Ethiopian Journal of Behavioral Studies, October 2024, 7(2), 218-225

Quality of Quantitative Data in Research

Tamirie Andualem [1]

DOI: https://doi.org/10.20372/ejobs.v7i2.10614
Received: 2 September 2023; Accepted: 13 October 2024

Since its establishment in 2018, the Ethiopian Journal of Behavioral Studies has published 45 empirical studies (why almost all of them empirical, not otherwise, will be discussed another time). These studies contributed securing national accreditation from the Federal Ministry of Education's task force that was established for third round journal evaluation in 2023. Of these manuscripts published since its inception in 2028, 25 of them employed quantitative data alone or together with qualitative data. These studies demonstrated data quality through using reliability coefficients. Quality of data is a necessary condition for a valid conclusion of a study. Quality of data in quantitative research is, simply and usually, shown using a quantitative index called reliability. While reliability of data is a necessary condition for validity, it is not sufficient. Validity of the data (validity of the data-also called validity of an assessment tool- is shown using reliability analysis, correlational analysis, and logical and practical analysis. It can be understood to be the overall quality of the data. As a single validity index is not used in research, single reliability index is presented commonly in quantitative empirical studies. This short communication aims to examine the common

[1] Associate Professor of Psychology, School of Psychology CEBS, AAU. Email tamirieand@yahoo.com

misunderstandings and solutions regarding reliability indices reported in EJOBS and other similar journals.

Data Collection Tools Requiring Reliability Indices

Data collection tools in quantitative research papers most commonly include questionnaire, inventory, scale, and test. It does not mean that data collected through these tools, even if standardized and validated in another country or place, qualify to meet the required reliability coefficient in the new context.

A tool with a set of questions resulting in frequencies may not necessarily produce a reliability coefficient. The first part of a tool, which is known to be demographic or socio-demographic questions, cannot have a reliability coefficient. Reliability coefficients are calculated for questions or items addressing the concept or construct under investigation. Therefore, reliability of a tool should be understood as reliability of data pertaining to those specific questions referring to the concept or construct.

Data collection tool requiring reliability coefficient is developed by other prior researcher/s, or it can be determined by the researcher for that particular research. As it is not easy to develop (including its validation) a quantitative research tool in a "theory testing" or "problem-solving" study, it is advisable to adapt or adopt a tool that was developed by others – either within or outside of a country.

It is advisable to present one or two items from each tool in the Instruments section of the Methods. Researchers should show these items in the way they would be answered by the respondents and scored by the researcher. This helps readers assess, at least, the logical or content validity of the tool.

Cronbach's Alpha and other Reliability Coefficients

Reliability of tools can be determined through different techniques depending on the uni-dimensionality of the construct: internal consistency of scores of items measuring the construct (internal consistency), the consistency of scores across times of administrations (test-retest), and the consistency of scores of raters or observers (inter-rater reliability).

Test-retest (also called stability measure), administered for a fairly long period of time between the two test conditions, is used when the construct is not assumed to be uni-dimensional. Besides, this reliability measure is necessary when the data collection tool is a single item measure.

Inter-rater reliability measure that considers two or more raters or observers or coders (though a single observer) is also possible when observing a phenomenon two or more than two times. An important condition to be met here is the equivalence of the competency of the experts or raters on the knowledge of the issue observed.

While internal consistency reliability has different types including Split-Half and Kuder-Richardson forms in the case of achievement tests scored as correct and incorrect, Cronbach's alpha is the popular one mainly because it is a default option in SPSS data analysis software (*SPSS: Analyse, Scale, Reliability Analysis*). Cronbach's alpha--where the item responses are scored as 0 or 1, or more-than-two scale values--is a generalization of Kuder-Richardson 20 (KR-20) -- where the item responses are scored as 0 or 1. As Cronbach's alpha is carried out on the assumption of uni-dimensionality of the construct of the study, this assumption should be checked using factor analysis: It can be taken as a generalization of internal consistency reliability measures (DeVellis, 2016). One has to use at least an exploratory factor analysis for

exploratory research methods, and at best a confirmatory factor analysis in the case of adopted or adapted tools, and as an extension of exploratory factor analysis when one develops own tool.

Level of Reliability Coefficient

Even if theoretically reliability coefficient is within 0 and 1, practically one can get negative reliability coefficient. Negative reliability coefficient results from items that are stated in opposite direction, but not reverse-scored when they are analyzed for the coefficient. Questions or items that are very different, or measuring different ideas, can have almost zero reliability coefficient. An acceptable reliability coefficient is 0.70 (Nunnally, 1978). This does not mean that coefficients below this number is a poor data or tool, nor is a coefficient more than .70 a guarantee for a good data or instrument.

