조사연구(2018. 5) 19권 2호: 73-104
DOI http://dx.doi.org/10.20997/SR.19.2.4

연구논문

Originators of Reliability Coefficients :
A Historical Review of the Originators of Reliability Coefficients Including *Cronbach's* Alpha*

Eunseong Cho[a) · Sungyong Chun[b)

The names of commonly used reliability coefficients, such as *Cronbach's* alpha, give the impression that we are expressing respect for the first developers of the formulas. However, few studies have investigated the identity of each person who first discovered each reliability coefficient from a neutral point of view. This study examines the history of reliability coefficients and presents conclusions regarding who should be credited for developing each reliability coefficient. For example, this study claims that credit for inventing the alpha formula should be awarded to Kuder and Richardson (1937) and that the merit of developing a reliability coefficient based on a unidimensional confirmatory factor analysis model should be returned to Jöreskog (1971). This study criticizes the existing names of reliability coefficients as *pseudo-historical* (i.e., not actually but having the appearance of being historical), suggesting the use of *ahistorical* (i.e., without concern for history) names instead.

* The present research has been conducted by the Research Grant of Kwangwoon University in 2017.

a) Professor, School of Business Administration, Kwangwoon University.

b) corresponding author: Associate Professor, Department of Business Administration, Dankook University.
E-mail: sychun@dankook.ac.kr.

Key words: Cronbach's alpha, coefficient alpha, the Spearman-Brown formula, composite reliability, McDonald's omega

Ⅰ. Introduction

Psychological studies routinely report reliability coefficients of test scores. For example, readers are likely familiar with at least some of the names of reliability coefficients, such as Cronbach's alpha, standardized alpha, the Spearman-Brown formula, composite reliability, and McDonald's omega. These conventional names give us the impression that we are expressing appreciation for the scholars who first developed the reliability coefficients. This study originates from the questions of whether the conventional names are historically legitimate and, if not, whether the practical benefits of continuing to use the name outweighs the lack of historical evidence.

Let us take the name Cronbach's alpha as an example. This name itself does not contain any information that might help psychologists use the formula. For individuals without background knowledge, the name Cronbach's alpha does not yield a clue regarding its meaning and function. The only possible conjecture based on the name is that Cronbach must have first proposed it. Therefore, this study aims to confirm two issues: (1) whether Cronbach was the researcher who first discovered this formula (and if not, who should be given the most credit for developing the formula) and (2) (if not) whether this name should continue to be used for a specific reason.

To answer the above questions, this study identifies the originator of each reliability coefficient. Few studies have raised this issue. Cronbach's (1951) and McDonald's (1999) studies have had a huge impact on how people call and use reliability coefficients. Past research that has had a great effect needs to be reviewed from various perspectives. However, few detailed studies have investigated the history of reliability coefficients, with the notable exceptions of Cronbach and Shavelson's (2004) own explanation of the history of alpha and Sijtsma's congratulatory comments (Heiser et al. 2016) on Cronbach's (1951) record number of citations. This study provides a comprehensive review of and a third-party perspective on the history of reliability coefficients, with a substantial component dedicated to a discussion of Cronbach's alpha, the most commonly used reliability coefficient.

This study is divided into two components. The first section argues that the current practice of recognizing the originator of alpha as Cronbach (1951) is incorrect. The second section explains the history of four other reliability coefficients, namely, the Spearman-Brown formula, Guttman's λ_4, standardized alpha, and McDonald's omega.

Ⅱ. Who First Developed Alpha?

Before proceeding with a discussion of alpha's history, it should be clarified that Cronbach and Shavelson (2004) himself declared that the expression Cronbach's alpha was inappropriate and stated that Kuder and Richardson (1937) had published a formula commonly called KR-20 and that alpha was "an easily calculated translation" (Cronbach & Shavelson: 397) of KR-20. Despite his rejection, Cronbach's alpha remains the most common name used to refer to this formula.

A reasonable explanation for this phenomenon is that while Cronbach's (1951) contribution to the alpha formula is well recognized, the contributions of studies that published the same formula before Cronbach (1951) are not well documented. Most textbooks describe Cronbach (1951) as the first to create the alpha formula. Cronbach (1951) and Cronbach and Shavelson (2004) vaguely explained previous studies other than that conducted by Kuder and Richardson (1937) to the extent that readers who are unfamiliar with the history of reliability coefficients might think that Cronbach (1951) was the first to publish a general formula of KR-20. For example, he noted the following: "So far as I recall, there was no one to offer the version that I offered in 1951, except for the Kuder-Richardson report, which did not give a general formula" (Cronbach & Shavelson 2004: 416). This study aims to help readers achieve a balanced view of alpha's history through a detailed review of pre-Cronbach (1951) studies.

1. Cronbach (1951) and Its Previous Studies

This study is not the first to argue that Cronbach (1951) did not first publish the alpha formula. McDonald (1999) states that Guttman (1945) published the formula for alpha before Cronbach (1951). Cho and Kim (2015) and Sijtsma (2009) assert that Hoyt (1941b) preceded both studies in discovering alpha. However, previous studies did not address alpha's history as an important topic and did not specify the commonalities and differences of the formulas proposed by Cronbach (1951), Guttman (1945), and Hoyt (1941b).

This study asserts that Cronbach (1951) is the sixth (not second or third) study to have discovered the general expression of KR-20. This study excavates three additional pre-Cronbach (1951) studies (Edgerton

& Thomson 1942; Gulliksen 1950; Jackson & Ferguson 1941) that contain the general expression of KR-20. In addition, this paper will explain the specific versions of the formula presented by both Kuder and Richardson (1937) and the papers that followed.

Kuder and Richardson (1937) developed various reliability formulas, each with different assumptions; however, they did not propose a special name for each reliability coefficient. They believed that the twentieth and twenty-first formulas would be the most useful. Subsequent studies referred to these formulas as Kuder-Richardson Formula 20 and 21, or KR-20 and 21 for short. Kuder and Richardson (1937) address conditions in which the test had dichotomously scored items (e.g., correct or incorrect). The test score X is the sum of the observed scores of the items (i.e., $X=\Sigma_{i=1}^{k}X_i$), σ_x^2 denotes the test score variance, p_i denotes the percentage of correct responses for item i, q_i denotes the percentage of incorrect responses for item i ($p_i+q_i=1$), $\overline{pq}$ denotes $\Sigma p_i q_i/k$, $\bar{p}$ denotes $\Sigma p_i/k$, and $\bar{q}$ denotes $\Sigma q_i/k$. The formulas for KR-20 and KR-21 are presented as follows.

