RESEARCH ARTICLE

A Primer on Reliability via Coefficient Alpha and Omega

Miguel A. Padilla

Author's affiliation:

Department of Psychology, Old Dominion University, 250 Mills Godwin Building, Norfolk, VA 23529.

***Corresponding author:** Miguel A. Padilla
Department of Psychology, Old Dominion University, 250 Mills Godwin Building, Norfolk, VA 23529.
Email: mapadill@odu.edu

ABSTRACT

There are several types of reliabilities when measuring constructs, each with its own purpose. However, the focus here is on reliability from the perspective of coefficients alpha and omega. Whether warranted or not, coefficient alpha has been the standard for establishing test reliability for over 60 years. Although coefficient omega was proposed as an alternative to coefficient alpha, it continues to be relatively unknown. To an extent, the popularity of coefficient alpha and obscurity of coefficient omega can be attributed to a misunderstanding of the models and assumptions associated with each. Additionally, until recently, research on developing each coefficient has been relatively stagnant, providing a void where the popularity and obscurity occurred. The aim of this article is twofold. First, a clear discussion is presented for the models and assumptions associated with both coefficients, and how they relate to test reliability. Second, a review is provided of some of the latest research in developing both coefficients. Taken together, the information will enable researchers to make informed decisions about applying coefficients alpha and omega in their own research.

Keywords: coefficient alpha and omega, reliability, composite reliability, bootstrap, confidence interval, interval estimate.

1. Introduction: Developing measurement instruments for psychological constructs (hereinafter referred to as constructs) requires that those instruments be evaluated. In the social/behavioral sciences, measurement instruments commonly take several forms including tests, questionnaires, surveys, or inventories consisting of items or components. The evaluation of measurement instruments requires the consideration of two important psychometric properties: reliability and validity. One thing to point out is that reliability and validity are properties of the set of items from a measurement instrument used to measure the construct, and not properties of the construct itself. Although each psychometric property has several variations and approaches, this article will focus on reliability from the point of view of coefficients alpha and omega. To appreciate coefficients alpha and omega, some ideas from validity will also be presented. Even so, the aim of this article is to provide a brief overview of the central ideas behind coefficients alpha and omega, along with a review of some of their current developments, to enable researchers to make informed decisions about the application of each coefficient.

From a general perspective, reliability is the idea of consistency (or stability) in responding to a set of items that measure a construct. There are several types of reliability, each with its own purpose, and each has a standard requirement for how many administrations are needed from the measurement instrument. For example, test-retest reliability is appropriate for assessing the stability of a construct over time. Test-retest reliability requires two test administrations of the same measurement instrument over a time interval. The length of the time interval is specific to the construct being measured. It is also possible to obtain two or more types of reliability information. Test-retest with alternate forms reliability is appropriate when alternate forms of a measurement instrument are required for assessing the stability of a construct over time. Test-retest with alternate forms reliability also requires two test administrations over a time interval, but uses one form at the first time point and a similar (alternate) form at the second time point. Therefore, the particular type of reliability depends on the particular type of stability that a researcher wants to measure from a construct. Before considering coefficients alpha and omega, two important concepts are first discussed: 1) dimensionality and internal consistency and 2) classical test theory and its connection to reliability.

2. Dimensionality and Internal Consistency: Dimensionality refers to the number of constructs that represent a set of items and is an important concept to consider when measuring constructs. For example, unidimensional refers to one construct, two-dimensional refers to two constructs, etc. Unfortunately, unidimensionality is also referred to as item homogeneity, and item homogeneity is often confused with internal consistency. Internal consistency refers to the interrelatedness of a set of items (Furr, 2018). The result is that unidimensionality is often confused with internal consistency (Furr, 2018; Lord, Novick, & Birnbaum, 1968). Unidimensionality and internal consistency are related but do not refer to the same idea. Unidimensionality implies internal consistency. However, internal consistency does not imply unidimensionality. In fact, it is possible for a set of items to be interrelated and multidimensional (not unidimensional or homogenous).

To demonstrate the relationship between dimensionality and internal consistency, consider a set of nine items that are represented by spheres (circles) in the following scenarios (Yu, 2001). In each scenario, a correlation between items is represented by an intersection between the corresponding spheres. In Figure 1, none of the items intersect. In this scenario, the items are neither internally consistent nor unidimensional. In Figure 2, all the items intersect. In this scenario, the items are internally consistent and unidimensional (homogenous). In Figure 3, some of the items intersect, but not all of them. Here, the items are internally consistent, but they are also multidimensional (not unidimensional or homogenous).

