Current Science Research Bulletin
Volume 02 (02), February 2025, pp. 11-16
Available on: https://csrbjournal.org/index.php/csrb
e-ISSN: 3050-7103
p-ISSN: 3050-709X

Understanding Cronbach's Alpha in Social and Management Studies

Adetayo O. ADENIRAN (Ph.D., MInsTA)

University of Ilesa, Ilesa, Osun State, Nigeria; Department of Transport Planning and Logistics
https://orcid.org/0000-0002-6870-1212

ABSTRACT: In any research, the accuracy of assessments or evaluations can be improved when social and management science researchers work to develop easy-to-understand, accurate, and precise questions and questionnaires. Two essential components in assessing a measuring tool are validity and reliability. Surveys, simulations, and traditional knowledge, skill, or attitude assessments are examples of instruments. Concepts, psychomotor capability, and emotive values can be measured using instruments. The essence of validity is to identify the extent to which an instrument measures what it is supposed to measure. Reliability is the capacity of an instrument to measure consistently. Both reliability and validity of instruments are strongly connected. Without reliability, an instrument cannot be considered legitimate. However, the validity of an instrument is independent of its dependability, while the dependability of an instrument may be objectively assessed. This study identifies Cronbach's Alpha as the most employed objective reliability metric within the social and management discipline.

KEYWORDS: Coefficient Alpha, Cronbach's Alpha, Reliability
JEL classification codes: B40, C82

INTRODUCTION

Assessing a certain event requires frequent scientific study using already-existing instruments or creating new ones. Scales and tests are the two types of instruments considered in this study. Attitudes and other notions in the emotional domain are thought to be measured via scales (Adeniran, 2019). According to Howe *et al.* (2014), the tests are supposed to assess cognitive traits, including knowledge and comprehension of scientific concepts and themes. A researcher is expected to factor in the instrument's quality and applicability to specific research topics when selecting or creating a new instrument for a study (Adams & Wieman, 2010).

Validity (the degree to which an instrument measures what it claims to measure, rather than something else) and reliability (the degree to which an instrument can be expected to give the same measured outcome when measurements are repeated) are two traditional ways to conceptualize quality (Taber, 2013). Repeated measurements are taken to check for consistency in readings as part of the dependability test in the physical sciences. Although high reliability does not guarantee accuracy (for instance, an improperly calibrated ammeter may produce very consistent repeat readings without being accurate), it does offer a foundation for concluding changes (for instance, an increase in the reading on an improperly calibrated ammeter that has demonstrated the ability to provide repeatable readings can be assumed to indicate an increased current).

However, it might be challenging to discern between actual changes in measurement and reading changes when equipment produces unreliable results. Because human beings are always changing as a result of experiences between instrument administrations and the measurement process itself, it can be challenging to test the reliability of an instrument, such as an attitude scale or a knowledge test, by simply taking repeated readings in social science research. As a result, a person may respond to a series of questions, and that activity alone may result in fresh perspectives or more information integration. For no other reason than the impact of social, economic, and environmental factors on their encounter with the topic, the participant may respond differently to the same questions a day, week, or month later.

This study is a methodological critique that focuses on Cronbach's Alpha, a metric frequently linked to instrument reliability in scientific studies as evidenced in social and management research. Cronbach's Alpha reliability estimate is also referred to as coefficient Alpha. One of the reliability estimates that are most frequently cited in the literature is Cronbach's Alpha (Adeniran, 2019). The various methods for estimating reliability, the suitability of Cronbach's Alpha for estimating reliability, and the interpretation of Cronbach's Alpha were all investigated in this study to provide a sufficient explanation of Cronbach's Alpha.

Corresponding Author: Adetayo O. ADENIRAN

When multiple-item measures of a concept or construct are used in social and management researches, Alpha calculation has become standard procedure. It is simpler to use than other estimates (such as test reliability estimates) because it only requires one test administration (Adams & Wieman, 2010). Even though Alpha is frequently used in literature, its definition, appropriate use, and interpretation are not entirely apparent; therefore, it's critical to provide additional context for the underlying assumptions of Alpha to encourage its more efficient application. It should be noted that Cronbach's Alpha is the primary dependability indicator that this study is focusing on.

LITERATURE REVIEW

Strategies for estimating reliability

There are three methods for assessing dependability, according to Brown (1997):

a. Test-retest reliability, which is determined by administering a test twice and calculating the correlation between the two sets of scores;
b. Equivalent (or parallel) forms reliability, which is determined by administering two test forms and calculating the correlation between the two sets of scores; and
c. Internal consistency reliability is determined by using one of the numerous internal consistency equations that are available to determine a reliability estimate based on a single form of test administered once.