$$\rho_{KR-20(\text{or}iginal)}=\frac{k}{k-1}\left(\frac{\sigma_X^2-k\overline{pq}}{\sigma_X^2}\right)=\frac{k}{k-1}\left(1-\frac{\Sigma_{i=1}^{k}p_iq_i}{\sigma_X^2}\right)\quad \text{and} \tag{1}$$

$$\rho_{KR-21}=\frac{k}{k-1}\left(\frac{\sigma_X^2-k\cdot\bar{p}\cdot\bar{q}}{\sigma_X^2}\right) \tag{2}$$

The general expression of KR-20 does not place limitations on the score of X_i. In the original expression, X_i may have a value of either 0 or 1. In the general expression, it may have all real number values (e.g., 2.47). Let σ_i^2 denote the variance of item i. The general expression of KR-20 is as follows.

$$\rho_{JF} = \rho_{ET} = \frac{k}{k-1}\left(\frac{\sigma_X^2 - \sum_{i=1}^{k}\sigma_i^2}{\sigma_X^2}\right) \text{ and } \lambda_3 = \rho_G = \alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k}\sigma_i^2}{\sigma_X^2}\right) \quad (3)$$

Hoyt (1941b) is the first to present the general formula, describing an idea to derive the KR-20 formula using analysis of variance (ANOVA), a method that generates exactly the same result as alpha. However, Hoyt (1941b) does not present Equation 3, instead explaining the entire process of the computation to explicate his method.

The second line of research to suggest the general expression is that of Jackson and Ferguson (1941). Because Hoyt (1941b) was included in the third issue of a quarterly academic journal, we assume it was published in July, August, or September of that year. Jackson and Ferguson (1941) was published in October. In contrast to Hoyt (1941b), Jackson and Ferguson (1941) clearly express Equation 3 (i.e., ρ_{JF}), making it the first paper to explicitly propose the current version of the alpha formula.

The third study that featured the general expression (i.e., ρ_{ET}) is Edgerton and Thomson (1942); however, it did not propose a new way of deriving KR-20 as the other studies introduced here.

Guttman (1945) is the fourth researcher to have published the general expression (i.e., λ_3). Based on the assumption that measurement errors are independent of each other, he deduces six reliability estimators, designating them $\lambda_1, \ldots, \lambda_6$. Guttman (1945) proves that these estimators are always equal to or smaller than the reliability, introducing the term *lower bounds* to describe this quality. He also offers mathematical proof that λ_2 is always a more accurate reliability estimator than λ_3 but notes that the calculation of λ_2 is more complex than that of λ_3; therefore, λ_3 can be used instead of λ_2 if the covariances are not significantly different (i.e., being tau-equivalent in modern terms).

The fifth study to have presented the general expression (i.e., ρ_G) is Gulliksen (1950), which proposes a new way to derive KR-20 based on "[t]he simplest and most direct assumption" (p. 223). In contemporary terms, his assumption is the same as the condition of being essentially tau-equivalent (Lord & Novick 1968).

Cronbach (1951), the sixth study to present the general expression (i.e., α), sparked the popular use of this reliability coefficient by eliminating concerns that made users hesitate to use it (Heiser et al., 2016). First, his proof of the relationship between alpha and split-half reliability has been highly responsive. Several reliability coefficients already existed at that time, but there was no clear conclusion as to which coefficient to use. Cronbach (1951) proved that alpha equals the average of split-half reliability (λ_4: Guttman 1945) values obtained from all possible split-halves. This proof is not significant given the study by Guttman (1945), which proved that alpha (i.e., λ_3) and λ_4 are not reliability coefficients in the strict sense but lower bounds of the reliability. However, the concept of lower bounds was not fully understood at the time (Heiser et al., 2016), and Cronbach's (1951) proof had the advantage of being intuitively easy to understand. This proof has recognized alpha as the representative reliability coefficient and not just one of several methods.

Second, Cronbach (1951) presented a comprehensive and encyclopedic (Cronbach & Shavelson 2004: 396) explanation for the interpretation and use of alpha. The length of this paper is 38 pages, making it not only the longest of all papers published in *Psychometrika* in 1951 but three times as long as the average paper. The most notable was Cronbach's (1951) assertion that a high value of alpha indicates the internal consistency or homogeneity of the data. In other words, alpha has been explained to be useful for informing of not only the reliability but also the unidimensionality of the data (Heiser et al. 2016; Sijtsma 2009).

Third, Cronbach (1951) adopted a different approach to alpha's prerequisites from previous studies. Pre-Cronbach (1951) studies focused on the mathematical proof of the assumptions of the alpha formula. However, because too-strict restrictions were needed to derive the formula, the concern that alpha's assumptions could not easily be met by real-world data was raised. For example, Cronbach (1943) criticized KR-20's assumption of unidimensionality as unrealistic, stating the following: "The basic assumption of the Kuder-Richardson method ... that the items measure only one general variable plus specific factors, is manifestly untrue for most achievement tests" (p.486). Cronbach (1951) took the opposite approach from the previous study. In fact, he focused his attention on its interpretation, assuming that the alpha formula had already been provided. Users were thus convinced that alpha could be used without regard to whether the data satisfied the assumptions of the alpha formula. What changed was his attitude toward the assumption of alpha, not the assumption itself.

Fourth, Cronbach (1951) suggested that the degree of alpha's underestimation was not worse than expected. Kuder and Richardson (1937) and Hoyt (1941a) regarded it as a major advantage of KR-20 that it does not overestimate the reliability. In contrast, Cronbach (1943) opposed the universal use of KR-20, criticizing it as producing "excessively conservative estimates of reliability" (p. 488) that are sometimes less than zero. In addition, Cronbach (1943) lamented that it was important to know the degree of underestimation of KR-20, but little information was available. Cronbach's (1951) proof that alpha is the mean of the split-half reliability values obtained from all split-halves seemingly gave clues to his own question. In other words, it was possible to conclude that alpha's tendency for underestimation is not very serious because alpha provides a value greater than approximately half of the split-half reliability estimates. Considering that the reference point of the comparison is the values of the split-half reliability coefficient, not other competitive alternatives such as λ_2, it is difficult to agree with this interpretation from a modern perspective.