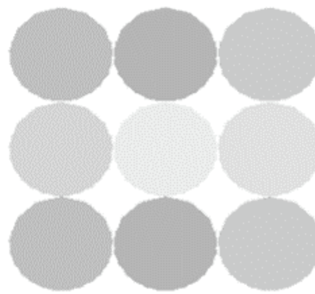

Figure 1. No internal consistency and not unidimensional.

Figure 2. Internally consistent and unidimensional.

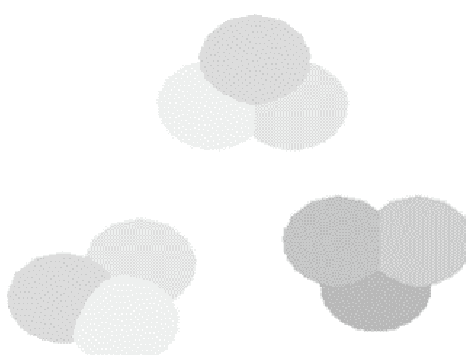

Figure 3. Internally consistent and multidimensional.

Determining dimensionality is a validity question established through validity techniques such as factor analysis. Broadly speaking, validity is concerned with the ability of a set of items to measure what they are supposed to measure for the context of interest (Tarkkonen & Vehkalahti, 2005). It is key to note that determining dimensionality must be established separately from internal consistency. This article does not cover validity techniques, but the interested reader can consult Furr (2018) and Crocker and Algina (1986). Next, reliability from the perspective of classical test theory based models is presented.

3. Classical Test Theory Based Models and Reliability: Classical test theory (CTT) is at the center of reliability and provides the foundation for three related models (Graham, 2006). The models will be presented from the most to least restrictive, which in this context refers to the assumptions of each model. For example, more restrictive indicates more or stronger assumptions and less restrictive indicates less or weaker assumptions. In addition, all models assume the items to be unidimensional, and index individuals with $i = 1, 2, ..., n$ and items with $j = 1, 2, ..., k$.

The first model is the most restrictive because it assumes the items to be parallel. As such, this model is referred to as the parallel model and is written as

$$x_{ij} = \tau_i + u_i, \tag{1}$$

where x_{ij} is the observed score for individual i on item j, and τ_i is the true score and u_i is the random measurement error for individual i. Here, the model specifies that each individual has their own true score τ_i and measurement error u_i, but that the true score for each individual is the same (or constant) across all items, which is implied by τ not having a j subscript. In addition, the measurement errors are assumed to be independent and have the same distribution with a mean of zero and constant variance $\left(\sigma_u^2\right)$. It then follows that the items (x_js) have equal variances $\left(\sigma_j^2 = \sigma^2\right)$, equal covariances $\left(\sigma_{ij} = \theta\right)$, and that the variances are not equal to the covariances $\left(\sigma^2 \neq \theta\right)$; i.e., the item covariance matrix is compound symmetric. In short, the parallel model assumes that all the items are equivalent in measuring the construct of interest.

The second model assumes the items to be tau-equivalent or essentially tau-equivalent. The tau-equivalence model is written as

$$x_{ij} = \tau_i + u_{ij}, \tag{2}$$

and the essentially tau-equivalence model is written as

$$x_{ij} = \left(a_j + \tau_i\right) + u_{ij}, \tag{3}$$

where a_j is a unique constant for item j. Here, the models specify that each individual has their own true score τ_i, but that the random measurement error u_{ij} for each individual is potentially different for each item. However, the essentially tau-equivalence model allows the true scores to differ by a unique constant a_j for each item. The measurement errors are assumed to be independent and have the same distribution with a mean of zero but a different variance for each item $\left(\sigma_{u_j}^2\right)$. In this case, the items (x_js) need not have equal variances $\left(\sigma_j^2\right)$ but have equal covariances $\left(\sigma_{ij} = \theta\right)$. Thus, the models in Equations 2 and 3 do not assume that all the items are equivalent in measuring the construct of interest. Specifically, these models assume that the items measure the construct of interest with differing amounts of measurement error. Hereinafter, (essentially) tau-equivalence will refer to the models in Equations 2 and 3.

The third model assumes the items to be congeneric. The congeneric model is written as

$$x_{ij} = \left(a_j + b_j \tau_i\right) + u_{ij}, \qquad (4)$$

where a_j is a unique constant and b_j a unique linear relationship for item j. Here, the model specifies that each individual has their own true score τ_i, and that the random measurement error u_{ij} for each individual is potentially different for each item. However, the model allows the true scores for each item to differ by a unique constant a_j and have a unique linear relationship b_j. The measurement errors are assumed to be independent and have the same distribution with a mean of zero but a different variance for each item $\left(\sigma^2_{u_j}\right)$. In this case, the items (x_js) need not have equal variances $\left(\sigma^2_j\right)$ or equal covariances $\left(\sigma_{ij}\right)$; i.e., the item covariance matrix is unstructured. Like the (essentially) tau-equivalent models, this model does not assume that all the items are equivalent in measuring the construct of interest. Specifically, this model assumes the items measure the construct of interest with differing amounts of measurement error and a differing linear relationship to the construct.

