Invited discussion of Cronbach (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334 (26,889 citations in Google Scholar as of 1/1/2016)

Sijtsma, Klaas

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334 (28,307 citations in Google Scholar as of 4/1/2016).

Klaas Sijtsma—Tilburg University, The Netherlands
E-mail: k.sijtsma@tilburguniversity.edu

By far the most cited article in *Psychometrika* is Lee Cronbach's famous study of coefficient alpha. I will address four questions: What makes the article special? Why do people cite the article more than other special articles? Was the article also influential beyond psychology? How does alpha fit in present-day psychometrics?

What makes the article special? In the first half of the 20th century, psychometricians distinguished two practically useful types of reliability, which were coefficients of stability and coefficients of equivalence. *Coefficients of stability* are correlations between two test scores obtained with the same test on two different time points, administered with a time interval that serves the psychologist's purpose. The coefficient provides information about the stability of the attribute across the time interval. *Coefficients of equivalence* are correlations between two test scores based on different sets of items intended to measure the same attribute, and express the degree to which the different sets are interchangeable. Cronbach mentioned two additional coefficient types. One type is the *coefficient of equivalence and stability* that refers to the correlation between test scores on different sets of items, one set administered first and the other set administered after a useful time interval. The other type is the *coefficient of precision*, referring to the correlation between the scores on the same test administered twice in one session to the same group of persons. The latter coefficient is hypothetical and not realizable in practice. Cronbach's article is about coefficients of equivalence.

Coefficients of equivalence were determined in two ways. First, one computed the correlation between equivalent forms of a test. However, constructing equivalent forms was impractical when one only intended to use one form. In clinical practice, many clinicians would prefer to avoid burdening their patients with twice the number of items necessary for diagnosing them and rather use one test form. In research, one would rather use the testing time to collect data on additional tests than spend precious time duplicating ones test battery. Hence, few researchers pursued the equivalent-form approach. Second, as a surrogate of equivalent forms,

one computed the correlation between the scores on two halves of the same test and then corrected for having obtained the coefficient of equivalence for only half the test rather than the whole test. This was the split-half method, only requiring one test form, and for reasons of efficiency, it became the dominant method for computing reliability. The method had two problems. First, several authors thought that the original method proposed independently by Spearman (1910) and Brown (1910) was based on unrealistically restrictive assumptions, and they proposed several, less restrictive, alternative methods. This raised the issue which method to use. Second, one could split a test of realistic length into a huge number of halves pairs, each pair producing a unique split-half reliability. This raised the issue which reliability to accept as final.

In the 1930s and 1940s, starting from different points of departure several authors suggested coefficients resembling what later became coefficient alpha (Kuder & Richardson, 1937), sometimes equating it (Guttman, 1945; Hoyt, 1941), but no one explored in detail the relationship of these coefficients to reliability obtained using equivalent forms and halves pairs. Cronbach (1951) did, hence producing his seminal 1951 contribution to psychometrics. He demonstrated two results, providing researchers with a tool he called coefficient alpha, that did not need equivalent test forms, or decisions about which split-half method to use and which halves pair to select. First, he demonstrated that coefficient alpha equals the mean of all possible split-half coefficients based on a method Guttman (1945) derived using realistic assumptions. Second, he argued that for most tests, even when the items in different test halves had different factor structures, the distribution of all possible split-half values shows modest spread, thus rendering the distribution's mean, which is coefficient alpha, an efficient summary of the distribution because it captures most split-half values with only little imprecision. Thus, in an era when computers were unavailable, Cronbach demonstrated that filling out a simple equation provides one with the mean of the distribution of split-half reliabilities, which is a quick and usually safe alternative to having to compute at least a sample of split-half values to check whether their spread really was small, justifying the use of coefficient alpha.

An auxiliary result of studying different factor structures' effect on split-half values and concluding the effect was small, was the conclusion that different factor structures, provided they were typical of real tests, had little effect on coefficient alpha. Moreover, Cronbach concluded that the test's general or dominant factor, in his words, the first-factor concentration in the test, was the main influence on alpha, and that group factors only affecting responses to subsets of

items in addition to the general factor's influence had a much smaller impact. Thus, Cronbach concluded that alpha quantifies the dominant factor among the items, which led him to relate coefficient alpha to what he called a test's internal consistency. Cronbach then suggested to quantify internal consistency by means of the mean inter-item correlation and related it to psychological interpretability. Relating alpha to internal consistency became the much-appreciated second main contribution the article made to psychometrics, test construction and test practice. Finally, Cronbach delimitated alpha from concepts such as Loevinger's homogeneity and Guttman's reproducibility.

