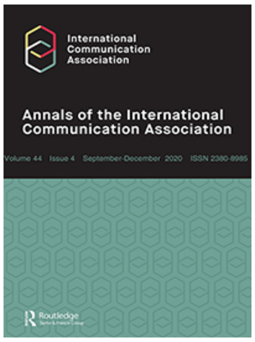

Annals of the International Communication Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rica20

Omega over alpha for reliability estimation of unidimensional communication measures

Alan K. Goodboy & Matthew M. Martin

To cite this article: Alan K. Goodboy & Matthew M. Martin (2020) Omega over alpha for reliability estimation of unidimensional communication measures, Annals of the International Communication Association, 44:4, 422-439, DOI: 10.1080/23808985.2020.1846135

To link to this article: https://doi.org/10.1080/23808985.2020.1846135

ANNALS OF THE INTERNATIONAL COMMUNICATION ASSOCIATION
2020, VOL. 44, NO. 4, 422–439
https://doi.org/10.1080/23808985.2020.1846135

Routledge
Taylor & Francis Group

REVIEW ARTICLE

Check for updates

Omega over alpha for reliability estimation of unidimensional communication measures

Alan K. Goodboy and Matthew M. Martin

Communication Studies, West Virginia University, Morgantown, WV, USA

ABSTRACT
Cronbach's alpha (coefficient α) is the conventional statistic communication scholars use to estimate the reliability of multi-item measurement instruments. For many, if not most communication measures, α should not be calculated for reliability estimation. Instead, coefficient omega (ω) should be reported as it aligns with the definition of reliability itself. In this primer, we review α and ω, and explain why ω should be the new 'gold standard' in reliability estimation. Using Mplus, we demonstrate how ω is calculated on an available data set and show how preliminary scales can be revised with 'ω if item deleted.' We also list several easy-to-use resources to calculate ω in other software programs. Communication researchers should routinely report ω instead of α.

ARTICLE HISTORY
Received 19 April 2020
Revised 26 October 2020

KEYWORDS
Reliability; Cronbach's alpha; McDonald's omega; Communication measurement

In graduate school we were both taught to always report Cronbach's alpha[1] (coefficient α; Cronbach, 1951) as a reliability estimate for our multi-item communication scales. We did what we were taught, and in fairness, we observed α reported in most of the published studies we read, so we assumed that α must be the best way to estimate scale reliability. In our own research, we have calculated and reported Cronbach's alpha in our studies with self-report communication measures, and like every other communication scholar we knew, we never thought to question why we did this or if what we were doing was the best practice. No editor or reviewer ever asked us to do otherwise. Unfortunately for us, and perhaps many other quantitative communication researchers, we were not properly educated about Cronbach's alpha and we used it blindly for years. To be frank, we now know there are better alternatives to alpha, and we believe it is time for a disciplinary change for how communication scholars estimate and report reliability of multi-item measurement instruments.

Cronbach's alpha has limited usefulness[2] for calculating reliability for most multi-item communication scales, because in general, alpha does not equal reliability (Raykov & Marcoulides, 2011). This is already well-understood in other social sciences, so we admit that what follows in this manuscript is nothing new to psychometricians. Indeed, there are numerous discussions on the limitations of alpha articulated elsewhere (see Bentler, 2009; Cho & Kim, 2015; Cortina, 1993; Crutzen & Peters, 2017; Gignac, 2014; Graham, 2006; Green & Hershberger, 2000; Green et al., 1977; Green & Yang, 2009; Hancock & An, 2018; McDonald, 1999; McNeish, 2018; Novick & Lewis, 1967; Raykov, 2001; 2019; Rodriguez et al., 2016; Savalei & Reise, 2019; Shevlin et al., 2000; Schmitt, 1996; Sijtsma, 2009; Trizano-Hermosilla & Alvarado, 2016; Zinbarg et al., 2005). Despite the availability of these in-depth and technical reviews of alpha's limitations in other sciences, these discussions of reliability

CONTACT Alan K. Goodboy agoodboy@mail.wvu.edu Communication Studies, West Virginia University, 108 Armstrong Hall, P.O. Box 6293, Morgantown, WV 26506–6293, USA
Supplemental data for this article can be accessed https://doi.org/10.1080/23808985.2020.1846135.

have not caught on (yet) in communication studies, evidenced by the fact that most quantitative scholars still report α in their published manuscripts (Hayes & Coutts, 2020). We hope to facilitate these discussions here, but to be fair, we are merely repeating what is known already and provide solutions that are already available in primers in other science disciplines (e.g. Komperda et al., 2018; McNeish, 2018; Viladrich et al., 2017). By bringing up these issues in a communication journal, we hope to encourage better practices for reporting reliability in multi-item communication scales. We review alpha's assumptions and limitations and then offer practical suggestions and solutions to replace the discipline's customary reliance on alpha as a reliability estimate. Ultimately, we argue that communication scholars should provide a calculation of reliability itself, which is coefficient omega, and we demonstrate how to calculate it with available data.

Reliability

It is important to define what is meant by the term 'reliability.' Under classical test theory (CTT; Lord & Novick, 1968; Novick, 1966), an observed score (X_{ij}) for a person (i) on an item (j) is the sum of two unobserved scores: the item true score (T_{ij}) and the item error score (E_{ij}), so that $X_{ij} = T_{ij} + E_{ij}$. That is, for an individual, the T is a constant and X and E are random variables, so that the error score for a measurement item is the difference between the observed score and the true score (Lord & Novick, 1968). Assuming that errors are uncorrelated, after obtaining a scale score, the variance of scores (${\sigma_X}^2$) have true variance (${\sigma_T}^2$) and error variance (${\sigma_E}^2$) partitions (${\sigma_X}^2 = {\sigma_T}^2 + {\sigma_E}^2$). In respect to these variance partitions, reliability (ρ_X), by definition, is the ratio of the true score variance to observed score variance ($\rho_X = {\sigma_T}^2/{\sigma_X}^2$). In other words, 'the square of the correlation between observed scores and true scores equals the ratio of the true-score variance to the observed score variance' (Lord & Novick, 1968, p. 57). Put simply, reliability is the percentage of true variance in observed variance. It is this definition of reliability ($\rho_X = {\sigma_T}^2/{\sigma_X}^2$) that we will return to later on as we discuss reliability estimation from a structural equation modeling framework using confirmatory factor analysis.

Limitations of alpha

Traditionally, and perhaps almost exclusively, communication scholars have calculated and reported coefficient (Cronbach's) alpha (α) for their reliability estimates, using all the items in a measurement instrument to calculate the mean of all split-half reliability coefficients (Cronbach, 1951). Researchers have circulated incorrect myths about alpha including that (1) alpha is equal to the reliability of a test score, (2) the value of alpha is independent of the number of items on a test, (3) alpha is an indication of the unidimensionality of a test score, (4) alpha is the best choice among reliability coefficients, (5) there is a particular level of alpha that is desired or adequate, and (6) if removing an item increases alpha, the test is better without that item (Hoekstra et al., 2019), among others. Some communication researchers might consider their measure to be 'reliable' if α exceeds some subjective cutoff value, such as .70, which has been popularized in communication scholarship as 'sufficiently high' (Johnson, 2017, p. 1415). However, this cutoff of $\alpha > .70$ for a multi-item scale is low, so scholars might require reliability to at least be .80 to be considered satisfactory (Raykov & Marcoulides, 2011), or perhaps even higher.

It is our opinion that reliability does not receive enough attention in published communication scholarship and that authors do not consider how substantial error variance in their measurements might appreciably attenuate their results. In published communication manuscripts, reliability is sometimes treated as follows: researchers (us included) report α as if it is reliability, are satisfied if α is .70 or higher (which is a considerable amount of error variance), and never mention reliability again unless it is considered a limitation in the discussion section (i.e. α was mentioned again as lower than desired). This fleeting mention of scale reliability poses a considerable problem in the statistical analyses that follow when reliability estimates are low. Communication scholars often use multiple summed scale scores (i.e. composite variables) from several communication measures in multivariate statistical analyses such as ordinary least squares regression, which assumes the predictors (*X1, X2 … Xk*) are all perfectly measured by the scales that are chosen (i.e. reliability = 1.0), when we know that reliability estimates are much lower. We need to pay more attention to reliability and ensure that it is satisfactory (e.g. perhaps .80+; Raykov & Marcoulides, 2011).