The classical definition of reliability was developed for a pair of parallel tests x_1 and x_2. Under the assumptions of the parallel model (i.e., Equation 1), the correlation between the two parallel tests takes on the following definition

$$\rho_{x_1 x_2} = \frac{\sigma^2_\tau}{\sigma^2_x} = \frac{\sigma^2_\tau}{\sigma^2_\tau + \sigma^2_u} = \rho_{xx'} \qquad (5)$$

where σ^2_x is the observed score variance, σ^2_τ is the true score variance, and σ^2_u is the measurement error variance (Allen & Yen, 1979; Crocker & Algina, 1986). According to Equation 5, reliability indicates the proportion of observed score variance attributable to true score variance. With dimensionality, the three CTT based models, and classical reliability in place, coefficients alpha and omega can now be discussed.

4. Test Reliability via Coefficient Alpha and Omega: Although the classical definition of reliability is foundational, of far more interest is the reliability for a set of items; i.e., going beyond the reliability for a pair of items (or tests). Here, reliability is obtained for a composite/sum score of all the items

$$x = \sum_{j=1}^{k} x_j \,. \qquad (6)$$

Depending on the context, this type of reliability is often referred to as composite reliability, test score reliability, test reliability, or the reliability of a measurement instrument. Note that equation 6 is not the definition of composite reliability. Composite reliability can take on several forms, but this article focuses on coefficients alpha and omega. Of notice is that CTT makes no distinction between tests, questionnaires, surveys, inventories, items, or components. For example, obtaining composite reliability for a set of items is just as plausible as obtaining composite reliability for a set of tests (Allen & Yen, 1979; Crocker & Algina, 1986).

4.1 Coefficient Alpha: When interest is in composite reliability in applied work, it is difficult to assume that all items are equivalent in measuring the construct of interest; i.e., difficult to assume the parallel model. It is more viable to assume that items measure the construct of interest with differing amounts of measurement error; i.e., assume the (essentially) tau-equivalence model. The appropriate form of composite reliability in this situation is coefficient alpha.

The population parameter for coefficient alpha is defined as

$$\alpha_C = \frac{k}{k-1}\left[1 - \frac{\sum_{j=1}^{k}\sigma_{jj}}{\sum_{i=1}^{k}\sum_{j=1}^{k}\sigma_{ij}}\right] \tag{7}$$

where k is the number of items, σ_{jj} are the variances, and σ_{ij} are the covariances from the $k \times k$ item covariance matrix. Coefficient alpha is computed (estimated) from data by using the corresponding elements $\left(\hat{\sigma}_{jj} \text{ and } \hat{\sigma}_{ij}\right)$ from the estimated $k \times k$ item covariance matrix.

Coefficient alpha is one of the most popular methods for assessing the composite reliability for a set of items (Hogan, Benjamin, & Brezinski, 2000). A simple internet search will reveal its use in disciplines such as sociology, anthropology, medicine, political science, economics, etc. (Blinkhorn, 1997). Coefficient alpha's popularity is remarkable given that it has remained unchanged since its first appearance in the literature (Cronbach, 1951; Guttman, 1945). There are three key contributors to its popularity. First, coefficient alpha is computationally easy as it only requires the item covariance (correlation) matrix (see Equation 7). Second, coefficient alpha is appropriate for continuous, ordinal (Likert-type), and binary items. This is a result of coefficient alpha being a more general form of the Kuder-Richardson formula 20 (KR 20), which is only appropriate for binary items (Kuder & Richardson, 1937). Third, coefficient alpha only requires one test administration. Estimating most other forms of reliability requires at least two test administrations. Two test administrations introduce methodological issues such as requiring more time, resources, and the increased potential for missing data due to attrition. Aside from these key features, coefficient alpha has other misunderstood features.

Coefficient alpha is an index of internal consistency (composite reliability) that assumes items to be at least (essentially) tau-equivalent. This directly impacts the properties (or interpretations) of coefficient alpha in two major ways. First, recall that all three CTT based models assume the items to be unidimensional, and this assumption propagates to coefficient alpha. As such, coefficient alpha is not a measure of unidimensionality, nor does it confirm unidimensionality (see Figure 3). Another way to think of this is that coefficient alpha provides no information about dimensionality. In fact, it is lower than composite reliability $\left(\text{i.e., } \alpha_C < \rho_{xx'}\right)$ if computed when the items are multidimensional (Cronbach, 1951).