Since there is no need to conduct an interview twice or have two versions of the interview, the internal consistency technique is the most straightforward logically. There are many types of internal consistency reliability estimations. The major ones are:

i. Split-half adjusted (i.e., adjusted using the Spearman-Brown prophecy formula, which is the focus of Brown, 2001);
ii. Kuder-Richardson formulas (also known as $(K\text{-}R)_{20}$ and $(K\text{-}R)_{21}$ (Kuder & Richardson, 1937); and
iii. Cronbach's Alpha (Cronbach, 1970).

The $K\text{-}R_{20}$ and Cronbach's Alpha are the internal consistency estimates that are most commonly published. Both offer a reliable underestimation of a set of test findings' dependability, which is a safe or cautious estimate. But only when the test items are graded dichotomously (that is, right or wrong), can the $(K\text{-}R)_{20}$ be used. The advantage of Cronbach's Alpha over $(K\text{-}R)_{20}$ is that it can be used when test items are weighted. For example, an item scoring 0 points would have a functionally and grammatically incorrect answer, 1 point would have a functionally incorrect but grammatically correct answer, 2 points would have a functionally correct but grammatically incorrect answer, and 3 points for a functionally and linguistically accurate response). Hence, Cronbach's Alpha is more flexible than $K\text{-}R_{20}$ and is generally the ideal reliability estimate for language test creation projects and language testing studies.

What is Cronbach's Alpha?

To quantify the internal consistency of a test or scale, Lee Cronbach created Alpha in 1951. It is represented by a number between 0 and 1. The degree to which every item in a test measures the same idea or construct is known as internal consistency, and it is related to how related the test's components are to one another. To guarantee validity, a test's internal consistency should be assessed before its use in a study or examination (Adeniran, 2019; Tavakol & Dennick, 2011).

Furthermore, reliability estimates reflect a test's measurement inaccuracy. This notion of dependability may be summed up as the test's correlation with itself. The index of measurement error is calculated by squaring this correlation and deducting it from 1.00 (Casanoves et al., 2015). With a reliability of 0.75, for instance, the scores have a 0.44 error variance (random error) ($0.75 \times 0.75 = 0.56$; $1.00 - 0.56 = 0.44$). The percentage of a test score that can be attributed to a mistake will decrease as the reliability estimate rises (Mun *et al*., 2015). Notably, a test's dependability indicates how measurement error affects a cohort of students' reported scores rather than a single student. The standard error of measurement (SEM) must be computed to determine the impact of measurement error on a single student's observed score (Casanoves *et al*., 2015; Mun *et al*., 2015).

A test's Alpha score rises as its elements are connected. However, according to Mumba et al. (2015), a high coefficient Alpha does not always indicate a high level of internal consistency. This is because test length has an impact on Alpha as well. The Alpha value decreases if the test duration is too short. Therefore, the test should include more related questions that assess the same idea to boost Alpha. It's also critical to remember that Alpha is a characteristic of test results from a particular test sample (Griethuijsen *et al*., 2014). As a result, researchers should assess Alpha each time the test is given rather than depending on published Alpha estimates.

Use of Cronbach's Alpha

When Alpha is incorrectly used, it can result in instances where a test or scale is incorrectly rejected or when the test is condemned for producing unreliable findings. Understanding the related ideas of internal consistency, homogeneity, or unidimensionality can help to optimize the usage of Alpha and prevent this scenario. While homogeneity relates to unidimensionality, internal consistency focuses on how a sample of test items relates to one another. If a measure's items assess only one latent characteristic or concept, it is referred to be unidimensional.

Corresponding Author: Adetayo O. ADENIRAN

According to Tavakol et al., internal consistency is a required but insufficient criterion for determining if a sample of test items is homogeneous or unidimensional (Cronbach's Alpha).

Dependability is fundamentally based on the assumption that a sample of test items is unidimensional; if this assumption is not met, dependability is significantly underestimated. The fact that a multidimensional test does not always have a lower Alpha than a unidimensional test is widely established. Therefore, a more rigorous interpretation of Alpha holds that it cannot be used as a simple indicator of a test's internal consistency. A test's dimensions can be found via factor analysis. Therefore, Alpha may be used to verify if a sample of things is truly unidimensional 0.5 rather than just measuring the unidimensionality of a group of items.