Indeed, in the social sciences, the average coefficient *α* for scales is low. Peterson (1994) conducted a meta-analysis of 4286 *α* coefficients in the behavioral sciences, revealing an average coefficient *α* of .77 (i.e. 23% of the average variance in these scales was measurement error). Peterson (1994) also found that 25% of these *α* coefficients were below .70. And although coefficient (Cronbach's) alpha has ruled the realm in communication measurement, scholars may not be aware of (or may ignore) the limitations of and restrictive conditions required for *α* to accurately estimate reliability. Consequently, they may provide estimates that are not reliability at all. Simply, they just report *α*, without realizing that it is not actually estimating true reliability in most studies featuring multi-item communication measurement instruments. For *α* to equal scale reliability, a unidimensional factor structure of the scale must be retained in a measurement model and the scale items must be at least essentially tau-equivalent with no correlated errors among items. If these restrictive conditions are unfulfilled in a measurement model, alpha will not equal reliability. We explain these ideas in more detail below.

Unidimensionality

Coefficient *α* rests on the assumption that a scale is unidimensional; that is, all of the items measure the same construct from a single factor. It is a widespread myth that *α* provides evidence for the unidimensionality of a communication measure and it is also untrue that a high alpha indicates that items are measuring the same construct. Alpha should not be used to estimate reliability of multidimensional scales (Gignac, 2014), so communication scholars who employ multidimensional scales (i.e. subscales for separate factors) typically compute *α* for each dimension. While unidimensionality is required for using *α*, obtaining a high value for *α* is *not* evidence for the dimensionality of a scale. Alpha cannot evaluate dimensionality (Green et al., 1977), and scales with many items (e.g. 20 items) tend to yield an 'acceptable' *α* because of the scale length (*α* in general becomes larger with more scale items of at least equal quality), even if the scale is treated as unidimensional when it is actually a multidimensional measurement instrument (Streiner, 2003).

Because dimensionality of a scale is poorly communicated by *α*, evidence of unidimensionality is obtained instead by testing a hypothesized measurement model using factor analysis and evaluating model fit to determine if the characteristics of the observed item score data align with a 1-factor specification. To assess the unidimensionality assumption for established scales, we recommend using confirmatory factor analysis (CFA) to test the hypothesis that all scale items load on a single latent variable (i.e. one latent variable is causing the shared variation in the scale items). It is important for scholars who calculate alpha to first test the dimensionality of their scales and ensure that scale items match the dimensional structure they are supposed to have. However, scale unidimensionality may be obtained in measurement models that impose various restrictions on item loadings and error variances (i.e. unidimensional measurement models may be more or less restrictive). Coefficient *α* makes another assumption within the unidimensional measurement model by requiring at least essential tau-equivalence.

Measurement models and essential tau-equivalency

Multi-item communication measures with a unidimensional factor structure may be considered *parallel, tau-equivalent,* or *congeneric* measurement models. As we discuss these models, we are reminded that under CTT, item scores are comprised of both true scores and error scores ($X_{ij} = T_{ij} + E_{ij}$). The *parallel* model is the most restrictive, assuming items 'have identical true scores and linearly experimentally independent errors having equal variances' (Lord & Novick, 1968, p. 47). From a factor-analysis perspective, this means that all the scale items have equal factor loadings (i.e. item 1 λ = item 2 λ = item 3 λ … = item k λ), and likewise, all error variances for each item are equal too (item 1

θ = item 2 θ = item 3 $\theta \ldots$ = item k θ). This model might sound unrealistic in practice because communication measures almost never consist of items that all measure the latent construct precisely to the same degree. Thus, the parallel model is unlikely to be tenable for most communication measures.

In CFA, to test the viability of the parallel model for a unidimensional scale, all factor loadings and all error variances are fixed to be equal and then model fit is assessed. To assess global fit in CFA (Brown, 2015), consult the model χ^2 (or scaled Satorra-Bentler or Yuan Bentler χ^2 with nonnormal data; Satorra & Bentler, 1994; Yuan & Bentler, 2000) and standardized root mean squared residual (SRMR), Steiger-Lind root mean square error of approximation (RMSEA) with its 90% confidence interval, and the Bentler Comparative Fit Index (CFI). To inspect local fit, examine the standardized, normalized, or correlation residuals (Goodboy & Kline, 2017). In most cases, the parallel model would provide poor model fit and be rejected because it is too restrictive.

Slightly less restrictive, a *tau-equivalent* model has items with 'the same true scores but (possibly) different error variances' (Lord & Novick, 1968, p. 47). From a factor-analysis perspective, this means that all the scale items still have equal factor loadings (i.e. scale items have equal slopes as manifestations of the latent variable being measured), but now, items could have different error variances. This model is important for coefficient *α,* which assumes essential tau-equivalence (sometimes called true score equivalence), meaning that the latent variable contributes the same amount of variance to all items; that is, all items' factor loadings are identical. Cortina (1993) explained that these measurements are linearly related and only differ by a constant (i.e. the true scores for any two scale items is within a constant of each other). As Graham (2006) explained, 'coefficient alpha assumes that all items measure the same latent trait on the same scale, with only variance unique to an item being comprised wholly of error' (p. 935). Like the parallel model, the tau-equivalent model may also be too restrictive for many communication measures. Any researcher who has factor-analyzed a scale before has probably witnessed in practice that factor loadings are generally not equal because some items are better explained by a latent construct than others (e.g. item 1 might have an unstandardized factor loading of .42 whereas item 2 might have a loading of .89).

To test a tau-equivalent model in CFA, a researcher would fix all factor loadings to be equal but freely estimate the error variances and then assess model fit. However, this model might provide poor fit because it is quite reasonable to expect that items do not produce equal factor loadings, especially when capturing the breadth of a communication construct. Communication scale items may tap into different facets of the same underlying latent variable, usually with different degrees of precision, because scholars who create these scales are tasked with ensuring that the full scope of the construct is sufficiently represented. During scale construction, researchers are not so concerned with their items sharing equal factor loadings. Communication scales are not necessarily designed to be tau-equivalent and quite often, we should expect some items to tap different aspects of a unidimensional construct better than other items. Items might be retained for the validity of test content even if they produce unequal factor loadings (e.g. .52) from other items (e.g. .92). Dropping items from a scale that produces lower factor loadings, but were important to include during the original formulation of a scale, can compromise the validity of the measure (e.g. evidence of test content) and is not a practice we recommend. The fact that coefficient alpha assumes an essentially tau-equivalent model is an issue for communication researchers who calculate alpha for scales that produce unequal factor loadings. Unfortunately, many communication researchers calculate alpha without providing evidence of unidimensionality in the first place, let alone additional evidence for meeting essential tau-equivalence. Calculating alpha without essential tau-equivalence will provide an inaccurate estimate of reliability which will be discussed shortly.

Most often, the *congeneric* model is appropriate for communication measures because it only assumes that scale items 'have linearly related true scores' (Jöreskog, 1971, p. 109). This model is the most flexible, meaning that items measure 'the same latent dimension in possibly different units of measurement and with possible different precision' (Raykov, 1997b, p. 174). In terms of factor analysis, items from a congeneric model may have different factor loadings and different error variances, which is the most realistic for communication measures. In CFA, a researcher would specify a unidimensional model where all factor loadings and all error variances are free to vary (i.e. unique values are estimated for each). In our experience, this is usually the single-factor model that is reported by researchers who publish unidimensional measurement models using CFA (i.e. a unidimensional CFA without imposing any equality constraints). If the congeneric model is retained because the data fit this model well, but then coefficient *α* is calculated from the scale items, a biased estimate of reliability will be reported. We speculate that *α* is calculated and reported quite often for communication scales that fit a congeneric model, or worse, are not unidimensional at all. Figure 1 displays and contrasts the parallel, tau-equivalent, and congeneric models from a CFA framework in terms of factor loadings and error variances.