Second, coefficient alpha is generally considered to be a lower bound to composite reliability. Part of this stems from coefficient alpha being appropriate for parallel and (essentially) tau-equivalent items. In this situation, coefficient alpha is equal to composite reliability $\left(\text{i.e., } \alpha_C = \rho_{xx'}\right)$. When items are parallel, coefficient alpha is then equal to the mean of all split-half reliabilities. However, when items are congeneric, coefficient alpha is lower than composite reliability $\left(\text{i.e., } \alpha_C < \rho_{xx'}\right)$. Considering dimensionality and the three CTT based models together leads to coefficient alpha being the lower bound to composite reliability $\left(\text{i.e., } \alpha_C \le \rho_{xx'}\right)$. On the other hand, when the assumption of independent measurement errors is violated (i.e., the measurement errors are correlated), coefficient alpha can be lower or higher than composite reliability (Raykov & Marcoulides, 2011; Zimmerman, Zumbo, & Lalonde, 1993). Lastly, coefficient alpha provides no information with respect to other forms of reliability, such as test-retest reliability (or stability over time).

4.2 Coefficient Omega: Although (essentially) tau-equivalence is a more realistic assumption than the parallel assumption, it may still not be enough in applied work. A more realistic assumption is that of congeneric items. The appropriate form of internal consistency (composite reliability) for congeneric items is coefficient omega

The population parameter for coefficient omega is defined as

$$\omega = \frac{\left(\sum_{j=1}^{k} \lambda_j\right)^2}{\left(\sum_{j=1}^{k} \lambda_j\right)^2 + \sum_{j=1}^{k} \psi_j} \tag{8}$$

where λ_j and ψ_j are the j^{th} factor loading and uniqueness, respectively, from a factor analytic model (McDonald, 1970, 1999). Coefficient omega is computed (estimated) from data by using the corresponding coefficients $\left(\hat{\lambda}_j \text{ and } \hat{\psi}_j\right)$ from the estimated confirmatory factor analysis (CFA) model. The similarities between Equations 5 and 8 make coefficient omega more intuitive than coefficient alpha.

Coefficient omega was first proposed in the literature 19 years after coefficient alpha (McDonald, 1970). Like coefficient alpha, coefficient omega is appropriate for continuous, ordinal (Likert-type), and binary items, and only requires one test administration. Unfortunately, coefficient omega does not share the popularity of coefficient alpha for three reasons. First, it is dwarfed by the popularity of coefficient alpha that comes from over 60 years of being cited in the literature of numerous disciplines. Second, it is not computationally easy as it requires modeling a CFA through specialized structural equation modeling (SEM) software. Recall that coefficient alpha only requires the item covariance (correlation) matrix. As such, coefficient alpha is common in standard statistical software (e.g., SPSS, SAS, R etc.), but coefficient omega is not. Third, knowledge of its theoretical properties is not as comprehensive. For example, unlike coefficient alpha, the asymptotic distribution of coefficient omega remains unknown. However, computational methods along with current computing power are making coefficient omega research viable. This research will be discussed later.

Coefficient omega is appropriate for parallel, (essentially) tau-equivalent, and congeneric items. It has the same properties (or interpretations) for parallel and (essentially) tau-equivalent items as coefficient alpha. However, for congeneric items, coefficient omega is the appropriate composite reliability index. It is therefore a more general form of composite reliability as it "captures the notion of reliability of a test score" (McDonald, 1999, p. 90). Like coefficient alpha, it provides no information with respect to other forms of reliability (e.g., test-retest reliability).

In addition to being appropriate for items that are at least congeneric, coefficient omega has the additional advantage of modeling flexibility. Recall that coefficient omega uses the factor loadings (λ_i) and uniqueness (ψ_j) from a CFA. This means that coefficient omega is more general in that it is not bound by the unidimensionality assumption. As such, coefficient omega can be obtained for items that are multidimensional. However, in the same way that coefficient alpha is not a measure of unidimensionality, coefficient omega is not a measure of multidimensionality. It is a reliability index that accounts for items that may be multidimensional. Even so, the focus here is on unidimensionality. See McDonald (1999), Raykov and Shrout (2002), and Zinbarg, Revelle, Yovel, and Li (2005) for a discussion of coefficient omega with multidimensional items.

