# Clarity on Cronbach's Alpha Use

Jack Barbera
*Portland State University*, jbarbera@pdx.edu

Nicole Naibert
*Portland State University*

Regis Komperda
*San Diego State University*

Thomas C. Pentecost
*Grand Valley State University*



Citation Details

Barbera, J., Naibert, N., Komperda, R., & Pentecost, T. C. (2020). Clarity on Cronbach's Alpha Use. Journal of Chemical Education. DOI: 10.1021/acs.jchemed.0c00183

# Clarity on Cronbach's alpha use

[a]Jack Barbera, [a]Nicole Naibert, [b]Regis Komperda, and [c]Thomas C. Pentecost

[a]Department of Chemistry, Portland State University, Portland, Oregon
[b]Department of Chemistry and Biochemistry, San Diego State University, San Diego, California
[c]Department of Chemistry, Grand Valley State University, Allendale, Michigan

## Abstract

The Cronbach's alpha statistic is regularly reported in science education studies. However, recent reviews have noted that it is not well understood. Therefore, this commentary provides additional clarity regarding the language used when describing and interpreting alpha and other estimates of reliability.

**Reference Information:**
Jack Barbera, Nicole Naibert, Regis Komperda, and Thomas C. Pentecost, "Clarity on Cronbach's Alpha Use," *Journal of Chemical Education*, 2020, DOI: 10.1021/acs.jchemed.0c00183.

## Commentary

It has been previously established both in this *Journal*[1] and in other reviews of science education research[2] that the alpha statistic attributed to Cronbach[3] is not well understood, even though it is widely reported in studies. Part of this issue may be the vague and often imprecise language used to explain what information is communicated by an alpha value. Describing what alpha represents in a simple and straightforward way is not a trivial task. This is because there are a number of statistical assumptions and research decisions underlying its use that should be described when interpreting the value. However, in practice this is not often the case.[2] In our previous work on alpha, we related it to the laboratory measurements more commonly encountered by chemists, but even that analogy has severe limitations when applied to measurements of human subjects rather than chemicals.[4]

To briefly summarize, alpha is one of many methods of evaluating the reliability of data collected from an assessment instrument (also referred to as a test). Reliability in a broad sense refers to the consistency, or precision, of measurements.[5] There are multiple ways to conceptualize consistency, one is measurement *constancy over time*, described by a coefficient of stability such as test-retest reliability, which is the correlation between scores at two time points.[6] Another is the measurement *consistency between forms*, described as a coefficient of equivalence such as alternate-forms (or parallel-forms) reliability.[7] Like test-rest reliability this is simply calculated as a correlation between the two forms of the instrument.

However, the burden of developing and administering multiple highly similar forms of an instrument to calculate a coefficient of equivalence led to the development of *single-administration coefficients of equivalence* where the forms were not items on different instruments but rather a splitting of items from the same instrument. These commonly known single-administration coefficients include Spearman-Brown split-halves reliability, the Kuder Richardson (KR) 20 formula, KR 21, and alpha.[6] These coefficients were designed to evaluate the consistency among items on a single instrument for different types of splits and different types of items. The development of alpha was a way to avoid concerns about how the items were split when evaluating consistency. Instead, alpha mathematically represents all possible combinations of items (i.e., splits) for evaluating the consistency of the single test. As a result, alpha can be conceptualized as representing "the proportion of the test variance due to all common factors among the items".[6] Therefore, alpha values can be described as providing information about the proportion of a test score that represents the trait being measured, as long as all of the underlying assumptions for its use have been met.[4] It is this context that leads to the commonly used description of alpha as a measure of "internal consistency" between the items and the trait being measured, though there are numerous concerns with that term, and the reason that we promote the description of alpha as a measure of single-administration reliability.[4]

As an example of an unsupported use of the alpha statistic, we reflect on a recent study published in this *Journal*.[8] The selection of this study is not intended to negate the research conducted by the authors, but rather to allow for discussion of the use of alpha and provide an alternative given the goals of the research. The study utilized weekly quiz scores as an outcome variable but wanted to vary the content of the quizzes across class sections to avoid students sharing answers. Prior to presenting the results of the study on this outcome variable the authors note that the quizzes "were reliable within the Cronbach's alpha criterion of 0.7. This indicates that the quizzes were within an acceptable range of difficulty". This use of alpha appears to be an attempt to establish that the different forms of the quizzes were similar to each other.

