Received: 25 February 2022 | Accepted: 13 March 2022
DOI: 10.1002/cfp2.1138

INVITED REVIEW

WILEY

The alpha and omega of financial risk-tolerance assessment

Wookjae Heo[1] | **Abed Rabbani**[2] | **John E. Grable**[3] | **Michael Roszkowski**[4]

[1]Division of Consumer Sciences, Purdue University, West Lafayette, Indiana, USA
[2]Personal Financial Planning, University of Missouri, Columbia, Missouri, USA
[3]Department of Financial Planning, Housing and Consumer Economics, University of Georgia, Athens, Georgia, USA
[4]Cherry Hill, New Jersey

Correspondence
John E. Grable, Department of Financial Planning, Housing and Consumer Economics, University of Georgia, 205 Dawson Hall, Athens, GA 30602, USA.
Email: grable@uga.edu

Abstract
Over the past three decades, numerous scaling and attitudinal measurement techniques have been developed to facilitate the assessment of an individual's financial risk tolerance. Cronbach's alpha has traditionally been used as the primary measure of scale reliability for assessment tools that have been developed using classical psychometric theory. Recently, however, psychometricians have raised concerns about the ongoing use of Cronbach's alpha as a robust measure of scale reliability. In its place, some have argued that reliability estimates should be based on greatest lower bound (GLB) and omega estimations. The purpose of this paper is to describe and compare these alternative reliability measures to Cronbach's alpha for a widely used research-focused financial risk-tolerance scale. Using a dataset with 179,450 observations, findings from this study suggest that while estimates based on Cronbach's alpha, omega, and the GLB do differ, for the most part, reliability estimates across the measures are more similar than dissimilar.

KEYWORDS
Cronbach's alpha, financial risk tolerance, greatest lower bound, internal consistency, omega, reliability

1 | INTRODUCTION

As noted by Danner et al. (2016), measurement instruments often serve as the foundation for research and practice across the social sciences. An important element associated with the appropriate use of measurement instruments involves finding and reporting suitable reliability estimations. This is particularly true in situations where policy, research, and consumer decisions rely on measurement scores. Danner et al. additionally argued that "meaningful comparisons between groups can only be made if measured scores have the same meaning across groups" (p. 177). This insight has particular importance when considering the way financial attitudes—particularly risk tolerance—are measured and used in practice.

The tradition of psychometric scale development and use to assess the attitudes of consumers, investors, and financial decision-makers has a long and robust history (Folke et al., 2021; Hanna et al., 2008; Heo et al., 2020; Khoshouei, 2009; Lay & Furnham, 2018; Roszkowski et al., 2005). Of particular importance in this domain is the assessment of financial risk tolerance as an antecedent to the financial and investment decision-making process (Rubio et al., 2010). As a personal attitudinal construct, financial risk tolerance can be conceptualized as the maximum level of uncertainty someone is willing to endure when the outcomes associated with the decision are both uncertain and potentially negative (Nobre & Grable, 2015).

Nearly all worldwide regulatory agencies that oversee securities markets mandate that financial advisors assess and evaluate the willingness of clients to engage in risky behavior before making an investment recommendation or implementing specific financial management investment strategies. Consider, for example, Financial

Industry Regulatory Authority (FINRA) customer due diligence requirements in the United States.[1] Customer due diligence rules require financial advisors to take steps to understand the nature and purpose of a customer relationship. FINRA rules state that a firm or advisor must[2]:

> Have a reasonable basis to believe that a transaction or investment strategy involving securities that are or have been recommended to an investor are suitable … a reasonable basis underlying suitability is due diligence to assess an investor's investment profile, which can be comprised of a *risk-tolerance assessment*, an investor's age, other investments, financial needs, tax status, investment objectives, experience, time horizon, liquidity need, and other factors.

Similarly, the European Securities and Market Authority mandates that a financial advisor must.

> [when providing investment advice or portfolio management] … obtain the necessary information regarding the client's or potential client's knowledge and experience in the investment field relevant to the specific type of product or service, that person's financial situation including his (sic) ability to bear losses, and his (sic) investment objectives including his (sic) risk tolerance so as to enable the investment firm to recommend to the client or potential client the investment services and financial instruments that are suitable for him (sic) and, in particular, are in accordance with his (sic) risk tolerance and ability to bear losses.[3]

In Canada, the Investment Industry Regulatory Organization of Canada (IIROC) has a similar rule whereby a financial advisor must be able to demonstrate that an investor's willingness and ability to take risk is measured and evaluated before the development of an investment recommendation (Brayman et al., 2015). Internationally, the Chartered Financial Analyst (CFA) Institute, through the organization's Standards of Practice Handbook (11th edition), directs the following[4]:

> When members and candidates are in an advisory relationship with an investor, they must … make a reasonable inquiry into an investor's or prospective investor's investment experience, risk and return objectives, and financial constraints prior to making any investment recommendation or taking investment action and must reassess and update this information regularly.

When viewed from a regulatory and practice management perspective, the primary purpose underlying the assessment of client financial risk tolerance is to ensure that investment and financial recommendations match an investor's emotional and financial ability to take on household-level financial/investment risk. A secondary purpose associated with measuring financial risk tolerance involves providing a financial advisor with information that can be used to engage an investor in meaningful investment and financial planning discussions.