5. Confidence Intervals for Coefficient Alpha and Omega: Little research about further developing coefficients alpha and omega has been published since their inception. During that time, coefficient alpha became a staple for estimating the composite reliability of a measurement instrument while coefficient omega was mostly ignored. However, the research articles by Graham (2006) and van Zyl, Neudecker, and Nel (2000), and computational methods (e.g., the bootstrap; Efron & Tibshirani, 1998) helped jump start new research developments for both coefficients. Graham provided a clear presentation of the three CTT based models, their impact on reliability, and showed how to test if the items conform to these models. Raykov and Marcoulides (2011) also present methods for testing if items conform to the three CTT based models. In addition, in recent years, the advent of computational methods along with the work by van Zyl et al. provided the fuel for active confidence interval research concerning coefficients alpha and omega.

A point and interval estimate are two general methods of using data to estimate or compute a parameter (or population parameter). A point estimate is a single value that estimates a parameter. Point estimate examples are the mean, standard deviation, and coefficients alpha and omega $\left(\text{i.e., } \hat{\mu}, \hat{\sigma}, \hat{\alpha}_C, \& \ \hat{\omega}\right)$. Based on data, a point estimate is a best guess of a parameter, but it does not account for sampling error (or variability) about the parameter estimate. However, an interval estimate accounts for sampling error by providing a range of values that are likely to contain the parameter.

A confidence interval (CI) is a common form of an interval estimate. A CI is formed by creating an error structure around a point estimate based on the standard error (*SE*) of the corresponding sampling distribution with a confidence level (Hogg, Tanis, & Zimmerman, 2015). The consistency of a parameter estimate is indicated by the confidence level denoted as $100(1-\alpha)\%$ where α represents the probability of a type I error. For example, a CI with $\alpha = .05$ indicates that 95% of all CIs created in the same way will contain the corresponding parameter.

CIs provide at least two important pieces of information. First, they provide precision information about the parameter estimate. Specifically, narrower intervals indicate more precision and wider intervals indicate less precision. Second, they can be used for hypothesis testing (inference); e.g., $H_0 : \mu - \mu_0 = 0$. What follows is a review of some of the CI research for coefficients alpha and omega to provide context under which conditions or situations the CIs perform well (or are robust).

5.1 CI Research for Coefficient Alpha: The first coefficient alpha CI was proposed by Feldt (1965). This CI assumed the items to be parallel and normally distributed. The authors demonstrated its use in an example with a coefficient alpha of .70, 26 items, a sample size of 41, and $\alpha = .10$. However, the CI was not accurate if the parallel and normality assumptions were violated (Barchard & Hakstian, 1997). Unfortunately, because these assumptions may not hold in practice, this CI has not seen much use in applied settings (Duhachek & Iacobucci, 2004).

Van Zyl et al. (2000) showed that the original estimation of coefficient alpha from Equation 7 is a maximum likelihood (ML) estimate and that its asymptotic distribution is normal. A useful feature of the asymptotic distribution is that it does not require the items to be parallel. However, it does require the items to be normally distributed. Duhacheck and Iacobucci (2004) subsequently developed a normal-theory (NT) CI for coefficient alpha based on the work from van Zyl et al. Duhacheck and Iacobucci compared the NT CI with those from Feldt (1965) and others with normal unidimensional (parallel) and two-dimensional (heterogeneous) items in sets of 5 and 7 items with sample sizes of 30, 50, 100, and 200, and mean item correlations of .40 to .70. Their results showed that the NT CI consistently outperformed all other CIs they investigated across all simulation conditions.

Given that the previous CI studies required the items to be normally distributed, Yuan, Gaurnaccia, and Hayslip (2003) developed an asymptotically distribution free (ADF) CI for coefficient alpha based on the work from Yuan and Bentler (2002). In addition to the ADF CI, Yuan et al. also proposed a bootstrap CI for coefficient alpha. Yuan et al. applied the ADF, NT, percentile bootstrap (PB), and biased-corrected and accelerated (BCa) bootstrap CIs for coefficient alpha to each dimension of the Hopkins Symptom Checklist (HSCL; Derogatis, Lipman, Rickels, Uhlenhuth, & Covi, 1974). The HSCL consists of five dimensions with 58 4-point Likert-type items. The sample size for the study consisted of 419 individuals, and all bootstrap CIs were based on 1,000 bootstrap samples (or replications) with $\alpha = .10$. According to their results, the BCa and PB

CIs were the most accurate followed by the ADF and NT CIs. Even so, the differences between the bootstrap and ADF CIs were all within three decimals (Maydeu-Olivares, Coffman, & Hartmann, 2007).