However, if a test has several concepts or constructs, it might not be a good idea to state the test's Alpha overall because the higher number of questions will inevitably cause the Alpha score to be inflated. Therefore, in theory, Alpha should be determined for each idea rather than for the test or scale as a whole. It follows that Alpha should be determined for every instance in a summative test with diverse, case-based questions. More significantly, Alpha is based on the "tau equivalent model," which postulates that every test item uses the same scale to evaluate the same latent feature.

Therefore, this assumption is broken because Alpha understates the test's reliability if numerous factors or qualities underlying the scale's components, as shown by factor analysis. Reliability will be understated and the tau-equivalence assumption will be broken if there are too few test items. Alpha becomes closer to a more accurate dependability estimate when test items satisfy the tau-equivalent model's presumptions. Since diverse test items would go against the tau-equivalent model's presumptions, Cronbach's Alpha is a lower-bound estimate of reliability. A more thorough analysis of the comparable measurement in the data would be necessary if the "standardized item Alpha" calculation in SPSS is greater than Cronbach's Alpha.

Numerical values of Cronbach's Alpha

As previously mentioned, the Alpha value is influenced by the quantity of test items, item interrelatedness, and dimensionality. The allowable Alpha ranges from 0.70 to 0.95, according to various studies. A low value of Alpha might be attributed to a low number of questions, inadequate interrelatedness between items, or diverse constructs (Mumba et al., 2015). For instance, certain elements should be changed or eliminated if a low Alpha results from a poor correlation between them. Finding them is as simple as calculating each test item's correlation with the test's overall score; those with low correlations (around zero) are eliminated.

Given that they are evaluating the same things, an excessively high Alpha might indicate that some of the items are redundant. It has been suggested that the Alpha value should not exceed 0.90. According to studies by Adeniran (2019), Casanoves et al. (2015), Mun et al. (2015), Mumba et al. (2015), Griethuijsen et al. (2014), Howe et al. (2014), Taber (2013), and Adams & Wieman (2010), satisfactory internal consistency is indicated by Cronbach Alpha values of 0.7 or above. Schmitt (1996), however, proposed that instruments with relatively low Alpha values can still be helpful in some situations and that there is no universal threshold (like 0.70) at which Alpha becomes acceptable.

Interpretation of Cronbach's Alpha

Like other internal consistency estimates, Cronbach's Alpha estimate (typically represented by the lowercase Greek letter α) indicates the percentage of test score variance that can be ascribed to genuine score variance. To put it another way, Cronbach's Alpha is used to calculate the percentage of variation in a set of test results that is systematic or consistent. All values between 0.00 and 1.00 are likewise feasible, and it can vary from 0.00 (if no variation is consistent) to 1.00 (if all variance is consistent).

For example, if the Cronbach's Alpha for a set of scores turns out to be .90, it can be interpreted that the test is 90% reliable, and by extension that it is 10% unreliable (100% - 90% = 10%). However, when interpreting Cronbach's Alpha, one should keep in mind at least the following five concepts:

[1]. Cronbach's Alpha provides an estimate of the internal consistency of the test, thus
   - (a) Cronbach's Alpha does not indicate the stability or consistency of the test over time, which would be better estimated using the test-retest reliability strategy, and
   - (b) Cronbach's Alpha does not indicate the stability or consistency of the test across test forms, which would be better estimated using the equivalent forms reliability strategy.

[2]. Cronbach's Alpha is appropriately applied to norm-referenced tests and norm-referenced decisions (e.g., admissions and placement decisions), but not to criterion-referenced tests and criterion-referenced decisions (e.g., diagnostic and achievement decisions).

[3]. All other factors held constant, tests that have normally distributed scores are more likely to have high Cronbach's Alpha reliability estimates than tests with positively or negatively skewed distributions, and so Cronbach's Alpha must be

Corresponding Author: Adetayo O. ADENIRAN

interpreted in light of the particular distribution involved.

[4]. All other factors held constant, Cronbach's Alpha will be higher for longer tests than for shorter tests (Brown 1998 & 2001), and so Cronbach's Alpha must be interpreted in light of the particular test length involved.

[5]. When using test results to make actual judgments, the standard error of measurement, or SEM, is an extra reliability statistic that is derived from the reliability estimate (Brown, 1999) and might be more helpful than the reliability estimate itself. Because it estimates the amount of variability in actual test score points that can be anticipated around a specific cut-point due to unreliable variance, the SEM is useful (with a 68% probability if one SEM plus or minus is used, or with a 95% probability if two SEMs plus or minus are used, or with a 98% probability if three are used) (Brown 1996 or 1999).