Traditionally, financial risk tolerance has been measured using psychometrically-sound scales developed using aspects of classical test theory (CTT). A key assumption underlying the use of scale based on CTT is that the scale is reliable. As noted by Sullivan (2011, p. 119), "Reliability refers to whether an assessment instrument gives the same results each time it is used in the same setting with the same type of subjects. Reliability essentially means consistent or dependable results." When viewed this way, reliability is a necessary prerequisite to determining a scale's validity.

Numerous means can be used to evaluate the strength of a scale's reliability. A powerful measure of reliability relies on the evaluation of test–retest scores. Principally, this test compares, via correlation, scores generated by one person on a scale to scores on the same scale, by the same person, at a different point in time. When viewed across a group of test-takers, a reliable scale will produce similar pre- and post-test scores. In practice, test–retest reliability estimates are difficult to obtain (Kruyen et al., 2013). First, few risk-tolerance questionnaires are included in panel datasets, making it difficult to obtain pre- and post-test scores from the same respondents. Similarly, few financial advisors regularly retest clients after an initial financial risk-tolerance test has been administered. Second, obtaining acceptable test–retest scores requires that differences in test-taking situations be minimized across administrations of the test, which can be difficult to achieve in practice.

Another measure of reliability is known as interrater reliability. This reliability index relies on the expert opinion of professionals who have knowledge about the outcome assessed by a scale. Highly reliable scales, tests, and questionnaires will exhibit a strong kappa or Kendall tau test score. In the realm of financial risk tolerance, interrater reliability is less applicable because risk tolerance, as a trait-type factor, is difficult to assess via third-party observation. It is also challenging for a financial advisor to gather enough people knowledgeable enough about a particular individual to provide an informed opinion about one client or a group of clients.

Among currently available measures of reliability, Cronbach's alpha tends to be the most widely applied and reported in the literature (Sijtsma, 2009b; Trizano-Hermosilla & Alvarado, 2016). This measure of reliability was formalized by Cronbach (1951) as:

$$\alpha = K \div (K-1) \times 1 - \sum\nolimits_{i=1}^{K} \sigma_{Y_i}^2 \div \sigma_X^2 \tag{1}$$

where K is the sum of items in the scale, $X = Y_1 + Y_2 + \ldots + Y_K$, σ_X^2 is the variance of the total test scores, and $\sigma_{Y_i}^2$ is the variance of item i.

Perhaps a more intuitively appealing alternative formula for calculating Cronbach's alpha is the following:

$$\alpha = \frac{K_{\bar{c}}}{(\bar{v} + (K-1)\bar{c})} \tag{2}$$

where K is equal to the number of items, c-bar is the average inter-item covariance among the items, and v-bar equals the average variance. This formula shows that alpha is a function of the number of items, their variances, and their covariances. Inspection of this equation reveals that as the number of items and their covariance increases, alpha becomes larger.

In the context of Cronbach's alpha, reliability is sometimes referred to as the internal structure of a scale or as internal consistency (Sullivan, 2011). In its most basic form, Cronbach's alpha is based on the correlation values among questions included in a scale. A scale's reliability can theoretically range from zero to 1.0, with higher scores (0.70 or higher) suggesting a more robust level of scale reliability (see DeVellis, 1991).

Over the past decade, psychometricians have published concerns about the ongoing use of Cronbach's alpha as a vigorous measure of scale reliability (see Raykov & Marcoulides, 2019). In its place, some have argued that reliability estimates should be based on greatest lower bound (GLB) and omega. The purpose of this paper is to describe these alternative reliability measures and provide a comparison of reliability based on Cronbach's alpha, GLB, and omega for a widely used financial risk-tolerance scale.

2 | BACKGROUND

Cronbach's alpha, while an extensively used index of internal consistency reliability, is nonetheless a statistic whose subtleties continue to be debated by psychometricians. Experts continue to make conflicting claims regarding what alpha actually measures and which factors are responsible for alpha's value under different parameters (e.g., Cho & Kim, 2015; Cortina, 1993; Davenport et al., 2015; Peterson, 1994; Schmitt, 1996; Sijtsma, 2009a; 2009b; 2015; Streiner, 2003; Tavakol & Dennick, 2011). Even Cronbach's view about alpha evolved over the years (Cronbach & Shavelson, 2004).

The use of Cronbach's alpha as a measure of reliability is premised on three assumptions, all of which tend to be violated to some degree in practice. It is said that Cronbach's alpha is frequently misapplied because its assumptions are not met (Raykov, 1997; Sijtsma, 2009b; Tavakol & Dennick, 2011). Strictly speaking, alpha assumes a true-score equivalence ("tau equivalent") model, which requires that: (1) a single latent trait (factor) underlie the scale, (2) the items forming the scale have equal variances, and (3) the covariances among these items are the same (Maydeu-Olivares et al., 2007). If these conditions are not met (and they seldom are), alpha will underestimate reliability. Skewed data can also reduce alpha estimates (Greer et al., 2006).

In addition to reporting Cronbach's alpha, statistical packages (such as SPSS) also calculate the item-total correlation for each item. This is the Pearson correlation between an item and the total (composite) score formed on the basis of the remaining items. Because the particular item under analysis is excluded from the composite, the correlation is also called a "corrected item-total correlation" or an "item-remainder correlation." Next to each item-total correlation the output also reports a "Cronbach's Alpha if Item is Deleted" figure. This value is the Cronbach's alpha that would result if the given item were to be excluded from the composite.