Maydeu-Olivares et al. (2007) enhanced the coefficient alpha ADF CI (Yuan et al., 2003) and investigated it under several simulation conditions. Of interest was the performance of the ADF and NT CIs with non-normal items (i.e., different forms of skewness and kurtosis). In one simulation study, they investigated parallel items in sets of 5 and 20 with sample sizes of 50, 100, 200, and 400. In a second simulation study, they investigated congeneric items in sets of 7, 14, and 21 with sample sizes of 50, 100, 400, and 1,000. In both simulations, the items were non-normal and Likert-type with 2 and 5 categories.[1] Maydeu-Olivares et al. found the following similar results for both parallel and congeneric items. When the items were approximately normally distributed, the NT CIs had good coverage with a sample size as small as 50. However, when the items deviate from normality, the ADF CIs outperformed the NT CIs with as small a sample size as 100. Overall, when the sample size was more than 100, the ADF CIs performed well under non-normality.

Romano, Kromrey, and Hibbard (2010) compared the performance of the NT and ADF CIs among others. The study used a 3-parameter item response theory model to generate binary items with coefficient alphas of 0.50, 0.70, and 0.90. In addition, the sample sizes investigated were 10, 50, 100, and 1,000. Here, the authors point out that the ADF CI had the poorest coverage, but that coverage became adequate when the sample size increased to 1,000. Interestingly, two of the other CIs investigated had the best coverage performance: the Bonett (2003) and Fisher (1950). This result is noteworthy given that these two CIs are the easiest to compute, and the Fisher CI was specifically developed for the product-moment correlation instead of coefficient alpha.

Padilla, Divers, and Newton (2012) proposed coefficient alpha bootstrap CIs and compared them to other CIs in the literature. The bootstrap CIs included the normal theory bootstrap (NTB), PB, and BCa. The NTB CI assumes that the sampling distribution of coefficient alpha is normal. However, the PB and BCa CIs make no assumption about the sampling distribution of coefficient alpha. Other non-bootstrap CIs included were the NT and ADF. The study investigated normal and non-normal parallel and congeneric items in sets of 5, 10, 15, and 20 items. The items were Likert-type with 2, 3, 5, and 7 response categories. In addition, the sample sizes investigated ranged from 50 to 300 in increments of 50. If items were normally distributed or had little skewness, the NT CI had good coverage. However, the NTB CI had the best performance in that it had consistent acceptable coverage over all simulation conditions investigated; i.e., it was robust to all simulation conditions investigated.

5.2 CI Research for Coefficient Omega: Raykov (1997, 1998) proposed a coefficient omega PB CI for a unidimensional construct.

[1] For the remainder of the article, Likert-type items with 2 categories are binary items.

The PB CI with 1,000 bootstrap samples was demonstrated in a simulated data set of 6 multivariate normal items that were continuous, congeneric, and unidimensional with a sample size of 400 and $\alpha = .10$. The coefficient omega PB CI was able to capture the population parameter for composite reliability.

In a subsequent study, Raykov (2002) proposed an analytical *SE* via the delta method for coefficient omega measuring a unidimensional construct. This *SE* can then be used to estimate a delta method CI for coefficient omega. The delta method CI was compared to the PB CI with 2,000 bootstrap samples in a simulated data set of 5 multivariate normal items that were continuous, congeneric, and unidimensional with a sample size of 500. Both CIs had similar results in capturing the population parameter for coefficient omega.

Raykov and Marcoulides (2011) added a modification to the delta method CI for coefficient omega and made three recommendations for categorical items. First, the modification took the form of adding a logit transformation to keep the CI within the range of 0 to 1 (i.e., it does not exceed $[0,1]$). Second, for items with less than 5 response categories, a three-step parceling procedure was proposed. See Little, Cunningham, Shahar, and Widaman (2002) for a discussion on parceling items. Third, for non-normal items with 5 to 7 response categories, the robust maximum likelihood (MLR) estimator was suggested. The delta method CI with logit transformation and $\alpha = .05$ was illustrated on two data sets: one with a sample size of 400 and 4 items and the other with a sample size of 350 and 5 items. The delta method CI with all modifications/suggestions and $\alpha = .05$ was illustrated on a data set with a sample size of 823 and 6 items with 4 response categories. In each situation, both CIs had similar results in capturing the population parameter for coefficient omega.

Padilla and Divers (2013a, 2013b) developed two bootstrapped based CIs for coefficient omega: the BCa and normal theory bootstrap (NTB). They investigated the performance of the PB, BCa, and NTB CIs in a simulation study, which included normal and non-normal parallel and congeneric items in sets of 5, 10, 15, and 20 items. The items were Likert-type with 2, 3, 5, and 7 response categories. In addition, the sample sizes investigated ranged from 50 to 300 in increments of 50. The CIs were based on 2,000 bootstrap samples with $\alpha = .05$. Overall, the NTB CI had the best performance in that it had the most consistent acceptable coverage over all simulation conditions investigated. In terms of robustness across all simulation conditions, the following order was recommended: the NTB for $n > 50$, PB for $n \geq 100$, and BCa for $n \geq 150$. The only exception was that the NTB was still a good choice for $n = 50$ with more than 5 items.