To assess the accuracy of data provided in an examination or research project, high-quality tests are essential. One often-used measure of test reliability is Alpha. The dimensionality and length of the test have an impact on Alpha. The fundamentally tau-equivalent approach's presumptions should be followed by Alpha as a reliability indicator. If these presumptions are not met, a low Alpha is shown. Since test reliability depends on test length, Alpha does not only evaluate test homogeneity or unidimensionality. Regardless of whether a test is homogeneous or not, its dependability is increased by its length. A high Alpha score (> 0.90) might indicate that the test should be shorter and may indicate redundancy.

CONCLUSIONS

An essential idea in the assessment of tests and surveys is Cronbach's Alpha. Assessors and researchers must estimate this number to add validity and accuracy to their data interpretation. It is a practical and adaptable technique for examining the reliability of language test outcomes. In the process, it's critical to keep in mind that reliability regardless of the method employed to achieve it is an estimate of the consistency of a set of items when given to a specific group at a specific time under specific conditions for a specific purpose rather than a feature of the test itself.

The assumptions behind the calculation of Cronbach's Alpha, the factors influencing its magnitude, and the ways in which its value can be interpreted have been attempted to be explained in this study. However, Cronbach's Alpha has often been reported in an uncritical manner, without sufficient understanding and interpretation. Extrapolating reliability results obtained under a particular set of circumstances to other situations requires great care.

Author contributions: AOA: Conceptualization, Writing - original draft, Introduction, Discussions. The author read and approved the fnal manuscript.

Institutional Review Board Statement: Not applicable.

Funding: This research received no external funding.

Acknowledgement: The author appreciate the editor and reviewers for reviewing and publishing of this manuscript.

Informed Consent Statement: Not applicable.

Availability of data and materials: Not applicable

Confllicts of Interest: The author declares that there is no competing interest.

REFERENCES

1. Adams, W. K., & Wieman, C. E. (2010). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education*, 33(9), 1289–1312. doi:10.1080/09500693.2010.512369.
2. Adeniran, A. O. (2019). Application of Likert Scale's Type and Cronbach's Alpha Analysis in an Airport Perception Study. *Scholar Journal of Applied Science Research*, 2(4), 01-05. doi:10.3619/SJASR.1000223
3. Auewarakul, C., Downing, S., Praditsuwan, R., & Jaturatamrong, U. (2005). Item analysis to improve reliability for an internal medicine undergraduate OSCE. *Adv Health Sci Educ Theory Pract*., 10:105-13.
4. Bland, J., & Altman, D. (1997). Statistics notes: Cronbach's Alpha. BMJ., 314:275.
5. Brown, J. D. (1988). Understanding research in second language learning: A teacher's guide to statistics and research design. Cambridge: Cambridge University Press.
6. Brown, J. D. (1996). *Testing in Language Programs*. Upper Saddle River, NJ: Prentice Hall.
7. Brown, J. D. (1997). Statistics Corner: Questions and answers about language testing statistics: Reliability of surveys. Shiken: JALT Testing & Evaluation SIG Newsletter, 1 (2), 17-19. Available at: http://jalt.org/test/bro_2.htm
8. Brown, J. D. (1998). Statistics Corner: Questions and answers about language testing statistics: Reliability and cloze test length. Shiken: JALT Testing & Evaluation SIG Newsletter, 2 (2), 19-22. Available at: http://jalt.org/test/bro_3.htm
9. Brown, J. D. (1999). Statistics Corner: Questions and answers about language testing statistics: The standard error vs. the standard error of measurement. Shiken: JALT Testing & Evaluation SIG Newsletter, 3 (1), 15-19. Available at: http://jalt.org/test/bro_4.htm.