I speculate that these latter contributions stood in the shadow of alpha being put in the zenith of reliability theory and relating alpha to a test's internal consistency. Taken together these major contributions and the minor contributions lend the article the appearance of an intellectual tour de force, replacing a hodge podge of concepts and methods, many of which badly understood, with a surprisingly simple method that already existed but whose meaning researchers had not realized until Cronbach clarified it to them. His contribution was so convincing that it stood the test of time until the present day.

Why do people cite the article more than other special articles? Cronbach (1951, p. 300) wrote that the essential problem set in his paper was: "How shall α be interpreted?" No doubt the article discussed this topic but its real contribution, one Cronbach probably did not foresee at the time of writing the article, was that, in a research area where many people were active, it provided researchers with a simple method that replaced chaos with order of the simplest form—a coefficient. A great contribution indeed. Only few publications make contributions of this magnitude, solving generations' psychometric problems. Rare examples are Bryk and Raudenbush (1992, second edition in 2002) in the context of multilevel modelling (27000+ citations) and Hu and Bentler (1999) in the context of structural equation modelling (32000+ citations). (Both results retrieved from Harzing's *Publish or Perish* that also consults Google Scholar.)

Not only the article about coefficient alpha has been cited frequently, but Cronbach's total oeuvre in which the 1951 article also is the front runner, has been cited 84,308 times (until March 9, 2016; retrieved from Harzing's *Publish or Perish*). His number 2 most-cited publication, the seminal article with Paul Meehl (Cronbach & Meehl, 1955) on construct validity, received 8,079 citations, a dazzling number that most of us can only dream of and

beating *Psychometrika*'s number 2 most-cited article (Kaiser, 1974; 6,043 citations as of 4/1/2016). Moreover, in total 14 of Cronbach's publications received 1000+ citations and another six received between 500 and 1000 citations. Cronbach truly had and continuous to have tremendous impact.

Was the article also influential beyond psychology? Table 1 is based on Web of Science (data retrieved on March 21, 2016, 7,058 citations in ISI journals), and shows one third of the citations came from psychology journals and two thirds from a large variety of other areas. This demonstrates that the 1951 article had a wide influence within and beyond psychology. In addition, the spread across many research areas renders it highly unlikely that only psychometricians cited the article, and the vast majority citing must be researchers reporting alpha for their test scores.

Table 1: Frequencies and Percentages of ISI Journal Articles Citing Cronbach (1951).

Research Area	Frequency	Percentage
Psychology	2432	34.5
Business Economics	871	12.3
Public Environmental Occupational Health	619	8.8
Psychiatry	589	8.3
Health care Sciences Services	528	7.5
Education Educational Research	365	5.2
Neurosciences Neurology	342	4.8
Social Sciences Other Topics	311	4.4
Sport Sciences	247	3.5
Computer Science	235	3.3

How does alpha fit in present-day psychometrics? In 1951, the mathematics of classical test theory was not yet fully developed and psychometricians found it difficult to agree about the relation between alpha and reliability. For example, Cronbach (1951, p. 299) wrote "It has generally been stated that α gives a lower bound to 'the true reliability'—whatever that means to that particular writer". Defining reliability as the correlation between two mathematically parallel tests, Novick and Lewis (1967) proved that coefficient alpha is a lower bound to the

reliability (also, see Lord & Novick, 1968). This definition comes close to what Cronbach called a coefficient of precision, expressing the sheer influence of random error on measurement. Later authors (Bollen, 1989; McDonald, 1998) introduced systematic error represented by group and specific factors, suggesting a factor-analysis approach to reliability. Cronbach, Gleser, Nanda, and Rajaratnam (1971) introduced generalizability theory, suggesting coefficients expressing reliability of person measurement free of unwanted influences such as caused by different test forms and raters. This development led Cronbach (2004) to abandon coefficient alpha in favor of generalizability, but probably due to its relative complexity the latter approach never became as popular as the former. Unlike Cronbach, however, many psychometricians, test constructors and researchers have remained faithful to coefficient alpha.

References

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3,* 296-322.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical Linear Models in Social and Behavioral Research: Applications and Data Analysis Methods (First Edition*). Newbury Park, CA: Sage Publications.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297–334.

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

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281-302.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* New York: Wiley.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika, 10,* 255–282.

Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika, 6,* 153–160.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39,* 31-36.

Kuder, G. F., & Richardson, M. W. (1937). The theory of estimation of test reliability. *Psychometrika, 2,* 151–160.

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

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

Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika, 32,* 1–13.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3,* 271-195.