2. KR-20 and Alpha were Considered Identical

Studies before Cronbach (1951) described the original and general expressions as the same formula. Hoyt (1941b) states, "It may be interesting to some who are familiar with the work of Kuder and Richardson that the foregoing method of estimating the coefficient of reliability gives precisely the same result as formula (20) of their paper. This fact can be easily verified algebraically" (p. 156). Jackson and Ferguson (1941) state that Equation 3 (i.e., ρ_{JF}) "is identical with the Kuder-Richardson formula (20)" (p.74). Guttman (1945) indicates that "λ_3 resembles a formula developed separately by Kuder and Richardson and Hoyt. In fact, [λ_3] is algebraically identical to this formula (which is formula (20) in Kuder and Richardson's paper)" (p.274-275). Gulliksen (1950) also emphasizes that the formula presented in his paper is "identical" (p. 224) to the formula proposed in Kuder and Richardson (1937), Jackson and Ferguson (1941), and Guttman (1945). None of the studies discussed here describe the two expressions as different formulas.

It is common practice among scholars to attempt to differentiate their research by emphasizing its difference from previous studies. However, Hoyt (1941a) uses the fact that he derived KR-20 based on a different approach (Hoyt 1941b) from Kuder and Richardson (1937) to compliment the authors: "The theoretical soundness of the Kuder-Richardson derivation is indicated by the fact that analysis of variance techniques applied to this problem produce an identical formula" (p. 93). He does not boast of trivial differences from the previous literature as a virtue.

3. Kuder and Richardson (1937) are Likely to have Chosen the Original Version Intentionally

Current textbooks appear to indicate that the general expression overcomes the important limitations of the original expression of KR-20. Readers who are accustomed to this interpretation may experience difficulty understanding why pre-Cronbach (1951) studies described the original and general expressions as being identical. Indeed, the two expressions are different in only a minor fashion. The concept that $\Sigma p_i q_i$ means $\Sigma \sigma_i^2$ is an easy relationship that is known to individuals familiar with basic statistics. From today's perspective, the general expression is more useful than the original expression because whereas the original formula may be applied to only dichotomously scored items (that is, 0 or 1), the general expression may be used for other general data. Furthermore, current users typically analyze data not measured as dichotomously scored items. Why, then, did Kuder and Richardson (1937) propose the original version?

There is a strong possibility that Kuder and Richardson (1937) deliberately chose the original expression. It was not that they could not derive the general expression: If one follows the logic with which they derived the original formula, one can easily understand that mere modifications will also easily derive the general expression. For example, the authors referred to $\Sigma p_i q_i$ as the sum of the variances of the items (p. 154) to explicitly describe the relationship of $\Sigma p_i q_i = \Sigma \sigma_i^2$. Kuder and Richardson (1937) likely proposed the original expression because, at the time of publication, that expression was more helpful to users than was the general expression. To understand this reasoning, we must understand the conditions of the past, which differed from current conditions.

First, the data processed by the formula users of the time were measured with dichotomously scored items. The "reliability of persons, over items, on a single trial" is typically referred to as test score reliability, which is derived from the finding that the pioneers of reliability research were primarily interested in students' test scores. Unlike today, scoring and calculating the results of a test once required many hours. To simplify the scoring process, school tests at the time were configured as true or false (Vehkalahti 2000). The International Business Machine Counting Sorter, which made scoring and calculation four to eight times faster than manual processing, began to be used in 1937 (Bedell 1940). The IBM Counting Sorter also classified answers as only true or false; thus, when Kuder and Richardson (1937) was published, there was little need to propose the general version in place of the original version.

Second, the ease of calculation was thought to the most important consideration. Today's widespread use of statistical software packages enables us to obtain reliability coefficient values without having to understand the formula; in the past, however, because reliability coefficients had to be calculated by paper and pencil by users, the ease of calculation was considered critical. Thus, the academic community (1) preferred the formula for which the calculation was easier if the resulting value did not substantially differ and (2) preferred the more easily calculated version between two algebraically equivalent formulas.

The importance placed on the ease of calculation in the first sense is indicated by the fact that Kuder and Richardson (1937) proposed both KR-20 and KR-21 together. Because KR-21 produces less precise reliability estimates than KR-20, it is mathematically inferior, and from the contemporary perspective, KR-21 would not be deemed sufficiently valuable to merit presentation. However, although the calculation of KR-21 is easier and simpler, in most cases the resulting values of the two formulas are not very different. KR-21 had high usability in an era when computer-based computations were practically impossible.

Kuder and Richardson (1937) likely proposed the original expression instead of the general expression because of the ease of calculation in the second sense. If the general expression was proposed, users who did not understand the relationship of $\Sigma p_i q_i = \Sigma \sigma_i^2$ would have had experienced difficulty in calculation. In a situation in which most users analyzed dichotomously scored items, there was no specific need for the authors to suggest the general expression.

In those days, there were no arguments that the general expression is more useful than or superior to the original expression. Although many subsequent studies discuss Kuder and Richardson (1937; Cronbach 1943, 1947; Hoyt 1941a; Kelley 1942; Tucker 1949; Wherry & Gaylord 1943), no authors have described the fact that the original formula may be applied to only dichotomously scored items as a limitation. Ferguson (1951) argued that the original formula can also be expanded to general situations through the following statement:

> Hitherto the Kuder-Richardson [formula 20] has been largely used to provide a descriptive index of the internal consistency of tests constructed of items which permit only two categories of response, a pass or a fail, to which the values 1 and 0 are assigned, respectively. The use of this formula may, however, be legitimately extended to provide indices of the internal consistency of responses on personality inventories, attitude scales, and other types of tests which permit more than two categories of response.(p. 614)

4. Evaluation of the Achievements of Cronbach (1951) and Kuder and Richardson (1937)

Although Cronbach's (1951) historical achievements should be respected, the fact that his interpretation of alpha literally affects the present is undesirable. Name affects our perception. The name Cronbach's alpha gives the misleading impression that Cronbach (1951) is the most authoritative source of this reliability coefficient rather than only one of the many studies on this reliability coefficient. Perception determines our behavior. Cronbach (1951) is still the most influential source of this reliability coefficient. According to Google Scholar, nearly 3,000 studies per year cite Cronbach (1951). Numerous textbooks still illustrate Cronbach's (1951) mathematical proof and terminology (e.g., internal consistency) to explain the usefulness of alpha. The public perception of alpha stands at the level of 1951, like a broken clock.

The pace of scientific progress is rapid. For example, the paper by Watson and Crick (1953), which first identified the structure of deoxyribonucleic acid, is a great achievement, but its content is only at a basic level from the standpoint of modern biology. Cronbach's (1951) arguments and approaches have been criticized as ineffective or proven to be inaccurate (Bentler 2009; Cho & Kim 2015; Cortina 1993; Green, Lissitz, & Mulaik 1977; Green & Yang, 2009; Hunt & Bentler 2015; McDonald 1981; Osburn 2000; Revelle & Zinbarg 2009; Sijtsma 2009, 2015; van der Ark, van der Palm, & Sijtsma 2011; Yang & Green 2011). Cronbach (1951) should be recognized as having historical value, but his claim should not be misinterpreted as valid until now. In other words, one should refer to the latest research on alpha, not Cronbach's (1951), to find an accurate description of alpha.

