fixed at standard normal and the other at beta with skewness and kurtosis. Alpha range assumes tests with eight items. Bounds are rounded to three decimal points.

order to both reconcile seemingly disparate simulation results and offer best-practice recommendations. The idea that this reliability coefficient relies on any particular distributional assumptions, we believe, deserves a much more nuanced understanding from what is currently available in the published literature. Although it is true that, as mentioned in Raykov (2012) and Raykov and Marcoulides (2019), coefficient alpha (and reliability, in general) *is* defined irrespective of the distributional form of the data, the idea of the Fréchet–Hoeffding bounds introduces an intimate connection between the joint and marginal distributions that requires further exploration. The results from this article should not be interpreted as making the case that coefficient alpha is appropriate only for certain parametric families or distributional forms. Rather, it attempts to elucidate how the information contained within the joint probability distribution is bounded by the lower dimensional marginals and, since reliability attempts to quantify information pertaining to the joint distribution (see Markon, 2017), this same information must obey the restrictions implied by the Fréchet–Hoeffding bounds.

Calculating reliability when the data are ordinal is still an active area of research and not without its controversies (Chalmers, 2018; Zumbo et al., 2007; Zumbo & Kroc, 2019). Although a critical reader may wonder whether there is any value in treating ordinal data as continuous given the well-known drawbacks of this practice (Liddell & Kruschke, 2018), the main aim of this article is to situate the theoretical developments presented herein within the context of what researchers commonly do. Treating ordered, categorical, Likert-type responses as proxies for continuous quantities is standard when doing any type of item analysis or investigating the psychometric properties of a scale (Bishop & Herron, 2015; Carifio & Perla, 2007; Rhemtulla et al., 2012). Although we empathize with the idea of promoting “best practices” when doing data analysis, we would also like to provide useful guidelines and cautionary details that researchers may find useful when interpreting their results. Although there is nothing new about the idea that discrete data tend to yield smaller effect sizes (particularly covariances and correlations), what we consider

important is that we can characterize these smaller effect sizes by positing reasonable upper bounds on them. Moreover, as presented in the “Bounds on Coefficient Alpha: The Continuous Case” and “Empirical Simulations” sections these restrictions on effect sizes occur irrespective of whether the data are continuous or discrete. These are exclusively in terms of the CDFs of the marginal distributions and not the continuity (or lack thereof) of the scale or test.

We acknowledge that the idea of “range restriction” may also sound appealing to explain these limitations on effect sizes, but we would like to remind the reader that range restriction is neither necessary nor sufficient to reduce the correlation, even though it is popularly portrayed as such. In fact, Levin (1972) presents a counterexample where the restriction of range leads to an increase, rather than a decrease, in correlation. The Fréchet–Hoeffding bounds are much more general as they only require finite second moments and do not change unless the lower dimensional marginal distributions are transformed. A case could also be made from a pragmatic point of view—that, irrespective of how restricted the correlation/covariance range is, as the number of items keeps increasing the theoretical range of alpha keeps getting closer and closer to 1 (as long as the additional items do not only contain error variance). It just so happens that, if the full theoretical range is required, the number of items needed in the scale would change depending on their distribution. If they are skewed and discrete, then more items would be needed than if they were symmetric and continuous. That is certainly true but what we offer here is a way to characterize the rate of increase to reach the theoretical range of reliability $[0, 1]$, that is, how many more items would be needed to obtain a reasonable range of alpha. The estimation of Cronbach’s alpha now becomes more complex because it is no longer just in terms of the factor test structure or the ratio of true to error variance. The distribution of each marginal (i.e., the items) now comes into play to help inform the calculation and interpretation of the internal consistency of the test.