Generally, the elimination of an item with a low item-total correlation leads to an improved Cronbach's alpha. Conversely, eliminating an item with a high item-total correlation will typically result in a reduction in the magnitude of Cronbach's alpha. In other words, the general practice in test construction is to eliminate items where the alpha without a particular item in the mix is higher than the actual alpha when the item is part of the composite. However, some researchers warn that if applied indiscriminately, this practice can lead to inflated alphas (Kopalle & Lehmann, 1997) or loss of predictive validity (Raykov, 2008).

Based on these objections to Cronbach's alpha, Peters (2014) recommended that researchers "abandon" (p. 60) the use of Cronbach's alpha altogether. In practice, this is not likely, but the strength of objections to the concept of internal consistency suggests that researchers, and those who rely on the use of scales and questionnaires in practice, ought to consider the utility of alternatives to Cronbach's alpha. One alternative is called GLB.

Sijtsma (2009a) equated GLB to the lowest possible reliability score a scale can generate, but Revelle and Zinbarg (2009) argued that GLB is a misnomer and that GLB provides a reliability index that is systematically lower than omega but it is not the GLB of reliability. Reliability is an interval (Peters, 2014). Based on the work of Woodhouse and Jackson (1977), GLB can be expressed as:

$$GLB = 1 - tr(C_e) \div \sigma_x^2 \tag{3}$$

where σ_x^2 is the scale variance and $tr(C_e)$ is the trace of the inter-item error covariance matrix. It is worth noting that with small samples, GLB can overestimate the true reliability of a scale or psychometric measure (Trizano-Hermosilla & Alvarado, 2016).

GLB is seldom reported in the literature. This absence has less to do with its applicability and more to do with the fact that few statistical software packages are programmed to estimate GLB. A more widely available alternative measure of reliability is omega. McDonald (1978, 1999) proposed omega as a substitute to Cronbach's alpha. McDonald argued that omega provides greater flexibility because the estimation procedure relaxes assumptions regarding tau-equivalency (see Padilla, 2019 for a discussion of this and other related issues). Omega can be expressed as follows:

$$\begin{aligned}\omega_t &= \left(\sum \lambda_j\right)^2 \div \left[\left(\sum \lambda_j\right)^2 + \sum \left(1 - \lambda_j^2\right)\right] \\ &= \left(\sum \lambda_j\right)^2 \div \left[\left(\sum \lambda_j\right)^2 + \left(\sum \varphi\right)\right]\end{aligned} \tag{4}$$

where λ_j is the loading of an item j, λ_j^2 is the communality of item j, and φ equates to the uniqueness (see Trizano-Hermosilla & Alvarado, 2016).

The purpose of the present study was to compare Cronbach's alpha, GLB, and omega in describing the reliability of a financial risk-tolerance scale. The results from the analysis, which are described below, suggest that the tested scale's reliability falls within an acceptable range no matter how assessed. But the primary takeaway from this study is that while psychometricians and methodologists continue to debate the most appropriate way to measure scale reliability, the three reliability measures, in the context of financial risk-tolerance assessment, appear quite similar. The remainder of this paper describes the methods used to compare the three reliability approaches, the results from the comparison tests, and the findings in the context of measuring the willingness of investors to take a financial risk.

3 | METHODS

The methodological process used in this study follows that recommended by Peters (2014). Peters suggested that researchers adhere to the following six-step diagnostic process when evaluating the reliability of a scale or questionnaire (p. 65):

1. Compute Cronbach's alpha, omega, and the GLB, preferably with confidence intervals;

TABLE 2 Risk-tolerance scale item descriptives

Item	M	SD	Mdn	Skew	Kurtosis
1	2.67	0.70	3	−0.18	−0.13
2	2.08	0.87	2	0.62	−0.18
3	2.16	0.92	2	0.06	−1.19
4	2.03	0.83	2	−0.06	−1.54
5	1.82	0.67	2	0.23	−0.80
6	2.36	0.79	2	0.41	−0.21
7	1.85	0.75	2	0.69	0.30
8	2.58	0.86	3	−0.21	−0.59
9	1.83	0.99	1	0.35	−1.88
10	2.34	0.94	3	−0.73	−1.46
11	2.01	0.96	2	0.59	−0.65
12	1.75	0.66	2	0.32	−0.78
13	2.03	0.77	2	0.47	−0.04

TABLE 1 Alpha, omega, and the GLB for the risk-tolerance scale under two assumptions of level of measurement

	Interval assumption		Ordinal assumption	
	Point estimate	Confidence interval	Point estimate	Confidence interval
Cronbach's Alpha	0.73	0.72, 0.73	0.79	0.79, 0.79
Omega	0.73	0.72, 0.73	0.79	0.79, 0.79
Revelle's Omega	0.76			
GLB	0.71			

2. Conduct a factor analysis or principal component analysis and inspect all eigenvalues and the factor loadings (at least for the first factor);
3. Check the means, medians, and variances for each item;
4. Generate and evaluate a correlation matrix;
5. Examine the scatter plots of the associations between all items; and
6. Inspect histograms of each item's distribution.

Two types of analyses were used to estimate omega in this study. The first estimate used the traditional omega formula (McDonald, 1978). The second estimate was based on Revelle's omega (see Revelle & Zinbarg, 2009).[5] Generally, omega is assumed to be the same as Cronbach's alpha if the assumptions used to calculate Cronbach's alpha (e.g., all factor loadings are the same [unidimensionality]) are not violated. However, if any assumptions about the use of Cronbach's alpha are not met, the estimation of omega and Cronbach's alpha will likely differ. The process of checking the robustness of omega is as follows: conduct an exploratory factor analysis (EFA), followed by a descriptive analysis, an evaluation of the item correlation matrix, and a review of scatter plots and histograms.[6]

3.1 | Data used for the trials

Data for this study were obtained from an open-access online investment risk-tolerance survey. The dataset was cross-sectional ($n = 179{,}450$). The survey included 13 risk-tolerance items from a scale originally published by Grable and Lytton (1999). This scale was the basis for the investigation because scores derived from the questionnaire have been used extensively in published studies. Additionally, it was thought that supplementary evidence (or lack thereof) of the scale's reliability would be beneficial to others who may employ the scale for research and practice.