While the alpha value for an individual quiz can be used to discuss the consistency of the items within it, comparison of the alpha values across *different versions* to demonstrate their equivalence is not a theoretically supported use of the statistic. Though alpha does not provide support for the similarity of test forms, the previously discussed coefficient of equivalence known as *alternate-forms* (also called parallel-forms) reliability does.[7] Additionally, alpha is an instrument or scale-level statistic using data from all items whereas 'difficulty' is an item-level statistic describing the proportion of examinees who answer an item correctly, simply calculated and reported as the percent of correct responses to an item.[9] There is no established relation between item difficulty and alpha.[3]

Lastly, the selection of 0.7 as an 'acceptable' criterion highlights the need for precision regarding research decisions. Despite popular belief and frequent reporting, there is no standard, threshold, or criterion value for an acceptable alpha.[2] Researchers must make a case for what information they believe alpha is providing and why a specific value is being deemed as acceptable or unacceptable. As noted by Taber, "Authors often cite alpha values with little commentary to explain why they feel this statistic is relevant and seldom interpret the result for readers beyond citing an arbitrary threshold for an acceptable value".[2]

Our intent in discussing this recent use of alpha is to provide an example of the possible issues related to the reporting and interpretation of alpha, as it is often poorly understood, yet regularly reported. Cronbach's alpha can be a useful estimate of reliability conceptualized as item consistency, under certain very restrictive conditions,[4] but does not provide any information on score equivalency or item difficulty. Our hope with this commentary, and our prior work on reliability, is to aid the community in better understanding the statistical background for reliability in general and of alpha in particular along with the limitations of its use and interpretation.

## References


(1) Raker, J. R.; Emenike, M. E.; Holme, T. A. Using Structural Equation Modeling To Understand Chemistry Faculty Familiarity of Assessment Terminology: Results from a National Survey. *J. Chem. Educ.* **2013**, *90* (8), 981–987. https://doi.org/10.1021/ed300636m.

(2) Taber, K. S. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Res. Sci. Educ.* **2018**, *48* (6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2.

(3) Cronbach, L. J.; Shavelson, R. J. My Current Thoughts on Coefficient Alpha and Successor Procedures. *Educ. Psychol. Meas.* **2004**, *64* (3), 391–418. https://doi.org/10.1177/0013164404266386.

(4) Komperda, R.; Pentecost, T. C.; Barbera, J., Moving Beyond Alpha: A Primer on Alternative Sources of Single-Administration Reliability Evidence for Quantitative Chemistry Education Research. *J. Chem. Educ.* **2018,** *95* (9), 1477-1491, 10.1021/acs.jchemed.8b00220.

(5) American Educational Research Association; American Psychological Association; National Council on Measurement in Education. (2014). *Standards for Educational & Psychological Testing*; American Educational Research Association: Washington, DC.
(6) Cronbach, L. J., Coefficient Alpha and the Internal Structure of Tests, *Psychometrika*. **1951**, 16, 297-334.
(7) Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric Theory*. New York, McGraw-Hill.
(8) Thompson, M. M.; Lamanna, A. C., Catalyzing Group Work in Introductory Chemistry: Evaluation of Five Strategies. *J. Chem. Educ.* **2020**, *97* (2), 351-357, 10.1021/acs.jchemed.9b00454.
(9) Crocker, L., & Algina, J. (1986). *Introduction to Modern and Classical Test Theory*. New York, NY: Holt, Rinehart, and Winston.