This study acknowledges the contribution of Cronbach's (1951) article. However, at least some of the studies that published the alpha formula earlier than Cronbach (1951) should be recognized for having greater contributions than Cronbach (1951). Among them, Kuder and Richardson's (1937) work is the most decisive achievement.

Kuder and Richardson (1937) resolved an important and difficult problem that had long been a tangle. During the period in which the study was published, the only approach used to estimate the reliability of a test score was to artificially split the items in half and apply the formula proposed by Brown (1910) and Spearman (1910). The method was problematic in that the manner in which the items were split produced varying values of reliability for the same data set; however, no one identified a better approach for more than two decades before Kuder and Richardson (1937). For example, Kelley (1924) describes this situation as follows:

> "I know of no better simple way of securing an estimate of reliability of a college entrance test than to split it into halves and use the Spearman-Brown formula and though there are hazards in doing this I certainly think that such an estimate is very much better than none at all" (p. 200).

Kuder and Richardson (1937) proposed an innovative technique that opened the new era for reliability coefficients.

Ⅲ. Who First Developed Other Reliability Coefficients?

1. The Spearman-Brown or Brown-Spearman Formula

The name Spearman-Brown does not indicate cooperation between the two scholars. Brown (1910) and Spearman (1910) simultaneously published algebraically equivalent formulas in the *British Journal of Psychology*. If these two individuals were alive today, they would have been sensitive to the issue of whose name comes before the other because they were not on amicable terms. Charles Spearman was hostile to Karl Pearson, a renowned statistician who taught at the same school, the University of London, and the two continued to publish articles that criticized and ridiculed each other (Cowles 2005). William Brown was Pearson's student. Brown's doctoral dissertation, which was later published as a book (Brown 1911), devoted most of the space to criticism of Spearman (1904). Decades ago, the name Brown-Spearman formula was used in some cases; however, in recent times, most studies refer to it as the Spearman-Brown (prophecy or prediction) formula. This study delves into the issue of which name is more valid.

It is difficult to rationalize why Spearman's name should appear before Brown. One seemingly fair explanation is that Spearman is a better-known scholar than Brown. Spearman left huge marks on the field of research methods by developing rank correlation and pioneering a statistical analysis technique known as factor analysis. In particular, Spearman (1904) developed a formulaic definition of reliability to open new doors to the history of reliability research. Cronbach, Rajaratnam, and Gleser (1963) described him as "the father of the classical reliability theory in psychology" (p. 138). Thinking about the study in question, however, without considering each scholar's prestige, Brown's name must precede that of Spearman.

First, Brown (1910) presented the version of the formula that is currently used. The two studies both developed a formula that may predict the reliability of a test that has the length of pa, when it is known that the reliability of a test with the length qa is $\rho_{XX'}$. Let k denote the ratio of p to q. Most textbooks express this formula in Brown's (1910) version (i.e., Equation 5) instead of the version of Spearman (1910; i.e., Equation 4). This formula is often used to calculate the split-half reliability; however, only Brown (1910) suggests applying a formula in case $k=2$(i.e., Equation 6). Let ρ_{12} denote the Pearson product-moment correlation between the split-halves:

$$\rho_{Spearman} = \frac{p\rho_{XX'}}{q+(p-q)\rho_{XX'}} \tag{4}$$

$$\rho_{Brown} = \frac{k\rho_{XX'}}{1+(k-1)\rho_{XX'}}, \quad \text{and} \tag{5}$$

$$\rho_{Split-half} = \frac{2\rho_{12}}{1+\rho_{12}} \tag{6}$$

Second, Brown's (1910) proof is superior to that of his competitor. Traub (1997) made an assessment that "Brown's proof of the formula is the more elegant" (p. 10). Compared with Spearman′s (1910) proof that includes two pages, Brown's proof (1910) is simpler and more intuitive.

Third, there is a high likelihood that Brown (1910) was written before Spearman (1910). Brown (1910) is a part of the author's doctoral dissertation, and when the paper was published, Brown had already obtained a doctoral degree from the University of London. Spearman (1910) criticized Brown (1910), which indicates that Spearman was well aware of the contents of Brown (1910). However, Brown (1910) criticized only Spearman (1904), not Spearman (1910). It is unlikely a coincidence that the two rivals who belonged to the same university published the same formula in the same journal at the same time; it is likely that Brown (1910) influenced Spearman (1910).

Finally, Brown comes before Spearman in alphabetical order. Determining whose research achievements are superior or whose proof is more elaborate may depend on subjective judgments, which makes it necessary to rely on objective principles for a delicate determination such as the current issue. According to the criteria set by the American Psychological Association, two or more researchers should be listed in alphabetical order. The Brown-Spearman formula is the name that meets this principle.

2. The Flanagan-Rulon Formula and Guttman's L4

The history of split-half reliability, which was presented after the Brown-Spearman formula, is also not well known. The assumption to use the Brown-Spearman formula as a split-half reliability coefficient is that the variances between each split half are equal. It has been explained that Flanagan (1937), Guttman (1945; λ_4), Rulon (1939), and Mosier (1941) independently developed reliability coefficients that may be used if the variances between the two halves are unequal (Cho 2016; Cronbach 1951; Raju & Guttman 1965). However, the manner in which Flanagan and Rulon contributed to the development of this formula has not been described in detail.

The manner in which this formula was first developed and publicized is unique: Rulon (1939) published the formula first developed by Flanagan. It is difficult to recognize Flanagan (1937) as the first researcher to present this formula because the study did not explicitly state the reliability formula or explain the calculation process prior to presenting several reliability estimates. Rulon (1939) is the first study that proposed the formula, which presented the two formulas, with an indication that the second formula is easier to calculate: Let σ_1, σ_2, σ_X^2, and σ_D^2 denote the variances of X_1, X_2, X_{1+2}, and $X_1 - X_2$, respectively. That is,

$$\rho_{Rulon1} = \frac{4\rho_{12}\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 + 2\rho_{12}\sigma_1\rho_2} = \frac{4\rho_{12}\sigma_1\sigma_2}{\sigma_X^2}, \quad \text{and} \tag{7}$$

$$\rho_{Rulon2} = 1 - \frac{\sigma_D^2}{\sigma_X^2} \tag{8}$$

However, Rulon (1939) specified that Flanagan personally explained both formulas to him. While writing the paper, Rulon briefly went on sabbatical to work with Flanagan. In sum, Flanagan published the formula he developed in a paper published by his colleague, not in his own paper.