There are several future directions where one can extend this theoretical framework to help researchers interpret their effect sizes and complement best practices in data analysis. The theoretical framework elucidated in the appendix holds irrespective of the type of discrete random variable. We believe these results offer an improvement in interpretation, moving over simply stating that the effect sizes are generally low to providing an actual way of calculating how low they can be. Although this operates exclusively at the population level, one could imagine placing confidence or credible intervals on the sample Fréchet–Hoeffding bounds so that researchers can obtain a better idea of how high or how low their correlation-based effect sizes are expected to be. Although not elaborated further in this article, another future direction would be understanding and more carefully exploring how the bounds behave for the case of continuous random variables. Although the GSC algorithm proposed by Demirtas and Hedeker (2011) is a good first step, this method still suffers from the solution multiplicity problem of third-order polynomial transformations as presented in Astivia and Zumbo (2018, 2019). Different solutions to the Fleishman polynomials result in different bounds, so that two distributions could potentially have the same

exact four moments yet different correlation bounds. In general, it is more complicated to build a practical set of general recommendations that encompasses a wider range of continuous marginal distributions given that the bounds rely on the computation of their CDFs and need to be treated on a case-by-case basis. Consider, for instance, the results presented in Sheng and Sheng (2012) regarding the bias of Cronbach's alpha for negative excess kurtosis. Results from Table 1 consider the uniform distribution that also exhibits negative excess kurtosis ($-6/5$ to be precise), yet the approximated range spans the full theoretical alpha range of $(0, 1)$. Although seemingly contradictory, it is important to keep in mind that the data-generating process in Sheng and Sheng (2012) sets nonzero population values of the fifth and sixth moments (the c_4 and c_5 coefficients in Sheng and Sheng, 2012, equation 8, p. 3). This is not the case for the uniform distribution considered in the present article, which suggests the question of how much features of the distribution beyond excess kurtosis also play a role on the potential range restriction of Cronbach's alpha. It would perhaps be reasonable to consider the possibility of fitting different types of distributions to the available data and deriving the bounds once well-fitting distributions are obtained.

Finally, it is our hope that this article will help motivate applied researchers and quantitative social scientists alike to keep in mind the type of joint and marginal distributions they work with in an attempt to help promote more awareness of how these interact with common statistics found in day-to-day data analysis. The generality of the Fréchet–Hoeffding bounds implies that almost anything that operates on the correlation or the covariance matrix (from beta coefficients in multiple regression to loadings in factor analysis) is bounded. Certain effect sizes are simply not theoretically feasible (assuming certain marginals are fixed), and we think that, by acknowledging this, we can start creating more sensible guidelines to interpret effect sizes. A Pearson correlation of 0.5 is considered a large effect size (Cohen, 1988), yet (as shown in the “Bounds on Coefficient Alpha: The Continuous Case” section) if the marginal distributions are log-normal and opposite to each other in terms of skew, the maximum possible correlation is $1/e \approx 0.37$. What can be said about psychometric phenomena or correlational designs that rely on log-normal or other skewed distributions then? Perhaps it is now time to start bringing in advances in statistical theory from other areas to help us interpret these kinds of data in a more sophisticated manner.

Appendix

Setup, Definitions, and Relevant Properties

To restate the problem from the Results section, let (X, Y) be two jointly distributed discrete random variables, where X has k categories, with associated probabilities p_i , and Y has m categories with associated probabilities q_j . For notational convenience, we define the category count of X to begin at 1, not at 0. Therefore, any of the k categories take their values from the set $\{1, 2, \dots, k\}$. A similar setting also applies to

the m categories of Y . Moreover, (X, Y) can be arranged as a two-way contingency table with $k \times m$ cells, where (without loss of generality) X corresponds to the rows and Y to the columns. Define π_{ij} as the probability of belonging to cell (i, j) such that $Pr(x=i, y=j) = \pi_{ij}$. Remember that any row sum is equal to the marginal probability of x being equal to i : $\sum_{j=1}^m \pi_{ij} = p_i$ and, similarly, any column sum is equal to the marginal probability of y being equal to j : $\sum_{i=1}^k \pi_{ij} = q_j$.