In a subsequent study, Padilla and Divers (2016) compared the coefficient omega CIs in the literature. There were two types of CIs investigated: non-bootstrap and bootstrap. The non-bootstrap CIs included the delta method with logit transformation (DTLG) and three-step parceling with logit transformation (PRLG). The bootstrap CIs included the PB, BCa, NTB, and a new method using the bootstrap *SE* with the logit transformation (BTLG). The study included normal and non-normal congeneric items in sets of 6, 12, 18, and 24 items. The items were Likert-type with 2, 3, 5, and 7 response categories. In addition, the sample sizes investigated ranged from 50 to 500 in increments of 50. All CIs were based on $\alpha = .05$ and the bootstrap CIs had 2,000 bootstrap samples. As in the previous study, the NTB CI had the best performance as it had the most consistent acceptable coverage over all simulation conditions investigated. However, the BTLG, BCa, and PRLG CIs were good choices when $n \geq 100$.

6. Discussion: Reliability is an important property to establish for a measurement instrument. Although there are different types of reliability for different purposes, coefficient alpha has been the standard for establishing the test (composite) reliability of a measurement instrument for over 60 years. Noticing some of the shortcomings of coefficient alpha, McDonald (1970) proposed coefficient omega. However, coefficient omega has yet to reach the popularity of coefficient alpha. Regardless, until recently, neither coefficient received much research attention after their initial development. Therefore, this article provided a brief overview of the central ideas behind coefficients alpha and omega as well as a review of some of the current literature to enable applied researchers to make informed decisions about their use in various settings.

As discussed, recent years have seen more published work in developing coefficients alpha and omega. A lot of this is the result of Graham (2006), van Zyl et al. (2000), and the advent of computational methods (e.g., the bootstrap; Efron & Tibshirani, 1998). The work by Graham presented clear definition for the three CTT based models and how they relate to composite reliability (e.g., coefficients alpha and omega). Van Zyl et al. derived the asymptotic distribution of coefficient alpha, which provided a method to obtain the *SE* to construct coefficient alpha CIs for inference. Lastly, computational methods provided a flexible alternative to obtaining CIs for coefficients alpha and omega.

The coefficient alpha and omega CI research has been promising. To date, the NTB CI is the most optimal as it performed well under all conditions investigated (Padilla & Divers, 2016; Padilla et al., 2012). For example, although not explicitly discussed in the research, the coefficient alpha and omega CIs can be used for inference similar to regular CIs. For example, a researcher could examine the CI to test if the estimated coefficient alpha is significantly different from the standard acceptable alpha of .70 ($\alpha_0 = .70$; Nunnally & Bernstein, 1994). Statistical significance is met if the CI does not include $\alpha_0 = .70$. Otherwise, there is no statistical significance (i.e., the alpha value could equal .70).

Despite the recent advances, there is still more to learn about coefficients alpha and omega. Although not presented here, Bayesian methods also offer an alternative for interval estimation (Gelman, 2004; Lee, 2004). Bayesian methods provide a few advantages, but two stand out. First, prior information can be incorporated about parameter(s) in a model, which allows for a combination of data and prior information to learn more about the phenomenon under study. Second, Bayesian credible intervals (BCIs) can be generated, which are analogs to CIs. However, BCIs are more intuitive in that they provide the interpretation most researchers seek from CIs. In the context of this article, Padilla and Zhang (2011) developed a Bayesian coefficient alpha with corresponding BCIs. Given all that has been presented, and the continuing computational advancements, there is more coefficient alpha and omega research yet to come.

References:

1. Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Pub. Co.

2. 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. doi:10.1207/s15327906mbr3202_4

3. Blinkhorn, S. F. (1997). Past imperfect, future conditional: Fifty years of test theory. *British Journal of Mathematical and Statistical Psychology, 50*(2), 175-185. doi:10.1111/j.2044-8317.1997.tb01139.x

4. Bonett, D. G. (2003). Sample size requirements for comparing two alpha coefficients. *Applied Psychological Measurement, 27*(1), 72-74. doi:10.1177/0146621602239477

5. Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart, and Winston.

6. Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. doi:10.1007/bf02310555

7. Derogatis, L., Lipman, R., Rickels, K., Uhlenhuth, E., & Covi, L. (1974). The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. *Behavioral science, 19*(1), 1-15.