Uncorrelated errors

Under CTT, 'it is assumed that errors of measurement are uncorrelated with each other and with all true scores' (Lord & Novick, 1968, p. 493). That is, item errors ($E = X - T$) are distributed independently. Coefficient *α* assumes uncorrelated errors between any two items, or as Cronbach (2004) stated, 'it is expected that the experience of responding to one part (e.g. one item) will not affect performance on any subsequent item' (p. 402). The uncorrelated errors assumption may be tested in CFA, and frequently by default, measurement models are specified to have uncorrelated errors (i.e. in CFA, the researcher would have to take the extra step to specify an error correlation and have a compelling reason to add that parameter and lose a degree of freedom).

It is possible, however, that communication measures have correlated errors that are not produced by random measurement error, but instead, can reflect reliable variance (Green & Hershberger, 2000). For example, correlated errors might arise from two (or more) scale items that share a common cause such as (a) being negatively-worded/reverse-coded items, (b) sharing similar idiosyncratic wording that prompts respondents to answer differently, or (c) tapping into the same construct but in a unique domain, among other reasons. When errors are correlated and coefficient *α* is calculated, the reliability estimate will be biased. Cronbach (2004) acknowledged that uncorrelated errors might not be tenable for some measures, stating that 'the assumption, like all psychometric assumptions, is unlikely to be strictly true' (p. 402) and he recommended 'when the problem is major, alpha simply should not be used' (p. 403).

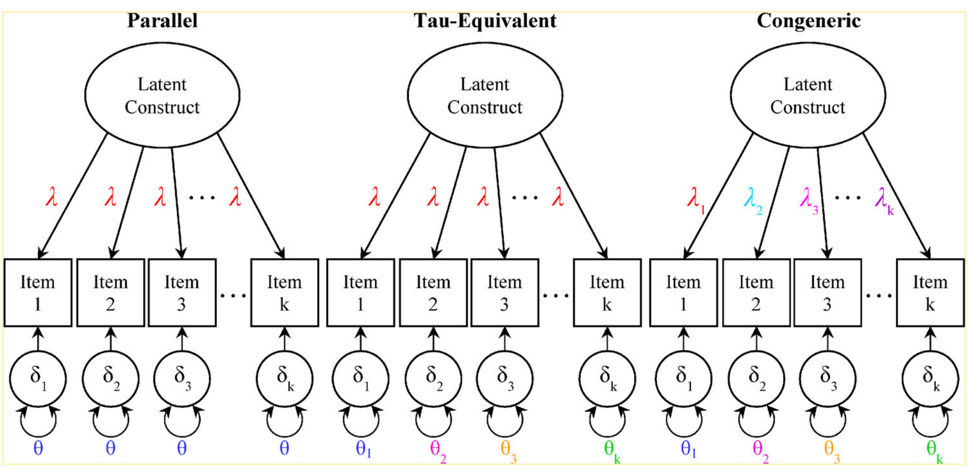

Figure 1. Unidimensional measurement models. Note: λ = unstandardized factor loading. θ = residual ariance. Factor loadings and residual variances with different subscripts (and different colors) are free to differ. Loadings and residual variances with no subscripts (and same color) are fixed to be equal. The (essentially) tau-equivalent model is required for coefficient alpha, which assumes equal factor loadings for every item.

Alpha ≠ reliability: violations of essential tau-equivalence and uncorrelated errors

What happens to alpha when there are violations of its restrictive assumptions? As previously mentioned, assuming a unidimensional factor structure holds, and measurement errors are uncorrelated, alpha will not be reliability unless item scores are at least essentially tau-equivalent (i.e. not congeneric). Alpha will underestimate reliability in a unidimensional factor model with uncorrelated errors when factor loadings are dissimilar and/or low (Raykov, 1997a; Raykov & Marcoulides, 2019). With uncorrelated errors and low and/or unequal factor loadings, alpha will be the lower bound of the true reliability value (Raykov, 1997a), so in these cases, communication scholars are underreporting the actual reliability of their scales.

However, when measurement errors are correlated, alpha is not necessarily the lower bound of reliability. In fact, alpha can either overestimate or underestimate the true reliability in both unidimensional and multidimensional factor models depending on the model parameters (Raykov & Marcoulides, 2011). If errors are positively correlated, α can overestimate reliability, depending on the number of scale items and the size of the error correlations (Zimmerman et al., 1993). Raykov (1997a) demonstrated that coefficient α will be different from scale reliability in the population (i.e. there is slippage; denoted as ε) as a function of (a) individual scale item violations of essential tau-equivalence, (b) error variances, (c) congeneric true variance, and (d) the length of the scale. Raykov showed that short scales (e.g. 2–4 items) with lower average factor loadings will yield spillage. Likewise, Raykov demonstrated with uncorrelated errors, a single item violating essential tau-equivalence will contaminate the estimate (see p. 346, Table 2 in Rakyov, 1997a for population discrepancies of 1 item violating essential tau-equivalence to various degrees).

However, coefficient α may violate essential tau-equivalence but still approximate reliability as long as items load uniformly high on the latent factor (e.g. factor loadings may be unequal as long as they all exceed .60+; see Raykov, 2012; Raykov & Marcoulides, 2011). Although alpha will equal reliability only under the special and restrictive case reviewed earlier, Raykov and Marcoulides (2015) explained that 'while in the general case coefficient alpha is not a consistent estimator of composite reliability and has a number of downsides, under unidimensionality and uncorrelated errors alpha can be very close to reliability at large if the average construct loading—given unitary latent variance—is in excess of .7 and the component loading deviations from it are within the interval (-.2, .2)' (p. 152). Coefficient alpha, then, might approximate reliability in well-established communication scales with uniformly high but unequal item loadings (e.g. .90, .69, .84, …).

But is a reliability estimate from alpha that is 'close enough' or 'approximate' to true reliability what we want to report in communication science? The answer to that question depends on how high the factor loadings are for each item, because if they are not high enough, they must be equal (i.e. essentially tau-equivalent). Communication measures may not have equal or uniformly high factor loadings for all items, especially for newer measures that have not generated ample validity evidence in independent samples (i.e. evidence of internal structure), or even older and more established communication measures (e.g. created in the 1980s) that predate the modern modeling techniques (i.e. were created without CFA) required to test the aforementioned assumptions in measurement models.

Alpha if item deleted ≠ reliability if item deleted

Another coefficient α related issue is that communication scholars use readily available features of statistics software to examine improvements in α by dropping items and recalculating α after items are removed. This item deletion process, often referred to as 'alpha if item deleted,' may improve the value of coefficient α, but it does so by capitalizing on sample specific variation in the data that may not replicate in another sample. Using 'alpha if item deleted' to drop scale items does not equate to reliability improvements at the population level, and even if the population α does improve from dropping an item, the actual scale reliability may be inaccurate because as we have reviewed, α is likely biased in the first place (Raykov, 2007). Even worse, trimming items to increase α may result in a (sometimes substantial) loss of criterion validity (see Raykov, 2008 for details). For these reasons, communication researchers should be critical of using α in the first place to estimate reliability, and if they do report α because its assumptions are fulfilled, they should discard the practice of using 'alpha if item deleted' to drop scale items.