Guttman's (1945) λ_4 is one of the six lower bounds the author proposed. Although the study suggested the utility of maximal λ_4 with the following statement, it was difficult to previously push the idea further given the lack of computer technology: "It is desirable, of course, to try to split the test in such a manner as to maximize [λ_4]" (p. 260).

$$\lambda_4 = 2\left(1 - \frac{\sigma_1^2 + \sigma_2^2}{\sigma_X^2}\right) \tag{9}$$

The split-half reliability formula is referred to differently depending on how it is used. The three formulas previously described are algebraically equivalent (Cho 2016). The name λ_4 is mainly used when Guttman's (1945) lower bound concept is used to obtain many split-half reliability values to choose from rather than calculating only one split-half reliability (e.g., Hunt & Bentler 2015; Osburn 2000). Formulas used for other objectives invoke the names Flanagan-Rulon or Rulon (e.g., Cortina 1993; Green 2003; Miller 1995). Thus, the same formula is referred to differently in different situations.

3. Standardized Alpha

The name standardized alpha results in a misconception in regards to the features of the coefficient. First, the name gives the impression that this reliability coefficient is a type of alpha. Previous studies have not strictly distinguished between alpha and standardized alpha in their use. For example, Cho and Kim (2015), who provide examples to explain the features of alpha, use the formula for standardized alpha instead of alpha.

Second, the name alpha induces users to prefer standardized alpha to alpha. Previous studies explain that there are two types of alpha. For example, Yu (2001) suggests that raw alpha and standardized alpha are two components of Cronbach's alpha. The word standardized has a more positive association than terms such as unstandardized or raw; thus, users without background knowledge may prefer standardized alpha to the other alpha.

If the name alpha was not included in the formula in question, this confusion and misunderstanding may not have occurred. Considering the characteristics of the formula, there is no reason to include the word alpha in the name of the coefficient. The relationship between standardized alpha and alpha is analogous to the relationship between the Brown-Spearman formula (i.e., Equation 6) and the Flanagan- Rulon formula (i.e., Equation 7 or 8). Thus, the two formulas are independent of each other. Historical evidence also suggests that there is no reason to include the Greek letter in the name. Cronbach (1951) did not use the term "standardized alpha" or recommend the use of the formula. The term "standardized alpha" is inappropriate for the characteristics and history of this reliability coefficient, whose records must be reviewed to understand this mislabeling.

Few previous studies delineate the history of standardized alpha, and studies that address the reliability coefficient (Falk & Savalei 2011, Hayashi & Kamata 2005) do not mention the origin of the formula. Unlike other reliability coefficients, standardized alpha does not have an uncontroversial developer in the records, resulting in a unique genesis.

SPSS (currently owned by IBM) contributed to the popularity of standardized alpha. SPSS was first developed for non-commercial use; however, it changed directions to the commercial world with the establishment of SPSS Inc. in 1975. A search on Google Scholar does not identify studies that used the term "standardized alpha" prior to 1975, which is when the number of papers that reported "standardized alpha" values increased. The common source cited by these papers is SPSS User's Guide (Specht 1975). SPSS not only named but also raised the level of utility of this formula, which previously had been little used.

This formula was rediscovered by SPSS; however, it would not be prudent to declare that it was first developed by a private company. The formula of standardized alpha has a similar form to the Brown-Spearman formula. If we assume that the reliability of the previous test ($\rho_{XX'}$) is the same as the average of the Pearson correlation coefficient ($\bar{\rho} = \Sigma_{i \neq j} \rho_{ij} / k(k-1)$) in Equation 5, the result will be the standardized alpha formula subsequently presented. The difference between the two lies not in the formula itself but in the interpretation of the formula. The Brown- Spearman formula has customarily been used to estimate the split-half reliability only when $k = 2$; when $k \geq 3$, it has not been used as an independent reliability coefficient. Considering the form of the formula, the first developers of standardized alpha are Brown (1910) and Spearman (1910). McDonald (1999) also refers to standardized alpha as the Spearman-Brown formula.

$$\alpha_{std} = \frac{k\bar{\rho}}{1+(k-1)\bar{\rho}} \tag{10}$$

4. Composite Reliability and McDonald's Omega

Before beginning the discussion, a congeneric measurement model (Jöreskog 1971) is explained. The test score X is the weighted sum of the observed score X_i, from item i $(i=1, \ldots, k)$ (i.e., $X=\Sigma_{i=1}^{k} w_i X_i$). X_i is separated into the sum of two uncorrelated unobserved components of the true score T_i and the error score e_i. Similar to Jöreskog (1971), this study assumes that there is no specific factor. A congeneric model has a true score configured as $T_i=\mu_i+\lambda_i F$, which, as such, is $X_i=\mu_i+\lambda_i F+e_i$. This study assumes that the errors among items are uncorrelated with each other (i.e., $Cov(e_i,e_j)=0 \quad \forall i \neq j$) and the variance of the latent variable F is 1.0 (i.e., $Var(F)=1$), whereas the expected value of e_i is 0 (i.e., $E(e_i)=0$). λ_i is referred to as the factor loading of item i.

The reliability coefficient based on a congeneric model was first presented by Jöreskog (1971). Along with Jöreskog's (1971) original version (ρ_J), this study presents a non-matrix version ($\tilde{\rho}_J$). Typical users use a unit-weighted sum (i.e., $X=\Sigma_{i=1}^{k} X_i$) and are unfamiliar with matrix algebra. The version that most textbooks feature ($\tilde{\rho}_{WLJ}$) was first proposed by Werts, Linn, & Jöreskog (1974). The two studies described in this paragraph do not specifically label the formula. To express gratitude for the scholar who first proposed the formula, it should be named the Jöreskog's formula or the Jöreskog-Werts formula; however, it is referred to entirely differently.

$$\rho_J = \frac{(\alpha'\beta)^2}{(\alpha'\beta)^2 + \alpha'\theta^2\alpha} \quad (\tilde{\rho}_J = \frac{(\sum_{i=1}^{k} w_i\lambda_i)^2}{(\sum_{i=1}^{k} w_i\lambda_i)^2 + \sum_{i-1}^{k} w_i^2\sigma_{e_i}^2)}) \tag{11}$$

$$\hat{\rho}_{WLJ} = \frac{(\sum_{i=1}^{k} \tilde{\lambda}_i)^2}{(\sum_{i=1}^{k} \hat{\lambda}_i)^2 + \sum_{i=1}^{k} \hat{\sigma}_{e_1}^2} \tag{12}$$

This coefficient answers to different names depending on the characteristics of the research. Substantive studies typically refer to it as the composite reliability, and methodological studies most commonly refer to it as the omega coefficient or McDonald's omega. Composite reliability is shorthand for the reliability of composite scores and is an inappropriate name for a specific reliability coefficient (Cho & Kim 2015). Because of these problems, an increasing number of studies use the name omega. This study provides a criticism of the utility and historical basis of the term omega.