Since all probabilities π_{ij} are nonnegative, we know that any subset of the above sum is less than or equal to the whole sum. This implies,

$$\sum_{i \in I} \pi_{ij} \leq \sum_{i=1}^k \pi_{ij} = q_j (*),$$

where $I \subseteq \{1, \dots, k\}$. Also note that the total of all cells sums up to 1:

$$\sum_{i=1}^k \sum_{j=1}^m \pi_{ij} = \sum_{i=1}^k p_i = \sum_{j=1}^m q_j = 1.$$

The expected value of X is

$$\mathbb{E}[X] = \sum_{i=1}^k p_i \cdot i.$$

The variance is

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sum_{i=1}^k p_i \cdot i^2 - \left(\sum_{i=1}^k p_i \cdot i \right)^2.$$

And, finally, the covariance between X and Y is obtained as follows:

$$Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \sum_{i=1}^k \sum_{j=1}^m \pi_{ij} \cdot i \cdot j - \left(\sum_{i=1}^k p_i \cdot i \right) \left(\sum_{j=1}^m q_j \cdot j \right).$$

Using the definition of the Pearson correlation coefficient, $\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$, we obtain

$$\rho_{XY} = \frac{\sum_{i=1}^k \sum_{j=1}^m \pi_{ij} \cdot i \cdot j - \left(\sum_{i=1}^k p_i \cdot i \right) \left(\sum_{j=1}^m q_j \cdot j \right)}{\sqrt{\left(\sum_{i=1}^k p_i \cdot i^2 - \left(\sum_{i=1}^k p_i \cdot i \right)^2 \right) \left(\sum_{j=1}^m q_j \cdot j^2 - \left(\sum_{j=1}^m q_j \cdot j \right)^2 \right)}}.$$

The Upper Bound for the Pearson Correlation Coefficient

Recall from (*) in the previous section that any subset of a row (column) sum of π_{ij} is less than or equal to the row (column) total. This is also true for a subset of size 1: $\pi_{ij} \leq p_i$ (for partial row sums) and $\pi_{ij} \leq q_j$ (for partial column sums), which naively recovers the upper Fréchet–Hoeffding bound as well (i.e., $\pi_{ij} \leq \min\{p_i, q_j\}$).

However, using this inequality would be inefficient and cause redundantly large bounds. Take the case of $k=5$. Although the inequality $\pi_{11} \leq p_1$ is true, the following inequality is also true: $\pi_{11} + \pi_{12} + \pi_{13} \leq p_1$, a much tighter bound. To find a tighter upper bound then, one can extract the largest possible partial sums for

$\mathbb{E}[XY] = \sum_{i=1}^k \sum_{j=1}^m \pi_{ij} \cdot i \cdot j$ and bound them from above using their row or column sums.

Notice that, since i and j are natural numbers, they can take values greater than 1 if $m, k \geq 2$. Also, notice that one can take either i or j out of the summation in the other index such that

$$\mathbb{E}[XY] = \sum_{i=1}^k \sum_{j=1}^m \pi_{ij} \cdot i \cdot j = \sum_{i=1}^k i \left(\sum_{j=1}^m \pi_{ij} \cdot j \right) = \sum_{j=1}^m j \left(\sum_{i=1}^k \pi_{ij} \cdot i \right). \quad (8)$$

Since the equality above is symmetric for both i and j , the following argument will only consider the case when i is taken out first, since one can simply mirror the notation in the end result to illustrate the two possible outcomes.

The definition of $\mathbb{E}[XY]$ is now

$$\mathbb{E}[XY] = \sum_{i=1}^k i \left(\sum_{j=1}^m \pi_{ij} \cdot j \right) = \sum_{i=1}^k i \left(\sum_{j=1}^m \pi_{ij} + \sum_{j=2}^m \pi_{ij} \cdot (j-1) \right). \quad (9)$$

The first summation in the parentheses is a partial row sum for row i . So, in particular, (*) holds and we find

$$\mathbb{E}[XY] \leq \sum_{i=1}^k i \left(p_i + \sum_{j=2}^m \pi_{ij} \cdot (j-1) \right) = \sum_{i=1}^k i p_i + \sum_{i=1}^k \sum_{j=2}^m \pi_{ij} \cdot i \cdot (j-1). \quad (10)$$