The mean and standard deviation of risk-tolerance scores were 27.50 and 5.22, respectively. In addition to risk-tolerance questions, the survey asked each survey participant to provide basic demographic data. Among the 179,450 participants, 56% self-identified as female and 44% self-identified as male. Approximately 24% of participants reported living with a spouse or partner at the time of survey completion. In terms of income, the distribution was: 25% less than $25,000, 20% between $25,000 and $49,999, 18% between $50,000 and $74,999, 12% between $75,000 and $99,999, and 25% over $100,000. Survey responses were collected over the period 2015 through 2017 through an online open-access survey.[7]

The following discussion describes the analytical process used in this study and the corresponding results from these analyses. Major programming tools used to calculate Cronbach's alpha, omega, Revelle's omega, and the GLB were R-studio and Stata 15.1.

4 | RESULTS

4.1 | Step 1. Compute Cronbach's alpha, omega, and the GLB

Table 1 shows the Cronbach's alpha, omega, and GLB values, with corresponding 95% confidence intervals, for

TABLE 4 Correlation estimates among scale items

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13
Q1	1.00												
Q2	0.22	1.00											
Q3	0.15	0.16	1.00										
Q4	0.20	0.22	0.08	1.00									
Q5	0.19	0.19	0.09	0.39	1.00								
Q6	0.30	0.22	0.18	0.11	0.17	1.00							
Q7	0.13	0.15	0.14	0.09	0.11	0.18	1.00						
Q8	0.21	0.27	0.16	0.23	0.20	0.21	0.15	1.00					
Q9	0.15	0.23	0.12	0.12	0.09	0.16	0.12	0.19	1.00				
Q10	0.10	0.12	0.09	0.12	0.08	0.10	0.06	0.13	0.09	1.00			
Q11	0.17	0.19	0.11	0.39	0.23	0.12	0.15	0.19	0.12	0.11	1.00		
Q12	0.23	0.27	0.15	0.33	0.30	0.23	0.18	0.28	0.17	0.12	0.25	1.00	
Q13	0.20	0.26	0.17	0.14	0.16	0.24	0.22	0.23	0.16	0.09	0.18	0.24	1.00

TABLE 3 Alpha, omega, and the GLB estimates on each risk-tolerance factor under two assumed levels of measurement

	Factor 1 (risk comfort and experience)				**Factor 2 (investment loss aversion)**			
	Interval assumption				Ordinal assumption			
	Point Estimate	Confidence Interval	Point Estimate	Confidence Interval	Point Estimate	Confidence Interval	Point Estimate	Confidence Interval
Cronbach's Alpha	0.64	0.64, 0.64	0.71	0.71, 0.71	0.64	0.64, 0.64	0.72	0.72, 0.72
Omega	0.64	0.64, 0.64	0.71	0.71, 0.71	0.60	0.65, 0.66	0.73	0.73, 0.73
Revelle's Omega	0.67				0.69			
GLB	0.66				0.68			

the risk-tolerance scale. Two estimates are presented. The first estimate was based on an assumption that the items in the scale represent interval measures. The second estimate was based on an assumption that the items are ordinal.

In the case of the risk-tolerance scale evaluated in this study, it could be argued that the items are ordinal rather than interval in nature. The scale items were all multiple-choice questions that were measured using response categories. The results shown in Table 1 indicate that the scale's reliability was similar across the measures. The values were higher under the ordinal scale assumption. Based on normative interpretations, the scores represent an acceptable level of reliability (DeVellis, 1991; Nunnally & Bernstein, 1994).

4.2 | Step 2. Conduct a factor analysis or principal component analysis and inspect all Eigenvalues and the factor loadings

The exploratory factor analysis indicated that the scale was comprised of two factors. The first factor included items 1, 2, 3, 6, 7, 8, 9, and 13 (labeled as risk comfort and experience). The second factor included items 4, 5, 10, 11, and 12 (labeled as investment loss aversion). The SS loadings, the proportion of variance, and cumulative variance were 2.37, and 1.91, 0.18, and 0.15, 0.18, and 0.33, respectively, for the two factors.

4.3 | Step 3. Check the means, medians, and variances for each item

Table 2 shows the mean, standard deviation, median, skewness, and kurtosis for each item in the risk-tolerance scale.