A name's utility originates from increased precision and efficiency of communication; however, the term omega results in confusion. In literature on the subject of reliability, the omega coefficient refers to a wide variety of reliability coefficients. The omega of Heise and Bohrnstedt (1970) and McDonald's omega share common features; however, they are different formulas. McDonald (1978, 1985, 1999) referred to various unidimensional and multidimensional reliability coefficients based on an exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) as all omega. The use of the term omega coefficient without an explanation of the context will prevent the user from communicating the exact formula he or she is attempting to use.

To determine the historical basis for the omega coefficient, McDonald (1970, 1985) must be reviewed. McDonald (1970) included a reliability formula denoted as theta in the appendix of the paper. Its original version (θ) and non-matrix version ($\tilde{\theta}$) are as follows:

$$\theta = \frac{w' C_c w}{w' C w} \quad (\tilde{\theta} = \frac{(\Sigma_{i=1}^{k} w_i \lambda_i)^2}{\Sigma_{i=1}^{k} \Sigma_{j=1}^{k} w_i w_j Cov(X_i, X_j)}) \tag{13}$$

McDonald (1985) referred to the formula that is algebraically equivalent to Equation 12 as omega and declared that McDonald's (1970) theta will be renamed omega. When $\overline{\lambda} = \Sigma_{i=1}^{k} \lambda_i / k$, $\overline{e} = \Sigma_{i=1}^{k} \sigma_{e_1}^2 / k$, McDonald's (1985) formula is indicated as follows:

$$\omega = \frac{k(\overline{\lambda})^2}{k(\overline{\lambda})^2 + \overline{e}} \tag{14}$$

McDonald (1999) explicitly stated that his omega coefficient was first suggested in McDonald (1970). McDonald (1985, 1999) did not cite Jöreskog (1971) or Werts et al. (1974). He implied that the first study on this reliability coefficient is not Jöreskog (1971) but McDonald (1970), and this is the reason that this coefficient is referred to as McDonald's omega. The following sections contain a review of this assertion.

The formulas suggested by Jöreskog (1971) and McDonald (1970) appear similar; however, they mean different things, considering the context and periodic backgrounds in which the two formulas were presented. In this regard, three pieces of evidence are proposed.

First, McDonald (1970) proposed the formula in the context of EFA, not CFA. The title "the theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis" is telling of the characteristics of this paper. Bentler (1968) and Heise and Bohrnstedt (1970) also discussed reliability in terms of EFA. If McDonald's (1970) omega can be considered the general expression of Equation 11, other previous studies may be subject to the same line of reasoning.

Second, Jöreskog (1971) answered a more central question. The author explained how to produce reliability estimates (i.e., $\hat{\lambda}$, $\hat{\sigma}^2_{e_1}$) in contrast to McDonald (1970). Equation 13 appeared only in the appendix of McDonald (1970), without related comments in the body. If this formula was one that substantially stood out compared with previous achievements in the field, it would not have been presented in such a minor manner. While it was relatively less difficult to come up with a reliability coefficient, at the time, an important technical obstacle was the estimation of the parameters of the formula. In an attempt to resolve this problem, Jöreskog addressed the issue in multiple studies (e.g., Jöreskog 1969, 1970, 1971).

Third, the denominators of the formulas are different. In Jöreskog's (1971) formula, the denominator expresses fitted covariances. From a contemporary perspective, the denominator of McDonald's (1970) formula may be understood to be a general expression that may express both observed covariances and fitted covariances. However, the early 1970s was a time in which knowledge regarding parameter estimation of CFA was not sufficient. Heise and Bohrnstedt (1970), who expressed the denominator in a similar approach to that of McDonald (1970), interpreted it in terms of observed covariances. The denominator in McDonald's (1970) style must be understood as indicating observed covariances.

We add a comment to prevent misunderstandings about McDonald. The discussion so far on whose merit is greater is limited to the reliability coefficient based on a congeneric measurement model, or a unidimensional CFA model. McDonald (1999) pioneered reliability coefficients based on multidimensional CFA models, and his contribution and originality cannot be overemphasized. It is highly likely that he referred to the various reliability formulas as omega coefficients to help readers easily understand his book through consistent expression. His reader-friendly explanations were very effective, as can be observed from the high impact that his book has had on the field of psychometrics.

Ⅳ. Conclusion

The ideal name of a tool is informative and consistent. For example, iron clubs in golf are named from one to nine, with a difference in one number indicating a driving distance of ten yards. Under this system, remembering the driving distance of one club will enable the user to easily predict the driving distance of other irons. Iron clubs did not originally have this systematic naming system: until the 1920s, they had irregular names that did not indicate (at least not to individuals without background knowledge) each club's characteristics (e.g., Mashie-Niblick). As soon as the contemporary naming system was created, golf equipment companies did not hesitate to abandon the conventional system in their quest to attract new customers. If the industry has made success through name changes, could academia also benefit from change?

Reliability coefficients are also a type of tool. The goals of researchers who investigate tools do not stop short of developing good devices; instead, they extend far beyond to help users correctly utilize pre-existing tools. Users tend not to understand the mathematical formula that underlies the reliability coefficients; thus, our goal as researchers of the tool should be not only to help users understand the formulas but also to lead them to choose the correct reliability coefficient without a deep understanding of the formulas. The names of reliability coefficients should be considered not as a given constraint that cannot be changed but as a research topic that should be investigated.

We have looked at the history of reliability coefficients. The reason we examined history is to show that the current names are *pseudo-historical*. At first glance, it seems to be based on history, but it actually has a name that is against historical facts. We do not claim that the names of reliability coefficients should be *historical*. Knowing the history of each coefficient and the names of the originators does not help us to use the reliability coefficient. Our argument is that the names should be *ahistorical* (i.e., without concern for history). To keep the analogy of the golf club, it is not at all important to the user who originally invented the 7 iron. However, a naming system that gives information about when to pick a 7 iron is most helpful.