The solution is omega

What should researchers use instead because α only provides reliability only under restrictive conditions? The solution is to instead calculate reliability itself (i.e. composite reliability[3]), which is coefficient omega (ω) as a point and interval estimate (McDonald, 1970, 1985, 1999). Coefficient ω is calculated using CFA with factor loadings and error variances in its equation (McDonald, 1999). Unlike α, ω does not require essential-tau equivalence for reliability to be accurate; it can be calculated using a congeneric model where factor loadings and error variances are free to be unequal as is the case for many communication measures. It is calculated as composite (construct) reliability that reflects the *true score variance/total observed variance* in a unidimensional scale (Dunn et al., 2014). Notice that this is the very definition of reliability reviewed earlier under CTT. McDonald (1999) provides the formula for coefficient omega as:

$$\omega = \frac{(\Sigma\lambda_j^2)}{[(\Sigma\lambda_j)^2 + \Sigma\theta_j]}$$

In this formula, the numerator is the sum of all scale items' unstandardized factor loadings squared $(\Sigma\ \lambda_j)^2$ and the denominator is the same sum of the items' unstandardized factor loadings squared $(\Sigma\ \lambda_j)^2$ added to the sum of the items' residual variances $(\Sigma\ \theta_j)$. McDonald (1999) noted that 'in a homogenous test, omega is both a reliability coefficient and a (construct) validity coefficient' (p. 208). To test for unidimensionality of a measure, a CFA is conducted and global fit statistics are reviewed (χ^2, RMSEA, CFI, SRMR) in conjunction with local fit (i.e. standardized, normalized, or correlation residuals), and if unidimensionality of the measure holds (i.e. most likely congeneric), then the point and interval estimate for ω is interpreted as reliability of the measure. The advantage of using this method is that reliability is not biased, despite how low, high, or unequal the factor loadings are, and the 95% confidence interval accompanies coefficient ω to give highly probable values of reliability in the population. If data are multivariate normal, a 95% bootstrapped confidence interval should be calculated using maximum likelihood (ML) estimation (Raykov & Marcoulides, 2011). If data are nonnormal, robust maximum likelihood should be used to accommodate nonnormal item distributions (Raykov & Marcoulides, 2016). We encourage researchers using CFA to habitually select robust ML as a default estimator because many communication measures do not meet the assumption of multivariate normality in ML. Robust ML estimators include MLM with the Satorra-Bentler scaled chi-square (S-Bχ^2; Satorra & Bentler, 1994) or MLR with the Yuan-Bentler residual-based chi-square (Y-Bχ^2, also known as T_2^* test statistic; Yuan & Bentler, 2000). Both robust ML estimators will incorporate a scaling correction factor in the assessment of model fit and will adjust the model χ^2 and standard errors to be accurate.

Although we recommend abandoning the practice of dropping items based on improvements in α ('alpha if item deleted'), we do recommend using the point and interval method of estimating ω for improving reliability (i.e. calculating ω after the deletion of an item). This is analogous to a procedure that could be labeled 'omega (reliability) if item deleted' but it does not suffer from the same shortcomings of deleting items to improve α. If measures need to be revised by deleting weak items to improve reliability (e.g. in the early stages of measurement development or revision), ω can be recalculated after each item is deleted to reduce error variance. Raykov and Marcoulides (2011) provided the procedures to accomplish this by examining weak loadings and error variances, which answers 'what happens to reliability when I drop this (potentially weak) item?' Coefficient ω can be examined after items are deleted from the measurement instrument, but ultimately, changes to an operationalization of a communication variable are better off made by respecting the theory or body of work that stimulated the measure development in the first place. The focus should not be making small improvements to reliability at the expense of removing a theoretically important item from a scale. This suggestion speaks to test content as a source of validity evidence. We demonstrate how to calculate omega if deleted on a scale in our tutorial section.

Alpha to omega

Should alpha be discarded altogether by communication scholars? Technically no, because calling for a moratorium on alpha would ignore the fact that alpha is still a dependable reliability estimate when its restrictive conditions are fulfilled, and when they are, α and ω will be identical as scale reliability (Raykov & Marcoulides, 2019). But communication scholars are rarely testing the assumptions of α. And if they did regularly test assumptions, they might find that α should not be calculated anyway for many scales. Therefore, we believe it is time to move on to ω. There is a chorus of scholars who agree.

Sijtsma (2009) criticized that 'in practice, alpha attains values that are outside the range of possible values of the reliability that can be derived from a single test administration' (p. 118). Cho and Kim (2015) stated 'alpha is a relatively inferior method despite its widespread use' (p. 224); 'substituting alpha with a superior alternative is not merely a matter of personal choice but a matter of academia consciously responding to the issue' (p. 225) … it would be prudent for the editors of various academic journals on organizational research to recommend that their contributors use superior alternatives with or in place of alpha in their works' (p. 225). McNeish (2018) summarized 'although Cronbach's alpha is familiar, commonly reported, and easy to obtain in software, it is rarely an appropriate measure of reliability–its assumptions are overly rigid and almost always violated' (p. 422). Peters (2014) noted simply that alpha 'is a fatally flawed estimate of its reliability' (p. 68).

Overall, the advantages of calculating ω as a point and interval estimate far outweigh the limitations of α (Dunn et al., 2014; Komperda et al., 2018; Zinbarg et al., 2005). Coefficient omega should now be the gold standard for reliability in communication studies. This replacement of reliability currency will take some time for the communication discipline to adopt. We hope that journal editors, reviewers, researchers, and educators agree to help us catch up with other social sciences where omega has already replaced alpha. We offer a tutorial below for calculating omega with a data set that is freely available at: [File = Annals Omega Data.dat]. We also recommend completing a digital module (free to register from the National Council on Measurement in Education) by Hancock and An (2018) to learn about coefficient ω featuring a module companion article, narrated video lecture, integrated slides, worked examples, analysis code, data, exercises, and so on at: https://ncme.elevate.commpartners.com/products/digital-module-02-scale-reliability-in-structural-equation-modeling

Tutorial: steps for calculating ω

Below we offer some advice for best practices when calculating and interpreting omega as scale reliability.

(1) Start with the measurement model first. Conduct a CFA using ML (if data are normal) or robust ML (MLM or MLR) estimation (if data are nonnormal) specifying a unidimensional factor structure for the scale or subscales (Crutzen & Peters, 2017; Green et al., 1977; Raykov, 1997a; 2001; Savalei & Reise, 2019; Yang & Green, 2011). All items should be specified to be caused by a single latent variable/factor (i.e. the single variable that the scale is designed to measure). Because omega accommodates a congeneric model, factor loadings can be unequal. Remember that calculating scale reliability requires that a 1-factor model adequately represents the data.

(2) Interpret the data to model correspondence by assessing global fit (i.e. overall fit of the model) and local fit (i.e. localized areas of misfit). To assess global fit in CFA (Brown, 2015), consult the model χ^2, (or scaled Satorra-Bentler or Yuan-Bentler χ^2 with nonnormal data), SRMR, RMSEA with its 90% confidence interval, and CFI. To inspect local fit, examine the standardized, normalized, or correlation residuals. Standardized and normalized residuals have significance tests to indicate misfit, whereas correlation residuals larger than an absolute value of .10 may indicate localized strain in the model (Kline, 2016). If the global and local fit statistics suggest that the unidimensional model be retained, the unstandardized factor loadings and residual variances are used to calculate a point and interval estimation of reliability, calculated as: $\omega = (\Sigma\ \lambda_j)^2/[(\Sigma\ \lambda_j)^2 + \Sigma\ \theta_j]$. Following these steps, we provide easy software solutions to estimate ω using this formula.

(3) Along with the point estimate of ω as scale reliability, a 95% bootstrapped confidence interval (CI) should be calculated if ML estimation is used with multivariate normal data (Padilla & Divers, 2016). If data are nonnormal, robust ML estimation is used (i.e. MLM with no missing data or MLR with missing data) and a CI (without bootstrapping) is calculated instead (bootstrapping is not conducted with robust ML). The advantage of providing confidence intervals is to assure readers that population reliability is highly probable to be within the lower and upper limits of the 95% CI. Most communication scholars who report reliability only provide the point estimate, but the CI is important to report too.

An example of ω in Mplus

Using the software Mplus 8.0 (Muthén & Muthén, 2017), we conducted a CFA with ML estimation on the Academic Entitlement Questionnaire (AEQ; Kopp et al., 2011) from a data set (Annals Omega Data.dat) provided by Goodboy and Frisby (2014). This scale consists of eight items and uses a 7-point Likert response format ranging from 1 (strongly disagree) to 7 (strongly agree). It measures ‘the expectation that one should receive certain positive academic outcomes (e.g. high grades) in academic settings, often independent of performance’ (Kopp et al., 2011, p. 106). In a sample of 222 students, Goodboy and Frisby (2014) reported a coefficient alpha of .71 for the scale, and unfortunately, they did not conduct a CFA to test for unidimensionality, nor did they impose the required equality constraint that the 8 factor loadings were equal (i.e. tau-equivalence) needed by alpha to be an appropriate estimate of reliability. Like many other communication scholars then and now, they just simply reported alpha, which is not a procedure we recommend today.