Since these sums are over a finite number of terms, and thus necessarily (absolutely) convergent, Fubini's theorem allows us to switch the order of summation at will. To wit,

$$\mathbb{E}[XY] \leq \sum_{i=1}^k i p_i + \sum_{j=2}^m (j-1) \sum_{i=1}^k \pi_{ij} \cdot i. \quad (11)$$

We can now repeat this maneuver in the final sum of Equation 11 by once again applying Equation 8 to find

$$\mathbb{E}[XY] \leq \sum_{i=1}^k ip_i + \sum_{j=2}^m (j-1) \left(\sum_{i=1}^k \pi_{ij} + \sum_{i=2}^k \pi_{ij} \cdot (i-1) \right). \quad (12)$$

Note that $\sum_{i=1}^k \pi_{ij}$ is a partial column sum and is less than or equal to the column total q_j by (*). Plugging this in gives

$$\mathbb{E}[XY] \leq \sum_{i=1}^k ip_i + \sum_{j=2}^m (j-1)q_j + \sum_{j=2}^m \sum_{i=2}^k \pi_{ij} \cdot (i-1) \cdot (j-1). \quad (13)$$

Repeating this procedure for i and j sequentially results in a unit increase in starting index and a unit decrease in the corresponding coefficient of π_{ij} at each step. Finally, we are left with the following:

$$\begin{aligned} \mathbb{E}[XY] \leq & \sum_{i=1}^k ip_i + \sum_{i=2}^k (i-1)p_i + \cdots + \sum_{i=k-1}^k 1 \cdot p_i + \sum_{j=2}^m (j-1)q_j \\ & + \sum_{j=3}^m (j-2)q_j + \cdots + \sum_{j=m}^m 1 \cdot q_j. \end{aligned} \quad (14)$$

Notice two important facts. First, that in every next sum, we have one less element of addition, that is, only the first summation contains p_1 , only the first two summations contain p_2 (where first sum contains two p_2 s and the second contains just one p_2), and so on. Second, that p_{k-1} is contained in every summation (first summation contains $(k-1)p_{k-1}$, second contains $(k-2)p_{k-2}$, etc.).

Following this pattern, one can see that every p_i is contained $1 + 2 + \cdots + i = \frac{i(i+1)}{2}$ times totally. Similarly for q_j (which starts from one index further, and thus contains one less element in every step) one has $1 + 2 + \cdots + (j-1) = \frac{j(j-1)}{2}$ elements. This can be expressed as

$$\mathbb{E}[XY] \leq \sum_{i=1}^k \frac{i(i+1)}{2} p_i + \sum_{j=2}^m \frac{j(j-1)}{2} q_j, \quad (15)$$

which bounds $\mathbb{E}[XY]$ from above. Moreover, this bound is sharp because it reduces to equality whenever $X = Y$; that is, if $X = Y$ then $\mathbb{E}[XY] = \mathbb{E}[X^2]$, and

$$\sum_{i=1}^k \frac{i(i+1)}{2} p_i + \sum_{i=2}^k \frac{i(i-1)}{2} p_i = \sum_{i=1}^k i^2 p_i = \mathbb{E}[X^2].$$

One could have started the above construction from j instead of i in which case, by symmetry, the bound would have been

$$\mathbb{E}[XY] \leq \sum_{j=1}^m \frac{j(j+1)}{2} q_j + \sum_{i=2}^k \frac{i(i-1)}{2} p_i. \quad (16)$$

To ensure the tightest bound, one just takes the minimum of both cases:

$$\mathbb{E}[XY] \leq \min \left(\sum_{i=1}^k \frac{i(i+1)}{2} p_i + \sum_{j=2}^m \frac{j(j-1)}{2} q_j, \sum_{j=1}^m \frac{j(j+1)}{2} q_j + \sum_{i=2}^k \frac{i(i-1)}{2} p_i \right). \quad (17)$$

The bound in Equation 17 is now referred to as $\mathbb{E}[XY]_{\max}$. Substituting it in the definition of the Pearson correlation for discrete random variables, one obtains

$$\rho_{XY} \leq \frac{\mathbb{E}[XY]_{\max} - \left(\sum_{i=1}^k p_i \cdot i \right) \left(\sum_{j=1}^m q_j \cdot j \right)}{\sqrt{\left(\sum_{i=1}^k p_i \cdot i^2 - \left(\sum_{i=1}^k p_i \cdot i \right)^2 \right) \left(\sum_{j=1}^m q_j \cdot j^2 - \left(\sum_{j=1}^m q_j \cdot j \right)^2 \right)}}, \quad (18)$$

and, in this case, the right-hand side of Equation 18 is referred to as $\rho_{\max}$.