An instrument that measures a single construct with relatively different items can have lower reliability than what one would expect. The number of items and the diversity of respondents also affect the reliability coefficient. Items with a larger number of questions can have a higher reliability coefficient than a test with lower number of items. Instruments with smaller number of items should also present the mean of the intercorrelation of the items (use, for instance, *SPSS: Analyse, Scale, Reliability Analysis, Statistics, Means*) besides the Cronbach's alpha coefficient, as Cronbach's alpha is usually less than .70 for a small number of items.

In addition,

> *Reliability is expected to be high even when the number of items is relatively small, provided the correlation among them is high. For example, a measure composed of 3 items whose average*

> *intercorrelation is .50 is expected to have an alpha of .75. The same alpha is expected for a measure composed of 9 items when the average intercorrelation among them is .25, and of 27 items when the average intercorrelation among them is .10 (Pedhazur & Schmelkin 1991, p.101).*

The characteristic of the group the instrument is administered to is also another factor in the degree or absence-or-presence, of reliability coefficient. A test administered to relatively homogeneous group on the trait has a low reliability coefficient. If a test is administered to a small number or large number of respondents with high level of similarity on the construct, the coefficient is low. For instance, very easy or very difficult exam questions can result in lower reliability coefficients.

Higher reliability coefficient should not be taken as a guarantee for a good test or data. We can get higher reliability coefficient for Cronbach's alpha by having many items or very similar (almost redundant) items. When the items of two unrelated tools (for instance, when one is self-report and the other maximal performance test like academic achievement) of distinct constructs with even low correlation are run for Cronbach's alpha, the resulting alpha can be more than the Cronbach's alpha of one or both or the average of the two measures.

The impact of lower reliability coefficients can be adjusted using coefficient of attenuation technique during correlational analysis. In this situation, the real correlation is higher than the observed correlation as reliabilities of the two variables' measures are not normally perfect, and the two measures of the two constructs are not similar.

Otherwise, method effects could inflate the correlation. Alternatively, structural equation modeling can be employed, controlling for measurement error through the measurement model of this advanced data analysis method.

Reliability Coefficient as a Variable

Reliability coefficient of a certain standardized or validated tool is understood, mistakenly by many, to be one and the same when administered to different groups of the population, or administered to different times or conditions. It, actually, varies from group to group, time to time, and condition to condition. Its variation should not be that high. A coefficient as much as .90 in the standardized tool is not expected to be less than .70 when the tool is used in another time and condition given the universality (similarity across countries) of the construct, careful adaptation/adoption and administration of the tool.

Reliability coefficient is not a property of the test but of the data. This means that even if a fixed reliability coefficient is reported during standardization or validation, one expects a different one when the tool is adapted or adopted. Let alone from one country to another, reliability should be reported for the final data of a research paper, besides the pilot reliability coefficient. Higher reliability coefficient in pilot studies is not a guarantee for the quality of the final data. The final data can be poor if the tool is administered carelessly by the test administrator, carelessly filled in by respondents, carelessly data-entered into a software, etc.

Reliability of a data for small sample size is not reliable, even if it is common to do it for test tryout or pilot testing. It should be of large sample size in which case testing it for statistical significance is not necessary (as testing of a correlation of .70

for statistical significance in sample size of, for instance, 100 or more is wastage of time: It is always statistically significant.

Reporting Reliability Coefficients in Research Papers

A reliability coefficient should be reported in the Method section of a paper. Even if every research paper may not use it, the following table can be taken as a comprehensive report of quality of the tool and its corresponding data.

Name of the construct or sub-construct	**Number of items**	**Reliability type**	**Reliability during standardization**	**Reliability of pilot study**	**Reliability of the final study**

If the groups of respondents (such as males or females) of the study are known to be different on the construct, reliability coefficients for each group should be separately presented. Even if they are not known to be different, it is safe to know if there is a difference in reliability coefficient across groups. A difference in reliability coefficient without the groups being different on the substance is an indication of "measurement error" (technically called absence of **measurement invariance** (Kline, 2011). One has to know that a difference between groups or conditions or times may not be a real difference, or a difference of content: It may be the result of measurement invaraince. As the concept and application of measurement invariance is not an easy task to be applied in journals like this one, presenting reliability coefficient for each group is a humble way of appreciating this complex process of achieving quality of a quantitative data.

As a conclusion, even if we do not have a single indicator for assessing the quality of a quantitative data in behavioral studies, it is a common practice to present reliability coefficient (mainly Cronbach's alpha) with not less than .70. As this coefficient or reliability in general is not sufficient for quality data, describing the tool clearly with at least one item, as an example, in the Methods section of the paper, and the detail procedures of administration and scoring of the instrument is important to give readers some sense of the quality of the data.

References

DeVellis, R. F. (2016). *Scale development: Theory and applications* (2nd ed.). Newbury Park: Sage.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed). New York: The Guildford Press.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill

Pedhazur, E. J. & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.