4.4 | Reevaluation of Step 2

A primary assumption with the use of any reliability measure is that of unidimensionality. As noted by Peters (2014), this assumption rarely holds in practice. Nearly all psychological, educational, health, and financial scales include correlated items that represent multiple dimensions of behavior or attitudes. Peters argued that "aggregating these measures despite the clear lack of unidimensionality is warranted on the basis of theory" (p. 63). Others contend that when a scale lacks unidimensionality, it needs to be divided into subscales (Dunn et al., 2013, p. 406). When split into distinct components, new estimates of Cronbach's alpha, omega, and the GLB

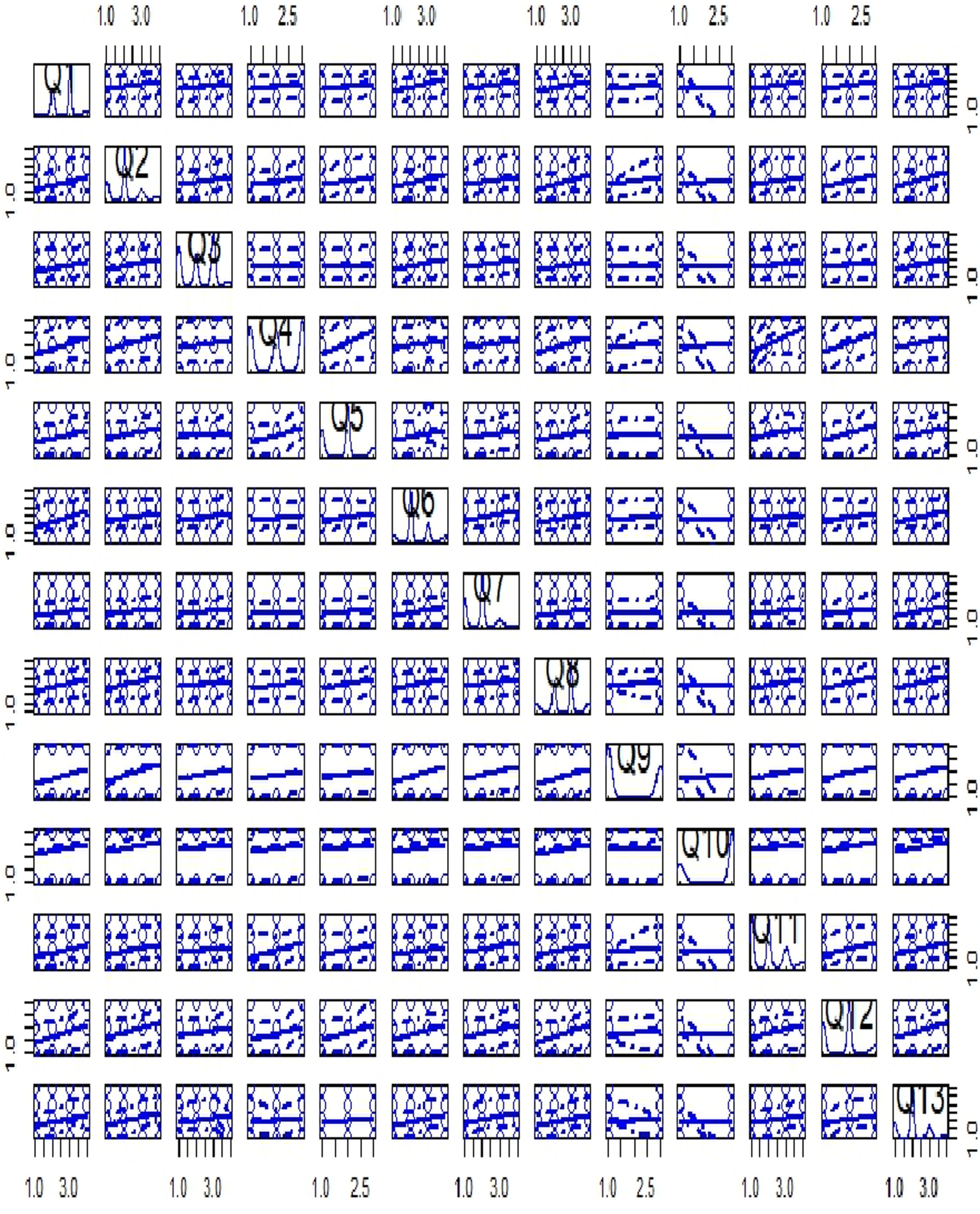

FIGURE 1 Bivariate scatterplots for combinations of all items in the scale

should be made. As noted in the narrative related to Table 1, the risk-tolerance scale was not unidimensional. Two factors were identified. As such, estimates of Cronbach's alpha, omega, and the GLB were re-estimated for each factor. Results of the re-estimations are shown in Table 3.