As a special case, consider $k=m=2$ for Bernoulli-distributed X and Y . Then for the first argument of $\mathbb{E}[XY]_{\max}$ one has

$$\begin{aligned} \mathbb{E}[XY]_{\max} &= \sum_{i=0}^1 \frac{i(i+1)}{2} p_i + \sum_{j=0}^1 \frac{j(j-1)}{2} q_j = \\ &= \frac{0(0+1)}{2} p_0 + \frac{0(0-1)}{2} q_0 + \frac{1(1+1)}{2} p_1 + \frac{1(1-1)}{2} q_1 = 0 + 0 + \frac{2}{2} p_1 + \frac{0}{2} q_1 = p_1. \end{aligned}$$

And in the second argument one obtains

$$\begin{aligned} \mathbb{E}[XY]_{\max} &= \sum_{j=0}^1 \frac{j(j+1)}{2} q_j + \sum_{i=0}^1 \frac{i(i-1)}{2} p_i = \\ &= \frac{0(0+1)}{2} q_0 + \frac{0(0-1)}{2} p_0 + \frac{1(1+1)}{2} q_1 + \frac{1(1-1)}{2} p_1 = 0 + 0 + \frac{2}{2} q_1 + \frac{0}{2} p_1 = q_1. \end{aligned}$$

By plugging these results in Equation 18, it results in

$$\rho_{XY} \leq \frac{\min(p_1, q_1) - p_1 q_1}{\sqrt{(p_1 - p_1^2)(q_1 - q_1^2)}}.$$

And since the minimum operator is first-order scale invariant over the positive real numbers, one can express it as

$$\rho_{XY} \leq \min \left(\frac{p_1 - p_1 q_1}{\sqrt{p_1(1-p_1)q_1(1-q_1)}}, \frac{q_1 - p_1 q_1}{\sqrt{p_1(1-p_1)q_1(1-q_1)}} \right).$$

Factor the numerators:

$$\rho_{XY} \leq \min \left(\frac{p_1(1-q_1)}{\sqrt{p_1(1-p_1)q_1(1-q_1)}}, \frac{q_1(1-p_1)}{\sqrt{p_1(1-p_1)q_1(1-q_1)}} \right).$$

And after canceling out equal terms and rationalizing the radicals, the final expression becomes

$$\rho_{XY} \leq \min \left(\sqrt{\frac{p_1(1-q_1)}{(1-p_1)q_1}}, \sqrt{\frac{q_1(1-p_1)}{p_1(1-q_1)}} \right),$$

which equals the upper bound derived in Ferguson (1941) and Guilford (1965) from the Results section.

The Upper Bound for Cronbach's Alpha

Since (standardized) Cronbach's alpha is a function of the nonredundant elements of the correlation matrix, define ρ_{max} as the Pearson correlation coefficient that achieves the maximum upper bound and plug it into Equation 3. Now, for A items, the following inequality follows:

$$\alpha_{std} = \frac{A}{\frac{1}{\frac{\rho_{1,2} + \dots + \rho_{A-1,A}}{A(A-1)/2}} + A - 1} \leq \frac{A}{\frac{1}{\frac{\rho_{max_{1,2}} + \dots + \rho_{max_{A,A-1}}}{A(A-1)/2}} + A - 1}. \quad (19)$$

And Equation 19 would be the maximum possible Cronbach's alpha exclusively in terms of the number of categories and marginal probabilities of each item.