To improve upon Goodboy and Frisby (2014), we tested the fit of the tau-equivalent model by imposing an equality constraint on the factor loadings but freely estimated the error variances that could be unequal. Global fit statistics unanimously indicated we should reject the tau-equivalent model: (χ^2 (27) = 51.833, p = .003; SRMR = .079; RMSEA = .064 [.037, .091], CFI = .901). Moreover, several normalized residuals were large and significant (i.e. above the absolute value of 2) indicating local fit issues. Because the tau-equivalent model was rejected, coefficient alpha was not an appropriate estimate of reliability for Goodboy and Frisby (2014). Next, we tested the congeneric model by releasing the equality constraint for the factor loadings and freely estimated the loadings. This congeneric measurement model is displayed in Figure 2.

The global fit of the congeneric model provided better data and model correspondence: (χ^2 (20) = 30.025, p = .069; SRMR = .040; RMSEA = .048 [.000, .081], CFI = .960). The χ^2 was not significant, indicating that the measurement model did not significantly deviate from a test of exact fit. The SRMR was below .08, suggesting that the average correlation residual was low. The RMSEA was below .05, giving support for close fit. The CFI indicated that the measurement model was 96% better than a baseline independence model. All four global fit statistics suggested that the measurement model be retained. Likewise, normalized residuals were inspected for local fit evaluation. There were no normalized residuals that revealed significant areas of concern (all residuals’ z values were below 1.96; p

> .05). Table 1 provides a matrix of normalized residuals, that should be regularly reported with CFA for local fit assessment (Kline, 2016).

Given that we retained the congeneric model as it fit the item data well, we can now calculate ω based on the unstandardized factor loadings and residual variances provided by the CFA output below. The unstandardized factor loadings and residual variances for each item are displayed in Table 2. Full Mplus input syntax and an output of the CFA results is provided at:[Files = Annals Unidimensional CFA.inp; Annals Unidimensional CFA Output.out].

First, we calculated the sum of the factor loadings from Table 2 and then squared the sum $(\Sigma\ \lambda_j)^2$:

$$(.905 + .645 + .896 + .928 + .514 + .513 + .769 + .752)^2 = 35.070$$

Second, we calculated the sum of the residual variances $(\Sigma\ \theta_j)$:

$$(1.478 + 3.183 + 2.088 + .919 + 1.004 + 1.583 + 2.214 + 1.290) = 13.759$$

Third, we used these values to calculate ω as:

$$\left(\frac{\text{sum of factor loadings squared}}{\text{sum of factor loadings squared} + \text{sum of residual variances}}\right) = \left(\frac{35.070}{35.070 + 13.579}\right) = \frac{35.07}{48.829}$$
$$= .718 = \text{scale reliability}.$$

Although we demonstrated these calculations using McDonald's (1999) formula, we can more easily perform this calculation in Mplus with the added bonus of requesting a bootstrapped confidence interval (using 10,000 samples) for ω (Raykov & Marcoulides, 2011). Our Mplus syntax for ω with bootstrapped CI is below in Table 3.

For those who would like to replicate this calculation in Mplus, the data file (Annals Omega.dat), input file, and output file are available at: [Files = Annals Omega.inp; Annals Omega Output.out].

Of interest are the values under 'New/Additional Parameters':

$$\text{OMEGA} = .718[95\%\ \text{CI}:.650,\ .767]$$

Goodboy and Frisby's (2014) scale reliability is .718, with a highly probable population value between .650 and .767. This is not impressive for a scale, given that 28% of the variance in the AEQ is measurement error. Of interest might be which items could be deleted to improve scale

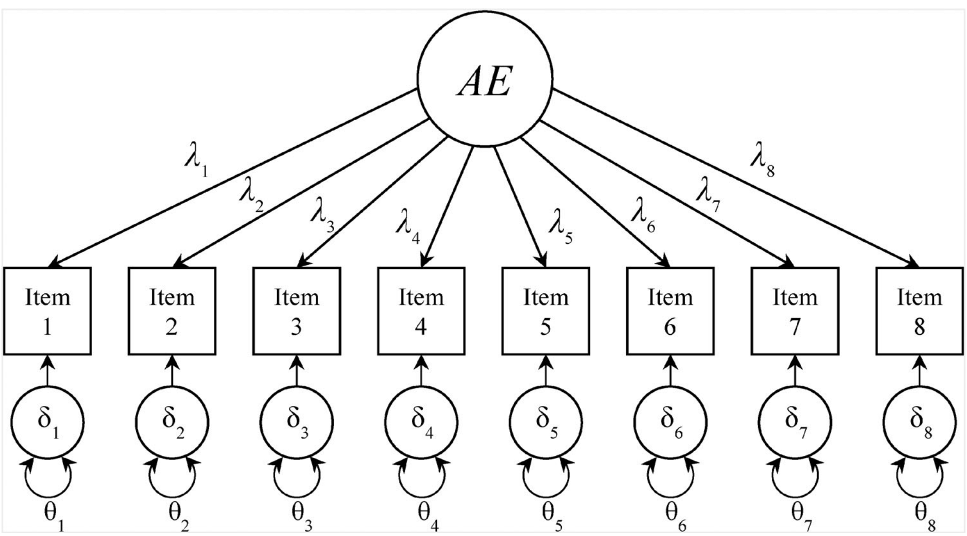

Figure 2. CFA of 8-item Academic Entitlement Questionnaire. Note: λ = unstandardized factor loading. θ = residual variance.

Table 1. Normalized Residuals.

	AE1	AE2	AE3	AE4	AE5	AE6	AE7	AE8
AE1	0.000							
AE2	0.841	0.000						
AE3	−0.526	−0.534	0.000					
AE4	−0.112	0.164	1.341	0.000				
AE5	−0.710	−1.502	−0.561	0.617	0.000			
AE6	0.090	0.802	−0.287	−0.939	0.947	0.000		
AE7	0.403	−0.195	−0.635	−1.170	0.641	1.223	0.000	
AE8	0.427	−0.022	−0.547	−0.344	−0.034	−0.237	1.009	0.000

reliability. Although we recommended that 'alpha if item deleted' procedures be abandoned, we can calculate 'omega if item deleted' instead to see if dropping an item would improve reliability. That is, we can evaluate the relative gain or loss to omega after deleting or adding items to a preliminary scale (Raykov, 2009). Although we are demonstrating this item deletion procedure for educational purposes, we caution readers against deleting items from established measures that have undergone validity testing for two reasons. First, deleting items from established scales likely jeopardizes the test content (i.e. content validity). Second, deleting items based on statistics alone ignores theoretical and substantive considerations that went into creating the items initially (i.e. these items were important to include for a substantive reason). To ensure that researchers are not exploiting sample specific variation (i.e. chance), scales that are shortened from item deletion procedures should be replicated in new data sets. Thus, the tutorial and syntax we provide below is more appropriate for newly developed measures that might undergo revision, not for established measures. Then again, the AEQ produced reliability of .72, which is under a satisfactory value of .80, so it might need improvement.

Omega if item deleted

If we pretended that the AEQ was tentative (e.g. assumed it could use revision) and wanted to improve its reliability by dropping an item, we would consult the unstandardized factor loadings and residual variances from Table 1. At first glance, item 2 appears to be a potential item to remove with the largest residual variance ($\theta_2 = 3.183$). Using the Mplus code below in Table 4, we added syntax to the model constraint statement to calculate ω and bootstrapped CI after the removal of each scale item 1–8. The full Mplus input file with all parameters and output file is available at: [Files = Annals Omega if Item Deleted.inp; Annals Omega if Item Deleted Output.out].