8. Duhachek, A., & Iacobucci, D. (2004). Alpha's standard error (ASE): An accurate and precise confidence interval estimate. *Journal of Applied Psychology, 89*(5), 792-808. doi:10.1037/0021-9010.89.5.792

9. Efron, B., & Tibshirani, R. (1998). *An introduction to the bootstrap.* Boca Raton FL: CRC Press.

10. Feldt, L. S. (1965). The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. *Psychometrika, 30*(3), 357-370. doi:10.1007/bf02289499

11. Fisher, R. A. (1950). *Statistical methods for research workers*. Edinburgh, UK: Oliver & Boyd.

12. Furr, R. M. (2018). *Psychometrics : An introduction* (3rd ed.). Los Angeles: SAGE.

13. Gelman, A. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, Fla.: Chapman & Hall/CRC.

14. 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. doi:10.1177/0013164406288165

15. Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika, 10*(4), 255-282. doi:10.1007/bf02288892

16. 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. doi:10.1177/00131640021970691

17. Hogg, R. V., Tanis, E. A., & Zimmerman, D. L. (2015). *Probability and statistical inference* (9th ed.). Boston: Pearson.

18. Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2*(3), 151-160. doi:10.1007/BF02288391

19. Lee, P. M. (2004). *Bayesian statistics: An introduction* (3rd ed.). London: Arnold.

20. Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 151-173. doi:10.1207/s15328007sem0902_1

21. Lord, F., Novick, M., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

22. Maydeu-Olivares, A., Coffman, D. L., & Hartmann, W. M. (2007). Asymptotically distribution-free (ADF) interval estimation of coefficient alpha. *Psychological Methods, 12*(2), 157-176. doi:10.1037/1082-989x.12.2.157

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

24. McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: L. Erlbaum Associates.

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

26. Padilla, M. A., & Divers, J. (2013a). Bootstrap interval estimation of reliability via coefficient omega. *Journal of Modern Applied Statistical Methods, 12*(1), 78-89. doi:10.22237/jmasm/1367381520

27. Padilla, M. A., & Divers, J. (2013b). Coefficient Omega bootstrap confidence intervals: Nonnormal distributions. *Educational and Psychological Measurement, 73*(6), 956-972. doi:10.1177/0013164413492765

28. 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. doi:10.1177/0013164415593776

29. Padilla, M. A., Divers, J., & Newton, M. (2012). Coefficient Alpha bootstrap confidence interval under nonnormality. *Applied Psychological Measurement, 36*(5), 331-348. doi:10.1177/0146621612445470

30. Padilla, M. A., & Zhang, G. (2011). Estimating internal consistency using Bayesian methods. *Journal of Modern Applied Statistical Methods, 10*(1), 277-286. doi:10.22237/jmasm/1304223840

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

32. Raykov, T. (1998). A method for obtaining standard errors and confidence intervals of composite reliability for congeneric items. *Applied Psychological Measurement, 22*(4), 369-374. doi:10.1177/014662169802200406

33. Raykov, T. (2002). Analytic estimation of standard error and confidence interval for scale reliability. *Multivariate Behavioral Research, 37*(1), 89-103. doi:10.1207/s15327906mbr3701_04

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

35. Raykov, T., & Shrout, P. E. (2002). Reliability of scales with general structure: Point and interval estimation using a structural equation modeling approach. *Structural Equation Modeling, 9*(2), 195-212. doi:10.1207/s15328007sem0902_3

36. Romano, J. L., Kromrey, J. D., & Hibbard, S. T. (2010). A Monte Carlo study of eight confidence interval methods for coefficient alpha. *Educational and Psychological Measurement, 70*(3), 376-393. doi:10.1177/0013164409355690

37. Tarkkonen, L., & Vehkalahti, K. (2005). Measurement errors in multivariate measurement scales. *Journal of Multivariate Analysis, 96*(1), 172-189. doi:10.1016/j.jmva.2004.09.007

38. van Zyl, J., Neudecker, H., & Nel, D. (2000). On the distribution of the maximum likelihood estimator of Cronbach's alpha. *Psychometrika, 65*(3), 271-280. doi:10.1007/bf02296146

39. Yu, C. H. (2001). *An introduction to computing and interpreting Cronbach Coefficient Alpha in SAS*. Paper presented at the 26th SAS User Group International (SUGI) Conference, Cary, NC. http://www2.sas.com/proceedings/sugi26/p246-26.pdf

40. Yuan, K.-H., & Bentler, P. (2002). On robusiness of the normal-theory based asymptotic distributions of three reliability coefficient estimates. *Psychometrika, 67*(2), 251-259. doi:10.1007/bf02294845

41. Yuan, K.-H., Guarnaccia, C. A., & Hayslip, B., Jr. (2003). A study of the distribution of sample coefficient alpha with the Hopkins Symptom Checklist: Bootstrap versus asymptotics. *Educational and Psychological Measurement, 63*(1), 5-23. doi:10.1177/0013164402239314

42. 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. doi:10.1177/0013164493053001003

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