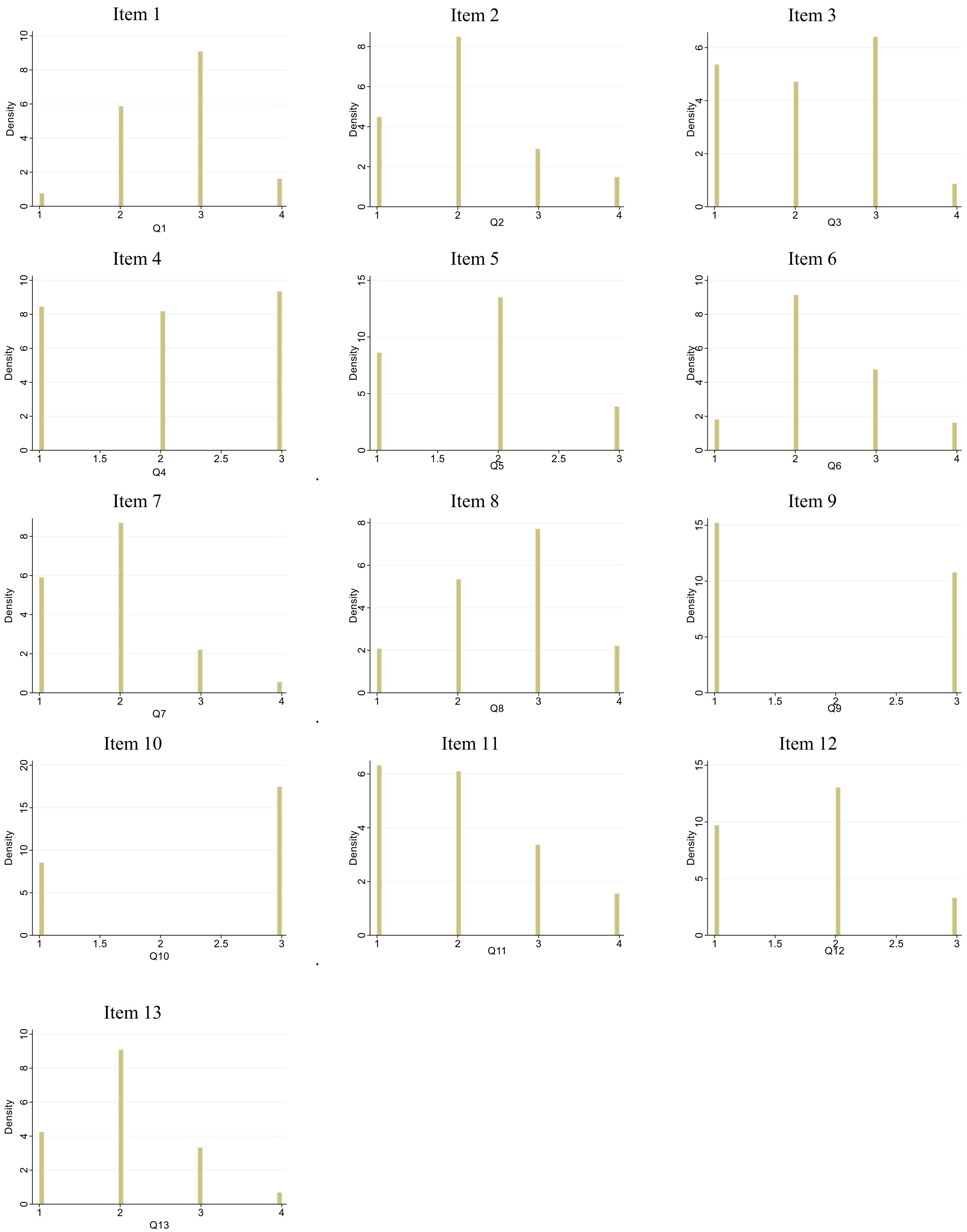

FIGURE 2 Univariate histograms for items in the scale

Examination of data from Table 3 leads to two observations. First, reliability estimates were lower for the two factors than the overall reliability estimates for the summated scale. This was not surprising given that the larger scale was comprised of a greater number of items. Second, the reliability estimates based on the assumption that the scale items are ordinal resulted in stronger estimates of reliability.

4.5 | Step 4. Generate and evaluate a correlation matrix

The next step in the analytical process involved evaluating the bivariate associations between the items comprising the scale. As shown in Table 4, the items were positively correlated. Each of the associations was statistically significant at the $p < 0.01$ level. Coefficients ranged from 0.06 to 0.39, with a median of 0.17.

4.6 | Steps 5 and 6. Examine the scatter plots of the associations between all items and inspect histograms of each item's distribution

The final analytical procedure involved checking the distribution of answers to each question to confirm that the general pattern of response was linear, from low to high-risk tolerance. Figure 1 shows the bivariate scatterplots for all combinations among the 13 risk-tolerance items. Figure 2 shows the univariate histograms of the 13 items. As depicted in Figure 1, the associations between variables were visually similar to the analytical correlations in Table 4; that is, responses were linearly associated. Figure 2 illustrates how answers were distributed among the items.

5 | DISCUSSION AND IMPLICATIONS

International practice and investment management standards mandate that financial advisors evaluate the financial risk tolerance of clients before making investment and financial recommendations. Although few of the criteria for practice prescribe the exact manner in which a financial advisor should evaluate a client's risk tolerance, nearly all approaches rely on psychometric scaling methodologies, most often based on classical test theory (CTT). A fundamental assumption associated with the use of any psychometric scale is that the scale is reliable. The dominant index of reliability, as typically reported in research studies, is Cronbach's alpha. Some psychometricians have expressed concerns about the use of Cronbach's alpha as a measure of scale reliability. The purpose of this research was to evaluate estimates of Cronbach's alpha against estimates of the GLB and omega—the two most widely recommended reliability alternatives to Cronbach's alpha—for a widely used financial risk-tolerance scale.