In this output, under 'New/Additional Parameters' Omega is calculated with zero items deleted, (Omega0), item 1 deleted (Omega1), item 2 deleted (Omega2), and so on. The 'Estimate' is ω, which is surrounded by the results of the bootstrapped confidence intervals with 10,000 samples. We recommend inspecting the lower and upper 2.5% which gives the 95% CI. However, from this output, one could derive the 90% CI or 99% CI as well. Part of this output is displayed below in Table 5.

Table 2. Unstandardized factor loadings, standard errors, and residual variances for AE items.

	λ	SE	θ
Item 1: If I don't do well on a test, the professor should make tests easier or curve grades.	.905	.108	1.478
Item 2: Professors should only lecture on material covered in the textbook and assigned readings.	.645	.142	3.183
Item 3: If I am struggling in a class, the professor should approach me and offer help.	.896	.124	2.088
Item 4: It is the professor's responsibility to make it easy for me to succeed.	.928	.094	.919
Item 5: I am a product of my environment. Therefore, if I do poorly in my class, it is not my fault.	.514	.083	1.004
Item 6: If I cannot learn the material for a class from lecture alone, then it is the professor's fault when I fail the test.	.513	.102	1.583
Item 7: I should be given the opportunity to make up a test, regardless of the reason for the absence.	.769	.125	2.214
Item 8: Because I pay tuition, I deserve passing grades.	.752	.098	1.290

Table 3. Mplus Syntax for Calculating Coefficient ω with 95% Bootstrapped CI.

```
TITLE: Omega for Academic Entitlement Scale
DATA: FILE IS Annals Omega.dat;
VARIABLE: NAMES ARE ae1–ae8;
USEVARIABLES = ae1–ae8;
ANALYSIS: estimator = ML;
          bootstrap = 10000;
MODEL: AE BY ae1*(P1)
         ae2–ae8 (P2–P8);
         ae1–ae8 (P9–P16);
         AE@1;
MODEL CONSTRAINT:
       NEW(OMEGA);
       OMEGA = (P1 + P2 + P3 + P4 + P5 + P6 + P7 + P8)**2/
               ((P1 + P2 + P3 + P4 + p5 + P6 + P7 + P8)
               **2 + P9 + P10 + P11 + P12 + P13 + P14 + P15 + P16);
OUTPUT: CINTERVAL(BCBOOTSTRAP);
```

Since we already know that ω is .718 [.650, .767] and reliability is low, an inspection of these new parameters will tell us if we can improve reliability by deleting an item. Only deleting item 2 (Omega2) will increase reliability to .725 [.661, .773]. Deleting any other item will result in a decrease of ω. Again, we are hesitant to recommend deleting items from validated scales, and in this example, the increase in reliability is trivial, so we would not recommend deleting item 2 in a research paper. Yet, we might not recommend using the AEQ at all based on its low reliability estimate. We are only demonstrating this procedure for educational purposes. In practice this procedure is better applied in preliminary stages of scale development to create a final item pool. Deleting scale items might appreciably affect the validity of a scale. Moreover, trivial increases in ω from item deletion might be due to sample specific variation, so replication in new samples is necessary. More on this procedure can be found in Raykov (2009) and Raykov and Marcoulides (2011).

Other software considerations

For those who are not Mplus users and do not want to calculate ω by hand, there are several easy to use options to estimate ω from other statistical software. For R users, there are several packages including Revelle's *psych* package (personality-project.org/r/psych/HowTo/omega.pdf) with a tutorial in R (Revelle & Condon, 2019). Other R tutorials for ω can be found in Dunn et al. (2014), Viladrich et al. (2017), McNeish (2018), and Peters (2014). For those less comfortable with R, there are two free open source software packages with a drag-and-drop interfaces (similar to SPSS and will import SPSS files) including JASP (jasp-stats.org) and jamovi (jamovi.org) that will calculate ω and ω if item deleted (but currently without bootstrapped CIs). An excellent option for SPSS and SAS users is Hayes and Coutts' (2020) OMEGA macro that will calculate ω using ML exploratory factor analysis (without bootstrapping), with the option to calculate ω using Hancock and An's (2020) closed form estimation that allows for a bootstrapped confidence interval. The OMEGA macro also produces ω if item deleted and can generate ω from all possible subsets of items. The OMEGA macro is the best option (and is easy to use) for communication scholars who prefer SPSS and SAS (see Hayes & Coutts, 2020). Indeed, ω is 'catching on' in social sciences and these software options make it easier to calculate and report. Nonetheless, we hope that ω becomes a standard feature in all statistical software packages soon.

Discrepancies between coefficient alpha and omega

How large of a reliability discrepancy will there be between coefficient alpha and omega in communication research? As in our worked example, the answer might be 'not much.' In a meta-analysis

Table 4. Mplus Syntax for Calculating ω if Item Deleted.

```
TITLE: Omega for Academic Entitlement Scale
DATA: FILE IS Annals Omega.dat;
VARIABLE: NAMES ARE ae1–ae8;
USEVARIABLES = ae1–ae8;
ANALYSIS: estimator = ML;
          bootstrap = 10000;
MODEL: AE BY ae1*(P1)
        ae2–ae8 (P2–P8);
        ae1–ae8 (P9–P16);
        AE@1;
MODEL CONSTRAINT:
NEW(OMEGA0 OMEGA1 OMEGA2 OMEGA3 OMEGA4 OMEGA5 OMEGA6
OMEGA7 OMEGA8);
 OMEGA0 = (P1 + P2 + P3 + P4 + P5 + P6 + P7 + P8)**2/
          ((P1 + P2 + P3 + P4 + P5 + P6 + P7 + P8)
          **2 + P9 + P10 + P11 + P12 + P13 + P14 + P15 + P16);
 OMEGA1 = (P2 + P3 + P4 + P5 + P6 + P7 + P8)**2/
          ((P2 + P3 + P4 + P5 + P6 + P7 + P8)
          **2 + P10 + P11 + P12 + P13 + P14 + P15 + P16);
 OMEGA2 = (P1 + P3 + P4 + P5 + P6 + P7 + P8)**2/
          ((P1 + P3 + P4 + P5 + P6 + P7 + P8)
          **2 + P9 + P11 + P12 + P13 + P14 + P15 + P16);
 OMEGA3 = (P1 + P2 + P4 + P5 + P6 + P7 + P8)**2/
          ((P1 + P2 + P4 + P5 + P6 + P7 + P8)
          **2 + P9 + P10 + P12 + P13 + P14 + P15 + P16);
 OMEGA4 = (P1 + P2 + P3 + P5 + P6 + P7 + P8)**2/
          ((P1 + P2 + P3 + P5 + P6 + P7 + P8)
          **2 + P9 + P10 + P11 + P13 + P14 + P15 + P16);
 OMEGA5 = (P1 + P2 + P3 + P4 + P6 + P7 + P8)**2/
          ((P1 + P2 + P3 + P4 + P6 + P7 + P8)
          **2 + P9 + P10 + P11 + P12 + P14 + P15 + P16);
 OMEGA6 = (P1 + P2 + P3 + P4 + P5 + P7 + P8)**2/
          ((P1 + P2 + P3 + P4 + P5 + P7 + P8)
          **2 + P9 + P10 + P11 + P12 + P13 + P15 + P16);
 OMEGA7 = (P1 + P2 + P3 + P4 + P5 + P6 + P8)**2/
          ((P1 + P2 + P3 + P4 + P5 + P6 + P8)
          **2 + P9 + P10 + P11 + P12 + P13 + P14 + P16);
 OMEGA8 = (P1 + P2 + P3 + P4 + P5 + P6 + P7)**2;/
          ((P1 + P2 + P3 + P4 + P5 + P6 + P7)
          **2 + P9 + P10 + P11 + P12 + P13 + P14 + P15);
OUTPUT: CINTERVAL(BCBOOTSTRAP);
```

from 24 journals (across psychology, marketing, management, and education) comparing 2525 pairs of alpha and composite reliability (i.e. omega) estimates, on average, omega reliability was .018 units higher (*SD* = .047) which amounts to 2.1% (Peterson & Kim, 2013). That is, from the same data, on average, alpha produced a .84 value whereas composite reliability (omega) yielded a value of .86.