Acknowledgment

The authors acknowledge the support of The University of British Columbia - Paragon Research Agreement.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID ID

Bruno D. Zumbo <https://orcid.org/0000-0003-2885-5724>

Note

1. A similar argument can be made using the covariance matrix for raw alpha, but results will be shown in the correlation metric to ease the presentation.

References

- Astivia, O. L. O. & Zumbo, B. D. (2017). Population models and simulation methods: the case of the Spearman rank correlation. *British Journal of Mathematical and Statistical Psychology*, 70, 347-367. <https://doi.org/10.1111/bmsp.12085>
- Astivia, O. L. O., & Zumbo, B. D. (2018). On the solution multiplicity of the Fleishman method and its impact in simulation studies. *British Journal of Mathematical and Statistical Psychology*, 71(3), 437-458. <https://doi.org/10.1111/bmsp.12126>
- Astivia, O. L. O., & Zumbo, B. D. (2019). A note on the solution multiplicity of the Vale–Maurelli intermediate correlation equation. *Journal of Educational and Behavioral Statistics*, 44(2), 127-143. <https://doi.org/10.3102/1076998618803381>
- Barchard, K. A., & Hakstian, A. R. (1997). The robustness of confidence intervals for coefficient alpha under violation of the assumption of essential parallelism. *Multivariate Behavioral Research*, 32(2), 169-191. <https://doi.org/10.1207/s15327906mbr32024>
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173-1182. <https://doi.org/10.1016/j.jom.2018.04.003>
- Bay, K. S. (1973). The effect of non normality on the sampling distribution and standard error of reliability coefficient estimates under an analysis of variance model. *British Journal of Mathematical and Statistical Psychology*, 26(1), 45-57. <https://doi.org/10.1111/j.2044-8317.1973.tb00505.x>
- Bishop, P. A., & Herron, R. L. (2015). Use and misuse of the Likert item responses and other ordinal measures. *International Journal of Exercise Science*, 8(3), 297-302.
- Bollen, K. (1989). *Structural equations with latent variables*. Wiley. <https://doi.org/10.1002/9781118619179>
- Brogden, H. E. (1946). The effect of bias due to difficulty factors in product-moment item intercorrelations on the accuracy of estimation of reliability by the Kuder–Richardson Formula Number 20. *Educational and Psychological Measurement*, 6(4), 517-520. <https://doi.org/10.1177/001316444600600408>
- Carifio, J., & Perla, R. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences*, 3(3), 106-116. <https://doi.org/10.3844/jssp.2007.106.116>
- Chalmers, R. P. (2018). On misconceptions and the limited usefulness of ordinal alpha. *Educational and Psychological Measurement*, 78(6), 1056-1071. <https://doi.org/10.1177/0013164417727036>
- Cohen, J. (1988). *Statistical power analysis for the social sciences*. Lawrence Erlbaum.

