

THE INTERNATIONAL JOURNAL OF HUMANITIES & SOCIAL STUDIES

Guidelines for Validating Research and Evaluation Instruments

Dr. Chikwe, Christian K.

Senior Lecturer, Department of Educational Psychology, Guidance and Counselling
Ignatius Ajuru University of Education, Rumuolumeni, Port Harcourt, Rivers State

Dr. Inko-Tariah, D. C.

Senior Lecturer, Department of Educational Psychology, Guidance and Counselling
Ignatius Ajuru University of Education, Rumuolumeni, Port Harcourt, Rivers State

Abstract:

The paper examined guidelines for validating research and evaluation instruments. Validity as the extent to which research and evaluation instruments measure what they were designed to measure, is a very important quality of any measuring instrument. The concepts of validity, research and evaluation were explained in the paper. The guidelines for validating an instrument were explored, and the paper identified and proposed the following: identification and definition of the construct derived from psychological theories, prior research, systematic observation and analysis of relevant behaviour domain; preparation of the items of the instrument to fit the construct definition; giving draft copies of the instrument items to experts in the area of the construct identified; giving the draft copies of the instrument to other experts different from those experts used earlier to determine the suitability and relevance of the items; selection of the most relevant items from the initial item pools using the mean statistic; correlation of the scores of the instrument with the scores of already existing standardized instrument on the same construct; and other appropriate internal analysis such as factor analysis of item cluster or subsets. The paper concludes that it is necessary that research and evaluation instruments be validated using any of the steps outlined.

1. Introduction

Validity is one of the most important qualities of a good measuring instrument, and its importance cannot be overemphasized. Validity determines the appropriateness, quality, and usefulness of research and evaluation instruments. The degree to which research and evaluation studies are utilized in improving the educational system and other areas of human endeavour depends on the validity of the instruments employed in carrying out such studies. Hence, the process of developing and validating instruments is of interest to researchers and evaluators. This process of validating an instrument is called validation. This paper, therefore, examines guidelines for validating research and evaluation instruments in education under the following subheadings: concepts of validity and validation; research and evaluation instruments, guidelines for validating research and evaluation instruments; and conclusion.

2. The Concepts of Validity and Validation

Research, evaluation studies, and literature are replete with the definition and meaning of validity and validation. Asuru (2014) defined validity as the degree to which a measuring instrument measures what it purports to measure. This implies that research and evaluation instruments are valid to the extent that they measure the attributes and variables they were designed to measure. An instrument designed to measure test anxiety is valid to the extent that it measures test anxiety and nothing else. LongJohn (2014) stated that validity of a test (instrument) is simply its relevance and the extent to which an instrument measures what is needed for a particular purpose and to which the results, as they are interpreted and used, meaningfully and thoroughly represent the specified knowledge and skill.

Also, Asuru (2014) noted that validity is a matter of degree (extent) such as high, moderate or low with reference to a specified purpose or use for which the test (instrument) is being considered. The index of measuring validity is the validity coefficient which ranges from zero to unity (0-1). Hence validity coefficient could be zero, positive or negative.

There are three types of validity. They are content validity, criterion-related validity, and construct validity. Although, there is face validity which is strictly speaking not a type of validity. Content validity is the extent to which an instrument covers a representative sample of the subject matter (topics) and the behavioural objectives. The content validity of an instrument is determined by the expert judgment of specialists in that subject area. Content validity is made more objective by preparing and using a table of specification or test blueprint.

Criterion-related validity is the extent to which an instrument can predict the behaviour of the respondent. Criterion-related validity determines how effective the scores of an instrument are in estimating the present status or predicting future performance in a specified situation (Asuru, 2014). There are two types of criterion-related validity (concurrent and predictive). Concurrent validity is

the extent to which an instrument agrees with a concurrent measure of a different type. In concurrent validity, an instrument is administered once and another already existing instrument that has similar characteristics with the new one under construction in terms of content and psychometric properties is also administered, and the scores obtained from the two instruments are correlated. If the correlation coefficient is reasonable, then the new instrument is said to be valid (LongJohn, 2014). In concurrent validity, the predictive score is correlated with a criterion at the same time.