Table 5. Coefficient ω if Item Deleted.

	New/Additional Parameters						
	LL .5%	LL 2.5%	LL 5%	Estimate	UL 5%	UL 2.5%	UL .5%
OMEGA0	0.624	0.650	0.661	0.718	0.759	0.767	0.779
OMEGA1	0.561	0.588	0.604	0.672	0.724	0.733	0.747
OMEGA2	0.635	0.661	0.671	0.725	0.765	0.773	0.785
OMEGA3	0.565	0.597	0.613	0.684	0.732	0.740	0.756
OMEGA4	0.535	0.573	0.589	0.660	0.710	0.719	0.735
OMEGA5	0.591	0.621	0.633	0.696	0.742	0.750	0.762
OMEGA6	0.608	0.634	0.648	0.706	0.748	0.757	0.769
OMEGA7	0.595	0.623	0.634	0.697	0.743	0.751	0.765
OMEGA8	0.577	0.603	0.618	0.682	0.729	0.737	0.751

Note: For each of the new parameters, we are leaving out 1 scale item and calculating a revised scale with omega.

Although this is not a large difference in reliability, it is certainly meaningful, and the difference was statistically significant ($p < .001$). To be fair, Peterson and Kim (2013) also noted that alpha and omega estimates were the same for 27% of observations, and discrepancies were within +/ − .02 of each other for 61% of observations. On the other hand, studies with small sample sizes (i.e. < 100) produced a larger discrepancy of .06 with an average alpha of .82 and an average omega of .88.

Recently, Hayes and Coutts (2020) examined data from 17 scales and calculated coefficient alpha and omega, finding only trivial discrepancies between the two. These trivial discrepancies (i.e. close to zero) will be the case with validated measures that produce uniformly high factor loadings (Raykov, 2012). Nonetheless, Hayes and Coutts (2020) pointed out that discrepancies can be larger and recommended 'it is high time to make the switch to ω' (p. 21). If a researcher is concerned about the discrepancy between α and ω in a data set, there are methods to determine if this difference is appreciable beyond sampling error (see Deng & Chan, 2017). In practice, greater discrepancies are likely due to the restrictive assumptions of alpha being violated. But even McDonald (1999) admitted, 'it is, in fact, difficult to invent a homogenous population structure in which alpha is a very poor lower bound to omega, or to find empirical examples in which the estimate of alpha is very much lower than that of omega' (p. 93). Regardless, although the discrepancies between alpha and omega might not be much on average, their difference can be greater than .02, and communication scholars should always report the more accurate estimate (omega).

Extensions of omega

We have discussed omega in the context of a congeneric scale where all approximately continuous scale items (i.e. 7-point Likert response format) are explained by a single factor (i.e. homogenous items). Briefly, we want to mention there are other applications of omega in different measurement models. There is categorical omega (see Kelly & Pornprasertmanit, 2016) when items have ordinal responses with few categories (e.g. 4-point response format). There are also extensions of omega for scales with multidimensional factor structures (i.e. heterogenous items), including a bifactor model (in CFA or ESEM; see Murray et al., 2019) in which omega is partitioned into omega hierarchical (ω_H) for a general factor and omega subscale (ω_{HS}) for specific factors (see Gignac, 2014). For nested data requiring multilevel CFA, there is within-level and between-cluster ω (see Geldhof et al., 2014). Thus, coefficient ω is not limited to unidimensional measures with continuous item response formats.

Summary

Communication scholars must make the switch from coefficient α to ω when reporting reliability of multi-item unidimensional measures. This should not be a statistical choice to make or a preference for researchers; coefficient ω will either be equal to, or most likely, a better estimate of reliability for communication measures (always). Other social sciences have already begun to make this switch. Schrodt (2020) recently editorialized that 'good research uses the right method, great research uses the right method *well* … using the right methods well also involves understanding the limitations of the method and explicating how those limitations contextualize the knowledge claims of the study (p. 2). We agree with Schrodt. To do great research, communication scholars need to report a reliability estimate with a confidence interval that aligns with the definition of reliability itself. This is coefficient ω, which does not have the same restrictive assumptions as α, so it is more flexible for scales that do not have equal factor loadings or uniformly high loadings. Relying on coefficient ω ensures that we are reporting scale reliability more accurately, while also requiring us to confirm our unidimensional measurement models, which speaks to construct validity. Coefficient ω can now be calculated with a scholar's choice of preferred software packages. We hope to see communication researchers, journal editors, manuscript reviewers, methods instructors, and graduate students, routinely report ω instead of α for scale reliability.

Notes

1. Although the article written by Lee Cronbach (1951) has been attributed for proposing the formula for coefficient α, the formula was proposed earlier by Guttman (1945). For a more detailed history and timeline of prior reliability formulas leading to α, see Cho and Kim (2015). Cronbach never named α after himself (i.e., Cronbach's alpha), although like many other social science disciplines, communication scholars have dubbed it so. Cronbach disliked that coefficient α was attributed to him and the label α altogether. Cronbach (2004) disclosed 'it is an embarrassment to me that the formula became conventionally known as Cronbach's α. The label alpha, which I applied, is also an embarrassment. It bespeaks my conviction that one could set up a variety of calculations that would assess properties of test score other than reliability, and alpha was only the beginning' (p. 397).
2. Even Cronbach himself no longer recommended alpha. After reflecting on his 1951 publication, he acknowledged 'my own views had evolved; I doubt whether coefficient alpha is the best way of judging the reliability of the instrument to which it is applied' (Cronbach, 2004, p. 393). Likewise, Cronbach admitted 'I no longer regard the alpha formula as the most appropriate way to examine most data' (p. 403).
3. This manuscript discusses composite reliability. There is also an entire literature on maximal reliability (construct replicability), calculated with coefficient *H* (Hancock & Mueller, 2001), as 'the squared correlation between a latent construct and the optimum linear composite formed by its indicators' (p. 195). *H* represents the highest reliability that could be obtained in a particular sample from an optimally weighted composite of items (i.e., maximal reliability is never lower than composite reliability). To estimate differences between composite reliability and maximal reliability coefficients, see Raykov et al. (2016).

Acknowledgement

The authors would like to thank Gregory R. Hancock for his advice regarding this manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, *74*(1), 137–143. https://doi.org/10.1007/s11336-008-9100-1

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.

Cho, E., & Kim, S. (2015). Cronbach's coefficient alpha: Well known but poorly understood. *Organizational Research Methods*, *18*(2), 207–230. https://doi.org/10.1177/1094428114555994

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(*1*), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

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

Cronbach, L. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, *64*(3), 391–418. https://doi.org/10.1177/0013164404266386

Crutzen, R., & Peters, G.-J. Y. (2017). Scale quality: Alpha is an inadequate estimate and factor-analytic evidence is needed first of all. *Health Psychology Review*, *11*(3), 242–247. https://doi.org/10.1080/17437199.2015.1124240

Deng, L., & Chan, W. (2017). Testing the difference between reliability coefficients alpha and omega. *Educational and Psychological Measurement*, *77*(2), 185–203. https://doi.org/10.1177/0013164416658325

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*(1), 72–91. https://doi.org/10.1037/a0032138

Gignac, G. E. (2014). On the inappropriateness of using items to calculate total scale score reliability via coefficient alpha for multidimensional scales. *European Journal of Psychological Assessment*, *30*(2), 130–139. https://doi.org/10.1027/1015-5759/a000181

Goodboy, A. K., & Frisby, B. N. (2014). Instructional dissent as an expression of students' academic orientations and beliefs about education. *Communication Studies*, *65*(1), 96–111. https://doi.org/10.1080/10510974.2013.785013

Goodboy, A. K., & Kline, R. B. (2017). Statistical and practical concerns with published communication research featuring structural equation modeling. *Communication Research Reports*, *34*(1), 68–77. https://doi.org/10.1080/08824096.2016.1214121

Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and Psychological Measurement*, *66*(6), 930–944. https://doi.org/10.1177/0013164406288165

Green, S. B., & Hershberger, S. L. (2000). Correlated errors in true score models and their effect on coefficient alpha. *Structural Equation Modeling*, *7*(2), 251–270. https://doi.org/10.1027/S15328007SEM0702_6

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, *37*(4), 827–838. https://doi.org/10.1177/001316447703700403

Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, *74*(1), 121–135. https://doi.org/10.1007/s11336-008-9098-4

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255–282. https://doi.org/10.1007/BF02288892

Hancock, G. R., & An, J. (2018). Scale reliability in structural equation modeling. *Educational Measurement: Issues and Practice*, *37*(2), 73–74. https://doi.org/10.1111/emip.12210

Hancock, G. R., & An, J. (2020). A closed-form alternative for estimating ω reliability under unidimensionality. *Measurement: Interdisciplinary Research and Perspectives*, *18*(1), 1–14. https://doi.org/10.1080/15366367.2019.1656049

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct validity with latent variable systems. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future – A festshrift in honor of Karl Jöreskog* (pp. 195–216). Social Scientific International.