<Table 1> shows the systematic nomenclature proposed by Cho (2016). It answers the question of under what conditions the formula should be used and follows the consistent format of "(data feature)+reliability". For example, the prerequisite for alpha to equal the reliability is that the data are tau-equivalent, so the name tau-equivalent reliability was proposed. The use of ahistorical names will encourage users to correctly implement reliability coefficients. Most users use alpha automatically for all data sets regardless of assumptions such as tau-equivalency. Despite criticism from many previous studies (e.g., Green & Yang 2009), this old habit has barely changed. As long as we continue to use the name alpha, it is difficult to expect fundamental changes in practice. If we begin to use the term tau- equivalent reliability instead of alpha, users will be able to clearly understand which conditions necessitate this coefficient.

〈TABLE 1〉 Conventional and proposed names of reliability coefficients

Data		Split-half	General
Parallel	Conventional	Spearman-Brown formula	Standardized alpha
	Proposed	Split-half parallel reliability	Parallel reliability
Tau-equivalent	Conventional	Flanagan-Rulon formula Guttman's	Cronbach's alpha
	Proposed	Split-half tau-equivalent reliability	Tau-equivalent reliability
Congeneric	Conventional	Angoff-Feldt coefficient	Composite reliability McDonald's omega
	Proposed	Split-half congeneric reliability	Congeneric reliability

REFERENCES

Bedell, R. 1940. "Scoring Weighted Multiple Keyed Tests on the IBM Counting Sorter." *Psychometrika* 5: 195-201. doi:10.1007/BF02288565.

Bentler, P.M. 1968. "Alpha-maximized Factor Analysis (Alphamax): Its Relation to Alpha and Canonical Factor Analysis." *Psychometrika* 33: 335-345. doi:10.1007/BF02289328.

Bentler, P.M. 2009. "Alpha, Dimension-free, and Model-based Internal Consistency Reliability." *Psychometrika* 74: 137-143. doi:10.1007/s11336-008-9100-1.

Brown, W. 1910. "Some Experimental Results in the Correlation of Mental Abilities." *British Journal of Psychology* 3: 296-322.

Brown, W. 1911. *The Essentials of Mental Measurement*. London: Cambridge University Press.

Cho, E. 2016. "Making Reliability Reliable: A Systematic Approach to Reliability Coefficients." *Organizational Research Methods* 19: 651-682.

doi:10.1177/1094428116656239.

Cho, E. and S. Kim. 2015. Cronbach's Coefficient Alpha: Well Known but Poorly Understood. ***Organizational Research Methods*** 18: 207-230. doi:10.1177/1094428114555994.

Cortina, J.M. 1993. "What is Coefficient Alpha? An Examination of Theory and Applications." ***Journal of Applied Psychology*** 78: 98-104. doi:10.1037/0021-9010.78.1.98.

Cowles, M. 2005. ***Statistics in Psychology: An Historical Perspective***. New York: Psychology Press.

Cronbach, L.J. 1943. "On Estimates of Test Reliability." ***Journal of Educational Psychology*** 34: 485-494. doi:10.1037/h0058608.

Cronbach, L.J. 1947. "Test Reliability; Its Meaning and Determination." ***Psychometrika*** 12: 1-16. doi:10.1007/BF02289289.

Cronbach, L.J. 1951. "Coefficient Alpha and the Internal Structure of Tests." Psychometrika 16: 297-334. doi:10.1007/BF02310555.

Cronbach, L.J., N. Rajaratnam, and G.C. Gleser. 1963. "Theory of Generalizability: A Liberalization of Reliability Theory." ***British Journal of Statistical Psychology*** 16: 137-163. doi:10.1111/j.2044-8317.1963.tb00206.x.

Cronbach, L.J. and R.J. Shavelson. 2004. "My Current Thoughts on Coefficient Alpha and Successor Procedures." ***Educational and Psychological Measurement*** 64: 391-418. doi:10.1177/0013164404266386.

Edgerton, H.A. and K.F. Thomson. 1942." Test Scores Examined with the Lexis Ratio." ***Psychometrika*** 7: 281-288. doi:10.1007/BF02288629.

Falk, C.F. and V. Savalei. 2011. "The Relationship between Unstandardized and Standardized Alpha, True Reliability, and the Underlying Measurement Model." ***Journal of Personality Assessment*** 93: 445-453. doi:10.1080/00223891.2011.594129.

Ferguson, G.A. 1951. "A Note on the Kuder-Richardson Formula." ***Educational and Psychological Measurement*** 11: 612-615. doi:10.1177/001316445101100409.

Flanagan, J.C. 1937. "A Proposed Procedure for Increasing the Efficiency of

Objective Tests." *Journal of Educational Psychology* 28: 17-21. doi:10.1037/h0057430.

Green, S.B.(2003. "A Coefficient Alpha for Test-retest Data." *Psychological Methods* 8: 88-101. doi:10.1037/1082-989X.8.1.88.

Green, S.B., R.W. Lissitz, and S.A. Mulaik. 1977. "Limitations of Coefficient Alpha as an Index of Test Unidimensionality." *Educational and Psychological Measurement* 37: 827-838. doi:10.1177/001316447703700403.

Green, S.B. and Y. Yang. 2009. "Commentary on Coefficient Alpha: A Cautionary Tale." *Psychometrika* 74: 121-135. doi:10.1007/s11336-008-9098-4.

Gulliksen, H. 1950. *Theory of Mental Tests*. New York, NY: John Wiley & Sons.

Guttman, L. 1945. "A Basis for Analyzing Test-retest Reliability." *Psychometrika* 10: 255-282. doi:10.1007/BF02288892.

Hayashi, K. and A. Kamata. 2005. "A Note on the Estimator of the Alpha Coefficient for Standardized Variables under Normality." *Psychometrika* 70: 579-586. doi:10.1007/s11336-001-0888-1.

Heise, D.R. and G.W. Bohrnstedt. 1970. "Validity, Invalidity, and Reliability." *Sociological Methodology* 2: 104-129. doi:10.2307/270785.

Heiser, W., L. Hubert, H. Kiers, H.-F. Köhn, C. Lewis, J. Muelman, … and Y. Takane. 2016. "Commentaries on the Ten Most Highly Cited Psychometrika Articles from 1936 to the Present." *Psychometrika* 81: 1177-1211. doi:10.1007/s11336-016-9540-y.

Hoyt, C.J. 1941a. "Note on a Simplified Method of Computing Test Reliability." *Educational and Psychological Measurement* 1: 93-95.

Hoyt, C.J. 1941b. "Test Reliability Estimated by Analysis of Variance." *Psychometrika* 6: 153-160. doi:10.1007/BF02289270.

Hunt, T.D. and P.M. Bentler. 2015. "Quantile Lower Bounds to Reliability Based on Locally Optimal Splits." *Psychometrika* 80: 182-195. doi:10.1007/s11336-013-9393-6.