Results from this analysis suggest that while estimates derived for Cronbach's alpha, omega, and the GLB did differ, for the most part, the estimates were very close. The GLB provided the most conservative estimate for reliability. The GLB is based on (a) the sum of the inter-item covariance matrix for item true scores and (b) the inter-item error covariance matrix (Trizano-Hermosilla & Alvarado, 2016), and as such, the GLB is most applicable to scales comprised of non-homogeneous items. In this study, the GLB provided a robust estimate of the scale's lowest level of reliability. Reliability estimates based on omega were almost identical to Cronbach's alpha reliability estimates. A number of other researchers (e.g., Savalei & Reise, 2019; Ventura-León, 2019) have also demonstrated that under an assumption of tau-equivalence, alpha (α) and omega (ω) provide nearly identical reliability estimates. Both alpha and omega are susceptible to distortion by outlier observations (Zhang & Yuan, 2016). Four takeaways are of particular interest in relation to this study. First, the risk-tolerance scale that was evaluated appears to provide a minimally acceptable level of reliability. Across the three estimates, 0.70 appears to be a reasonable minimum expectation for reliability. Second, given the question format within the scale, researchers and practitioners who use this or a similar scale could reasonably argue that items represent an ordinal level of measurement. When this assumption is accepted a higher level of reliability can be realized. Third, while the three measures of reliability did lead to slightly varying estimates of reliability, the approaches triangulated closely in the final analysis. Fourth, in alignment with Raykov and Marcoulides (2019) as well as Savalei and Reise (2019), findings from this study suggest that the abandonment of Cronbach's alpha, as proposed by some researchers (e.g., McNeish, 2018; Peters, 2014), appears to be premature and overstated. The response Mark Twain wrote in response to the notice about his death—"The report of my death was an exaggeration"—applies to Cronbach's alpha (Petsko, 2018).

The following insights and recommendations apply more broadly. First, Cronbach's alpha will underestimate reliability unless (1) there is one factor underlying the scale, and (2) all the items load about the same on that factor. Omega will give a more realistic estimate of reliability than Cronbach's alpha when: (1) there is more than one factor in the scale (i.e., a multifactorial scale), or (2) there is one factor underlying the scale but the items load differently on that factor (i.e., congeneric items). Generally, differences between Cronbach's alpha and omega tend to be rather small, unless one item loads very highly on the scale and the other items have very low factor loadings. Typically, this type of situation is rarely encountered, which is one reason that Cronbach's alpha and omega were so similar in this study and many others.

Normality may also be a consideration. Sheng and Sheng (2012) demonstrated that although non-normal score distributions can distort coefficient alpha in small samples, this can be corrected by increasing the sample size for its calculation. Based on the analysis by Trizano-Hermosilla and Alvarado (2016), researchers can rely on either omega (first choice) or alpha (second choice) when the test scores are normally distributed, but if scale skewness is low to moderate, GLB is the preferred measure. GLB is also recommended when the percentage of items in the scale with asymmetric (non-normal) distributions is high.

The findings from this study, and the resulting discussion, have direct implications for those who provide investment advice to individuals and households. As noted above, measuring financial risk tolerance is not only a best practice but also a required regulatory task. It is unlikely, however, that a financial advisor will be in a position to estimate and compare reliability scores for the financial risk-tolerance tool they use in their practice. This does not mean, however, that a financial advisor can dismiss issues of reliability and validity. A financial advisor needs to have confidence in the way risk tolerance is measured. One way to ensure that the assessment of a client's financial risk tolerance is robust is to ask a questionnaire's developer to provide evidence of scale reliability. Based on the results from this study, a financial advisor should look for estimates of reliability based on either Cronbach's alpha, GLB, or omega, Depending on the tool's particular characteristics, one of these reliability measures may be preferable.

ORCID

Wookjae Heo https://orcid.org/0000-0001-5807-9792
Abed Rabbani https://orcid.org/0000-0001-7647-8973
John E. Grable https://orcid.org/0000-0001-7093-8510
Michael Roszkowski https://orcid.org/0000-0002-9528-002X

ENDNOTES

[1] See FINRA's Customer Due Diligence Requirements for Financial Institutions (CDD Rule) and FINRA Rule 3310.

[2] Based on FINRA rules, investor financial risk profiles may consist of individualized risk scoring that allows an investor to be categorized into an appropriate investment classification.

[3] See European Securities and Market Authority. (n.d.).

[4] CFA Institute: https://www.cfainstitute.org/-/media/documents/code/code-ethics-standards/standards-practice-handbook-11th-ed-eff-July-2014-corr-sept-2014.ashx

[5] https://personality-project.org/revelle/publications/revelle.zinbarg.08.pdf

[6] If the sample used for computing the three estimates of reliability is not as large as that used in this analysis, it would make sense to test for statistical differences among them (Deng & Chan, 2017).

[7] https://pfp.missouri.edu/research/investment-risk-tolerance-assessment/

REFERENCES

Brayman, S., Finke, M., Bessner, E., Grable, J., Griffin, P., & Clement, R. (2015). *Current practices for risk profiling in Canada and review of global best practices*, prepared for the Investor Advisory Panel of the Ontario Securities Commission. http://www.osc.gov.on.ca/documents/en/Investors/iap_20151112_risk-profiling-report.pdf

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

Cortina, J. (1993). What is coefficient alpha: An examination of theory and applications. *Journal of Applied Psychology*, *78*, 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.

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, *64*, 391–418.

Danner, D., Blasius, J., Breyer, B., Eifler, S., Menold, N., Paulhus, D. L., Rammstedt, B., Roberts, R. D., Schmitt, M., & Ziegler, M. (2016). Current challenges, new developments, and future directions in scale construction. *European Journal of Psychological Assessment*, *32*, 175–180.

Davenport, E. C., Davison, M. L., Liou, P.-Y., & Love, Q. U. (2015). Reliability, dimensionality, and internal consistency as defined by Cronbach: Distinct albeit related concepts. *Educational Measurement: Issues and Practice*, *34*, 4–9. https://doi.org/10.1111/emip.12095

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

DeVellis, R. F. (1991). *Scale development*. Sage Publications.

Dunn, J. T., Baguley, T., & Brunsden, V. (2013). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *The British Journal of Psychology*, *105*, 399–412.