- Davenport, E. C., Jr., & El-Sanhury, N. A. (1991). Phi/phimax: Review and synthesis. *Educational and Psychological Measurement*, 51(4), 821-828. <https://doi.org/10.1177/001316449105100403>
- Demirtas, H. (2016). A note on the relationship between the phi coefficient and the tetrachoric correlation under nonnormal underlying distributions. *The American Statistician*, 70(2), 143-148. <https://doi.org/10.1080/00031305.2015.1077161>
- Demirtas, H., & Hedeker, D. (2011). A practical way for computing approximate lower and upper correlation bounds. *The American Statistician*, 65(2), 104-109. <https://doi.org/10.1198/tast.2011.10090>
- Feldt, L. S. (1965). The approximate sampling distribution of Kuder–Richardson reliability coefficient twenty. *Psychometrika*, 30(3), 357-370. <https://doi.org/10.1007/BF02289499>
- Ferguson, G. A. (1941). The factorial interpretation of test difficulty. *Psychometrika*, 6(5), 323-329. <https://doi.org/10.1007/BF02288588>
- Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43(4), 521-532. <https://doi.org/10.1007/BF02293811>
- Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont donnés. *Annales de l'Université de Lyon Section A: Sciences mathématiques et astronomie* [On correlation matrices with fixed margins. *Annals of the University of Lyon Section A: Mathematical sciences and astronomy*], 14, 53-77.
- Greer, T., Dunlap, W. P., & Beatty, G. O. (2003). A Monte Carlo evaluation of the tetrachoric correlation coefficient. *Educational and Psychological Measurement*, 63(6), 931-950. <https://doi.org/10.1177/0013164403251318>
- Guilford, J. P. (1965). The minimal phi coefficient and the maximal phi. *Educational and Psychological Measurement*, 25(1), 3-8. <https://doi.org/10.1177/001316446502500101>
- Headrick, T. C. (2002). Fast fifth-order polynomial transforms for generating univariate and multivariate non-normal distributions. *Computational Statistics and Data Analysis*, 40(4), 685-711. [https://doi.org/10.1016/S0167-9473\(02\)00072-5](https://doi.org/10.1016/S0167-9473(02)00072-5)
- Hoeffding, W. (1940). Scale-invariant correlation theory. In N. I. Fisher & P. K. Sen (Eds.), *The collected works of Wassily Hoeffding* (pp. 57-107). Springer. https://doi.org/10.1007/978-1-4612-0865-5_4
- Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement*, 60(4), 523-531. <https://doi.org/10.1177/00131640021970691>
- Joe, H. (2014). *Dependence modeling with copulas*. CRC Press. <https://doi.org/10.1201/b17116>
- Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36(2), 109-133. <https://doi.org/10.1007/BF02291393>
- Kelley, T. L. (1927). *Interpretation of educational measurements*. World Book.
- Kristof, W. (1963). The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts. *Psychometrika*, 28(3), 221-238. <https://doi.org/10.1007/BF02289571>
- Levin, J. (1972). The occurrence of an increase in correlation by restriction of range. *Psychometrika*, 37(1), 93-97. <https://doi.org/10.1007/BF02291414>
- Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, 79(2018), 328-348. <https://doi.org/10.1016/j.jesp.2018.08.009>
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

- Markon, K. E. (2017). *A generalized definition of reliability based on Lindley information*. <https://doi.org/10.31234/osf.io/vgpfh>
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412-433. <https://doi.org/10.1037/met0000144>
- Nelsen, R. B. (2010). *An introduction to copulas*. Springer Science & Business Media.
- Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32(1), 1-13. <https://doi.org/10.1007/BF02289400>
- Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, 21(2), 381-391. <https://doi.org/10.1086/209405>
- Raykov, T. (2012). Scale development using structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 472-492). Guilford Press.
- Raykov, T., & Marcoulides, G. A. (2019). Thanks coefficient alpha, we still need you!. *Educational and Psychological Measurement*, 79(1), 200-210. <https://doi.org/10.1177/0013164417725127>
- Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354-373. <https://doi.org/10.1037/a0029315>
- Sheng, Y., & Sheng, Z. (2012). Is coefficient alpha robust to non-normal data? *Frontiers in Psychology*, 3, Article 34. <https://doi.org/10.3389/fpsyg.2012.00034>
- Shultz, S. G. (1993). *A Monte Carlo study of the robustness of coefficient alpha* [Unpublished master's thesis]. University of Ottawa.
- Sun, W., Chou, C. P., Stacy, A. W., Ma, H., Unger, J., & Gallaher, P. (2007). SAS and SPSS macros to calculate standardized Cronbach's alpha using the upper bound of the phi coefficient for dichotomous items. *Behavior Research Methods*, 39(1), 71-81. <https://doi.org/10.3758/BF03192845>
- Zimmerman, D. W., Zumbo, B. D., & Lalonde, C. (1993). Coefficient alpha as an estimate of test reliability under violation of two assumptions. *Educational and Psychological Measurement*, 53(1), 33-49. <https://doi.org/10.1177/0013164493053001003>
- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, 6(1), 21-29. <https://doi.org/10.22237/jmasm/1177992180>
- Zumbo, B. D., & Kroc, E. (2019). A measurement is a choice and Stevens' scales of measurement do not help make it: A response to Chalmers. *Educational and Psychological Measurement*, 79(6), 1184-1197. <https://doi.org/10.1177/0013164419844305>