Jackson, R.W. and G.A. Ferguson. 1941. "Studies on the Reliability of Tests." *University of Toronto Department of Educational Research Bulletin* 12: 132.

Jöreskog, K.G. 1969. "A General Approach to Confirmatory Maximum Likelihood

Factor Analysis." *Psychometrika* 34: 183-202. doi:10.1007/BF02289343.

Jöreskog, K.G. 1970. "A General Method for Analysis of Covariance Structures." *Biometrika* 57: 239-251. doi:10.1093/biomet/57.2.239.

Jöreskog, K.G. 1971. "Statistical Analysis of Sets of Congeneric Tests." *Psychometrika* 36: 109-133. doi:10.1007/BF02291393.

Kelley, T.L. 1924. "Note on the Reliability of a Test: A Reply to Dr. Crum's Criticism." *Journal of Educational Psychology* 15: 193-204. doi:10.1037/h0072471.

Kelley, T.L. 1942. "The Reliability Coefficient." *Psychometrika* 7: 75-83. doi:10.1007/BF02288068.

Kuder, G.F. and M.W. Richardson. 1937. "The Theory of the Estimation of Test Reliability." *Psychometrika* 2: 151-160. doi:10.1007/BF02288391.

Lord, F. M. and M.R. Novick. 1968. "Statistical Theories of Mental Test Scores." *Reading,* MA: Addison-Wesley.

McDonald, R.P. 1970. "Theoretical Canonical Foundations of Principal Factor Analysis, Canonical Factor Analysis, and Alpha Factor Analysis." *British Journal of Mathematical and Statistical Psychology* 23: 1-21. doi:10.1111/j.2044-8317.1970.tb00432.x.

McDonald, R.P. 1978. "Generalizability in Factorable Domains: Domain Validity and Generalizability." *Educational and Psychological Measurement* 38: 75-79. doi:10.1177/001316447803800111.

McDonald, R.P. 1981. "The Dimensionality of Tests and Items." *British Journal of Mathematical and Statistical Psychology* 34: 100-117. doi:10.1111/j.2044-8317.1981.tb00621.x.

McDonald, R.P. 1985. *Factor Analysis and Related Methods*. Hillsdale, NJ: Lawrence Erlbaum.

McDonald, R.P. 1999. *Test Theory: A Unified Treatment*. Mahwah, NJ: Lawrence Erlbaum.

Miller, M.B. 1995. "Coefficient Alpha: A Basic Introduction from the Perspectives of Classical Test Theory and Structural Equation Modeling." *Structural Equation Modeling: A Multidisciplinary Journal* 2: 255-273.

doi:10.1080/10705519509540013.

Mosier, C.I. 1941. “A Short Cut in the Estimation of Split-halves Coefficients.” *Educational and Psychological Measurement* 1: 407-427. doi:10.1177/001316444100100133.

Osburn, H.G. 2000. “Coefficient Alpha and Related Internal Consistency Reliability Coefficients.” *Psychological Methods* 5: 343-355. doi:10.1037/1082-989X.5.3.343.

Raju, N.S. and I. Guttman. 1965. “A New Working Formula for the Split-half Reliability Model.” *Educational and Psychological Measurement* 25: 963-967. doi:10.1177/001316446502500402.

Revelle, W. and R.E. Zinbarg. 2009. “Coefficients Alpha, Beta, Omega, and the Glb: Comments on Sijtsma.” *Psychometrika* 74: 145-154. doi:10.1007/s11336-008-9102-z.

Rulon, P.J. 1939. “A Simplified Procedure for Determining the Reliability of a Test by Split-halves.” *Harvard Educational Review* 9: 99-103.

Sijtsma, K. 2009. “On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha.” *Psychometrika* 74: 107-120. doi:10.1007/s11336-008-9101-0.

Sijtsma, K. 2015. “Delimiting Coefficient α from Internal Consistency and Unidimensionality.” *Educational Measurement: Issues and Practice* 34: 10-13. doi:10.1111/emip.12099.

Spearman, C. 1904. “The Proof and Measurement of Association between Two Things.” *American Journal of Psychology* 15: 72-101.

Spearman, C. 1910. “Correlation Calculated from Faulty Data.” *British Journal of Psychology 1904-1920* 3: 271-295. doi:10.1111/j.2044-8295.1910.tb00206.x.

Specht, D.A. 1975. *SPSS: Statistical package for the social sciences, version 6: Users guide to subprogram reliability and repeated measures analysis of variance*. Ames, IA: Iowa State University.

Traub, R.E. 1997. “Classical Test Theory in Historical Perspective.” *Educational Measurement: Issues and Practice* 16: 8-14. doi:10.1111/j.1745-3992.1997.tb00603.x.

Tucker, L.R. 1949. "A Note on the Estimation of Test Reliability by the Kuder-Richardson formula (20)." ***Psychometrika*** 14: 117-119. doi:10.1007/BF02289147.

van der Ark, L.A., van der Palm, D.W., and K. Sijtsma. 2011. "A Latent Class Approach to Estimating Test-score Reliability." ***Applied Psychological Measurement*** 35: 380–392. doi:10.1177/0146621610392911.

Vehkalahti, K. 2000. ***Reliability of Measurement Scales: Tarkkonen's General Method Supersedes Cronbach's Alpha (Statistical Research Reports, Vol. 17)***. Helsinki, Finland: Finnish Statistical Society.

Watson, J.D. and F.H.C. Crick. 1953. "Molecular Structure of Nucleic Acids; A Structure for Deoxyribose Nucleic Acid." ***Nature*** 171: 737-738. doi:10.1038/171737a0.

Werts, C.E., R.L. Linn, and K.G. Jöreskog 1974. "Intraclass Reliability Estimates: Testing Structural Assumptions." ***Educational and Psychological Measurement*** 34: 25-33. doi:10.1177/001316447403400104.

Wherry, R.J. and R.H. Gaylord. 1943. "The Concept of Test and Item Reliability in Relation to Factor Pattern." ***Psychometrika*** 8: 247-264. doi:10.1007/BF02288707.

Yang, Y. and S.B. Green. 2011. "Coefficient Alpha: A Reliability Coefficient for the 21st Century?" ***Journal of Psychoeducational Assessment*** 29: 377-392. doi:10.1177/0734282911406668.

Yu, C.H. 2001. "An Introduction to Computing and Interpreting Cronbach Coefficient Alpha in SAS." In ***Proceedings of the Twenty-Sixth Annual SAS Users Group International Conference***, Paper 246. Cary, NC: SAS Institute, Inc.

<접수 2018/01/19, 수정 2018/04/16, 게재확정 2018/04/18>