European Securities and Market Authority. (n.d.). *FCA Handbook*. https://www.handbook.fca.org.uk/handbook/COBS/9A/?view=chapter.

Folke, T., Gjorgjiovska, J., Paul, J., Jakob, L., & Ruggeri, K. (2021). Asset: A new measure of economic and financial literacy. *European Journal of Psychological Assessment*, *37*, 65–80.

Grable, J. E., & Lytton, R. H. (1999). Financial risk tolerance revisited: The development of a risk assessment instrument. *Financial Services Review*, *8*, 163–181.

Greer, T. F., Dunlap, W. P., Hunter, S. T., & Berman, M. (2006). Skew and internal consistency. *Journal of Applied Psychology*, *91*, 1351–1358. https://doi.org/10.1037/0021-9010.91.6.1351

Hanna, S. D., Waller, W., & Finke, M. (2008). The concept of risk tolerance in personal financial planning. *Journal of Personal Finance*, *7*(1), 96–108.

Heo, W., Cho, S., & Lee, P. (2020). APR financial stress scale: Development and validation of a multidimensional measurement. *Journal of Financial Therapy*, *11*(1), 2. doi:10.4148/1944-9771.1216

Khoshouei, M. S. (2009). Psychometric evaluation of the Connor-Davidson resilience scale (CF_RISC) using Iranian students. *International Journal of Testing*, *9*, 60–66.

Kopalle, P. K., & Lehmann, D. R. (1997). Alpha inflation? The impact of eliminating scale items on Cronbach's alpha. *Organizational Behavior and Human Decision Processes*, *70*, 189–197.

Kruyen, P. M., Emons, W. H. M., & Sijtsma, K. (2013). On the shortcomings of shortened tests: A literature review. *International Journal of Testing*, *13*, 223–248.

Lay, A., & Furnham, A. (2018). A new money attitudes questionnaire. *European Journal of Psychological Assessment*, *35*, 813–822.

Maydeu-Olivares, A., Coffman, D. L., & Hartmann, W. M. (2007). Asymptotically distribution-free (ADF) interval estimation of coefficient alpha. *Psychological Methods*, *12*, 157–176.

McDonald, R. P. (1978). Generalizability in factorable domains: Domain validity and generalizability. *Educational and Psychological Measurement*, *38*(1), 75–79.

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

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

Nobre, L. H. N., & Grable, J. E. (2015). The role of risk profiles and risk tolerance in shaping investor decisions. *Journal of Financial Service Professionals*, *69*(3), 18–21.

Nunnally, J. C., & Bernstein, L. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Padilla, M. A. (2019). A primer on reliability via coefficient alpha and omega. *Archives of Psychology*, *3*(8), 1–15.

Peters, G.-J. Y. (2014). The alpha and the omega of scale reliability and validity. *The European Health Psychologist*, *16*(2), 56–69.

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

Petsko, E. (2018, November 1). Reports of Mark Twain's quote about his own death are greatly exaggerated. Retrieved from: https://www.mentalfloss.com/article/562400/reports-mark-twains-quote-about-mark-twains-death-are-greatly-exaggerated

Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research*, *32*, 329–353. https://doi.org/10.1207/s15327906mbr32042

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*, 275–285. https://doi.org/10.1348/000711007X188520

Raykov, T., & Marcoulides, G. A. (2019). Thanks coefficient alpha, we still need you! *Educational and Psychological Measurement*, *79*, 200–210.

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, Beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*, 145–154.

Roszkowski, M. J., Davey, G., & Grable, J. E. (2005). Insights from psychology and psychometrics on measuring risk tolerance. *Journal of Financial Planning*, *18*(4), 66–77.

Rubio, V. J., Hernández, J. M., Zaldívar, F., Márquez, O., & Santacreu, J. (2010). Can we predict risk-taking behavior. *European Journal of Psychological Assessment*, *26*, 87–94.

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. doi:10.1525/collabra.247

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

Sheng, Y., & Sheng, Z. (2012). Is coefficient alpha robust to non-normal data? *Frontiers in psychology*, *3*, 3. https://doi.org/10.3389/fpsyg.2012.00034

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

Sijtsma, K. (2009b). Correcting fallacies in validity, reliability, and classification. *International Journal of Testing*, *9*, 167–194.

Sijtsma, K. (2015). Delimiting coefficient α from internal consistency and unidimensionality. *Educational Measurement: Issues and Practice*, *34*, 10–13. https://doi.org/10.1111/emip.12099

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18

Sullivan, G. M. (2011). A primer on the validity of assessment instruments. *Journal of Graduate Medical Education*, *3*, 119–120.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology*, *7*, 769. https://doi.org/10.3389/fpsyg.2016.00769

Ventura-León, J. (2019). Back to content-based validity. *Adicciones*, *0*(0). https://doi.org/10.20882/adicciones.1213

Woodhouse, B., & Jackson, P. H. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items. II: A search procedure to locate the greatest lower bound. *Psychometrika*, *42*, 579–591. https://doi.org/10.1007/BF02295980

Zhang, Z., & Yuan, K. H. (2016). Robust coefficients alpha and omega and confidence intervals with outlying observations and missing data: Methods and software. *Educational and Psychological Measurement*, *76*(3), 387–411. https://doi.org/10.1177/0013164415594658

How to cite this article: Heo, W., Rabbani, A., Grable, J. E., & Roszkowski, M. (2022). The alpha and omega of financial risk-tolerance assessment. *Financial Planning Review*, *5*(1), e1138. https://doi.org/10.1002/cfp2.1138