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But … *Communication Methods and Measures*, *14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Hoekstra, R., Vugteveen, J., Warrens, M. J., & Kruyen, P. M. (2019). An empirical analysis of alleged misunderstandings of coefficient alpha. *International Journal of Social Research Methodology*, *22*(4), 351–364. https://doi.org/10.1080/13645579.2018.1547523

Johnson, A. J. (2017). Reliability, Cronbach's alpha. In M. Allen (Ed.), *The SAGE encyclopedia of communication research methods* (pp. 1415–1417). Sage.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, *36*(2), 109–133. https://doi.org/10.1007/BF02291393

Kelly, K., & Pornprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological Methods*, *21*(1), 69–92. https://doi.org/10.1037/a0040086

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Komperda, R., Pentecost, T. C., & Barbera, J. (2018). Moving beyond alpha: A primer on alternative sources of single-administration reliability evidence for quantitative chemistry education research. *Journal of Chemical Education*, *95*(9), 1477–1491. https://doi.org/10.1021/acs.jchemed.8b00220

Kopp, J. P., Zinn, T. E., Finney, S. J., & Jurich, D. P. (2011). The development and evaluation of the academic entitlement questionnaire. *Measurement and Evaluation in Counseling and Development*, *44*(2), 105–129. https://doi.org/10.1177/0748175611400292

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *The British Journal of Mathematical and Statistical Psychology*, *23*(1), 1–21. https://doi.org/10.1111/j.2044-8317.1970.tb00432.x

McDonald, R. P. (1985). *Factor analysis and related methods*. Erlbaum.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Erlbaum.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

Murray, A. L., Booth, T., Eisner, M., Obsuth, I., & Ribeaud, D. (2019). Quantifying the strength of general factors in psychopathology: A comparison of CFA with maximum likelihood estimation, BSEM, and ESEM/EFA bifactor approaches. *Journal of Personality Assessment*, *101*(6), 631–643. https://doi.org/10.1080/00223891.2018.1468338

Muthén, B. O., & Muthén, B. O. (2017). *Mplus user's guide (Version 8)*s.

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, *3*(1), 1–18. https://doi.org/10.1016/0022-2496(66)90002-2

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

Padilla, M. A., & Divers, J. (2016). A comparison of composite reliability estimators: Coefficient omega confidence intervals in the current literature. *Educational and Psychological Measurement*, *76*(3), 436–453. https://doi.org/10.1177/0013164415593776

Peters, G.-J. Y. (2014). The alpha and omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist*, *16*, 56–69. https://doi.org/10.31234/osf.io/h47fv

Peterson, R. A. (1994). A meta-analysis of Cronbach's alpha. *Journal of Consumer Research*, *21*(2), 381–391. https://doi.org/10.1086/209405

Peterson, R., & Kim, Y. (2013). On the relationship between coefficient alpha and composite reliability. *Journal of Applied Psychology*, *98*(1), 194–198. https://doi.org/10.1037/a0030767

Raykov, T. (1997a). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research*, *32*(4), 329–353. https://doi.org/10.1207/s15327906mbr3204_2

Raykov, T. (1997b). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, *21*(2), 173–184. https://doi.org/10.1177/01466216970212006

Raykov, T. (2001). Bias of coefficient *α* for fixed congeneric measures with correlated errors. *Applied Psychological Measurement*, *25*(1), 69–76. https://doi.org/10.1177/01466216010251005

Raykov, T. (2007). Reliability if deleted, not 'alpha if deleted': Evaluation of scale reliability following component deletion. *British Journal of Mathematical and Statistical Psychology*, *60*(2), 201–216. https://doi.org/10.1348/000711006X115954

Raykov, T. (2008). Alpha if item deleted: A note on loss of criterion validity in scale development if maximizing coefficient alpha. *British Journal of Mathematical and Statistical Psychology*, *61*(2), 275–285. https://doi.org/10.1348/000711007X188520

Raykov, T. (2009). Interval estimation of revision effect on scale reliability via covariance structure modeling. *Structural Equation Modeling*, *16*(3), 539–555. https://doi.org/10.1080/10705510903008337

Raykov, T. (2012). Scale construction and development using structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 472–492). Guilford Press.

Raykov, T. (2019). Strong convergence of the coefficient alpha estimator for reliability of multiple-component measuring instruments. *Structural Equation Modeling*, *26*(3), 430–436. https://doi.org/10.1080/10705511.2018.1515019

Raykov, T., Gabler, S., & Dimitrov, D. M. (2016). Maximal reliability and composite reliability: Examining their difference for multicomponent measuring instruments using latent variable modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(3), 384–391. https://doi.org/10.1080/10705511.2014.966369

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Routledge.

Raykov, T., & Marcoulides, G. A. (2015). A direct latent variable modeling based method for point and interval estimation of coefficient alpha. *Educational and Psychological Measurement*, *75*(1), 146–156. https://doi.org/10.1177/0013164414526039

Raykov, T., & Marcoulides, G. A. (2016). Scale reliability evaluation under multiple assumption violations. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(2), 302–313. https://doi.org/10.1080/10705511.2014.938597

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

Revelle, W., & Condon, D. M. (2019). Reliability from α to ω: A tutorial. *Psychological Assessment*, *31*(12), 1395–1411. https://doi.org/10.1037/pas0000754

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. https://doi.org/10.1037/met0000045

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye, & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Sage.

Savalei, V., & Reise, S. P. (2019). Don't forget the model in your model-based reliability coefficients: A reply to McNeish (2018). *Collabra: Psychology*, *5*(1), 36. https://doi.org/10.1525/Collabra.247

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353. https://doi.org/10.1037/1040-3590.8.4.350

Schrodt, P. (2020). What is the bar? Differentiating good from great communication scholarship. *Communication Monographs*, *87*(1), 1–3. https://doi.org/10.1080/03637751.2020.1709696

Shevlin, M., Miles, J. N. V., Davies, M. N. O., & Walker, S. (2000). Coefficient alpha: A useful indicator of reliability? *Personality and Individual Differences*, *28*(2), 229–237. https://doi.org/10.1016/S0191-8869(99)00093-8

Sijtsma, K. (2009). On the use, misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0

Streiner, D. L. (2003). Being inconsistent about consistency: When coefficient alpha does and does not matter. *Journal of Personality Assessment*, *80*(3), 217–222. https://doi.org/10.1207/S15327752JPA8003_01

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology*, *4*, 769. https://doi.org/10.3389/fpsyg.2013.00769

Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). A journey around alpha and omega to estimate internal consistency reliability. *Anales de Psicologia*, *33*, 755-782. https://doi.org/10.6018/analesps.33.3.268401

Yang, Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the twenty-first century? *Journal of Psychoeducational Assessment*, *29*(4), 377–392. https://doi.org/10.1177/0734282911406668

Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, *30*, 165–200. https://doi.org/10.1111/0081-1750.00078

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

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ωH: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*(1), 123–133. https://doi.org/10.1007/s11336-003-0974-7