Corresponding Author: Adetayo O. ADENIRAN

10. Brown, J. D. (2001). Statistics Corner. Questions and answers about language testing statistics: Can we use the Spearman-Brown prophecy formula to defend low reliability? Shiken: JALT Testing & Evaluation SIG Newsletter, 4 (3), 7-9. Available at: http://jalt.org/test/bro_9.htm.
11. Casanoves, M., González, Á., Salvadó, Z., Haro, J., & Novo, M. (2015). Knowledge and attitudes towards biotechnology of elementary education preservice teachers: the first Spanish experience. *International Journal of Science Education*, 37(17), 2923–2941. doi:10.1080/09500693.2015.1116718.
12. Cohen, R., & Swerdlik, M. (2010). Psychological testing and assessment. Boston: McGraw-Hill Higher Education.
13. Cortina, J. (1993). What is coefficient Alpha: an examination of theory and applications. *Journal of Applied Psychology*, 78:98-104.
14. Cronbach, L. (1951). Coefficient Alpha and the internal structure of tests. Psychomerika, 16:297-334.
15. Cronbach, L. J. (1970). Essentials of psychological testing (3rd ed.). New York: Harper & Row.
16. DeVellis, R. (2003). Scale development: theory and applications: theory and application. Thousand Okas, CA: Sage.
17. Eberhard, L., Hassel, A., Bäumer, A., Becker, J., BeckMußotter, J., & Bömicke, W. (2011). Analysis of quality and feasibility of an objective structured clinical examination (OSCE) in preclinical dental education. Eur J Dent Educ. 2011; 15: 1-7.
18. Graham, J. (2006). Congeneric and (Essentially) Tau-Equivalent estimates of score reliability: what they are and how to use them. Educational Psychological Measurement, 66:930-44.
19. Green, S., Lissitz, R., & Mulaik, S. (1977). Limitations of coefficient Alpha as an index of test unidimensionality. Educational Psychological Measurement, 37:827-38.
20. Green, S., & Thompson, M. (2005). Structural equation modeling in clinical psychology research In: Roberts M, Ilardi S, editors. Handbook of research in clinical psychology. Oxford: Wiley-Blackwell.
21. Griethuijsen, R. A. L. F., Eijck, M. W., Haste, H., Brok, P. J., Skinner, N. C., & Mansour, N., (2014). Global patterns in students' views of science and interest in science. *Research in Science Education*, 45(4), 581–603. doi:10.1007/s11165-014-9438-6.
22. Howe, C., Ilie, S., Guardia, P., Hofmann, R., Mercer, N., & Riga, F. (2014). Principled improvement in science: forces and proportional relations in early secondary-school teaching. *International Journal of Science Education*, 37(1), 162–184. doi:10.1080/09500693.2014.975168.
23. Kline, P. (1994). An easy guide to factor analysis New York: Routledge; 1994.
24. Kuder, G. F., & Richardson, M. W. (1937). The theory of estimation of test reliability. Psychometrika. 2, 151-160. Available at: https:// jalt.org/test/PDF/Brown13.pdfnti
25. Iramaneerat, C., Yudkowsky, R., & Downing, S. (2008). Quality control of an OSCE using generalizability theory and many-faceted Rasch measurement. Adv Health Sci Educ Theory Pract., 13:479-93.
26. Jasper, F. (2010). Applied dimensionality and test structure assessment with the START-M mathematics test. The International Journal of Educational and Psychological Assessment, 6:104-25.
27. Lawson, D. (2006). Applying generalizability theory to high stakes objective structured clinical examinations in a naturalistic environment. J. Manipulative Physiol Ther., 29:463-7.
28. Miller, M. (1995). Coefficient Alpha: a basic introduction from the perspectives of classical test theory and structural equation modeling. Structural Equation Modeling, 2:255-73.
29. Mumba, F., Mbewe, S., & Chabalengula, V. M. (2015). Elementary school teachers' familiarity, conceptual knowledge, and interest in light. *International Journal of Science Education*, 37(2), 185–209. doi:10.1080/09500693.2014.971906.
30. Mun, J., Mun, K., & Kim, S.-W. (2015). Exploration of Korean students' scientific imagination using the scientific imagination inventory. *International Journal of Science Education*, 37(13), 2091-2112. doi:10.1080/09500693.2015.1067380.
31. Nunnally, J., & Bernstein, L. (1994). Psychometric theory. New York: McGraw-Hill Higher, INC.
32. Schmitt, N. (1996). Uses and abuses of coefficient Alpha. Psychological Assessment; 8:350-3.
33. Schoonheim-Klein, M., Muijtjens, A., Habets, L., Manogue, M., Van der Vleuten, C., & Hoogstraten, J. (2008). On the reliability of a dental OSCE, using SEM: effect of different days. *Eur. J. Dent. Educ.*, 2008; 12: 131-137.
34. Streiner, D. (2003). Starting at the beginning: an introduction to coefficient Alpha and internal consistency. Journal of Personality Assessment, 80:99-103.
35. Taber, K. S. (2013). *Classroom-based research and evidence-based practice: an introduction* (2nd ed.). London: Sage.

Corresponding Author: Adetayo O. ADENIRAN

36. Tate, R. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. Applied Psychological Measurement, 27:159-203.
37. Tavakol, M, Mohagheghi, M. A., & Dennick, R. (2008). Assessing the skills of surgical residents using simulation. J Surg Educ. 65(2):77-83.
38. Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's Alpha. International Journal of Medical Education, 2, 53-55. doi: 10.5116/ijme.4dfb.8dfd53

Corresponding Author: Adetayo O. ADENIRAN