Predictive validity is the extent to which scores of an instrument predict future performance or behavior. The predictive score is correlated with the criterion score at a later date. In other words, predictive validity determines the extent to which instrument scores can predict ability in school, job, etc. at a later period on different assessment ability. Construct validity is the extent to which an instrument measures a psychological trait or construct. According to Asuru (2014: p. 79), "a construct in itself is a psychological quality which is assumed to exist in order to explain some aspects of behavior". Asuru further stated that in construct validity, the evaluator is concerned with analyzing the meaning of test (instrument) scores in terms of psychological constructs so as to rule out other unsatisfactory meanings or interpret results in terms of their psychological qualities. Nworgu as cited in Asuru (2014) identified two subclasses of construct validity. These are discriminate validity which ensures discrimination between traits that are dissimilar; and convergence validity which shows similarity between similar traits.

Chan (2014) explained that validity is the quality of inferences, claims or decisions drawn from the scores of an instrument. Validity, therefore, is concerned with inferences, claims or decisions based on instrument scores, not the instrument itself. American Psychological Association (1954) classified validity into four: content validity, predictive validity, concurrent validity and construct validity. They explained that construct validity is the focus or central component in validation and it is made up of five sources of evidence germane to the validation of the interpretation and use of the scores of an instrument. The five sources are evidence based on content; evidence based on response processes; evidence based on the internal structure; evidence based on relation to other variables, and consequences.

The content validity of an item is made up of the items, format, and wording of the items, response options, and the administration and scoring procedures. Evidence based on content can be determined by finding the relationship between the content of an instrument and the construct to be measured. Evidence based on response process is examining the cognitive or thinking process when responding to items. Methods such as think aloud protocols can be used to find out how items relate to each other by the use of a statistical method such as factor analysis and item response modeling. Evidence based on relations to other variables is the association between instrument scores and external variables. Consequences are the intended and unintended use of an instrument and how its unintended use weakens score inferences (Chan, 2014).

Validation is the process of gathering and evaluating the evidence to support the appropriateness, usefulness, and meaningfulness of the decisions and inferences made from instrument scores (Chan, 2014). It is an ongoing process which is used to accumulate and synthesize validity evidence to support the inferences, interpretations, claims, actions or decisions made. Anastasi (1986) explained that validation is the process of building a valid test. The validation of an instrument involves multiple procedures which are used systematically at various stages of validation (Jackson 1970, 1973 & Guion 1983 in Anastasi, 1986).

3. Research and Evaluation Instruments

Research is the systematic process of searching out something that can help in the solution of problems. Anyawu (2000) explained that research can be either simple or complex depending on the researcher's capability in terms of educational background, the purpose of the research, its significance, experiences and the funds available for the research. The important role research plays in education and other areas of human endeavour cannot be overemphasized. Hence, Anyawu (2001, p.4) has highlighted the roles of research as follows:

- (i) Research provides information to enable decision makers to solve identified problems;
- (ii) It provides insight into unknown processes;
- (iii) It provides training to participants in problem-solving and leadership;
- (iv) It assembles pool of data for planning and development; and
- (v) It encourages the application of scientific methods in social science and education, and other fields of study.

Adeleke (2010) defined research as the search for knowledge or any systematic investigation to establish facts. It is searching systematically for solutions to problems and helping to evaluate the findings of others. The purpose of research includes; to satisfy curiosity, to form the foundation of programme development and policies all over the world, to enhance how best to address the world's problems to have findings that should have implications for policy and project implementation (Adeleke, 2010).

Evaluation according to Nkwocha (2004 p. 4) is "the systematic placing of worth or value to attributes of an individual, object or event using specified standards". Evaluation is the systematic process of passing value judgment as to the worth of a course of study or programme. Evaluation can be formative or summative. Formative evaluation is evaluation carried out during or when the programme is in progress while summative evaluation is evaluation carried out at the end of a programme. Therefore, evaluation gives meaning to the result of research.

Research and evaluation are important in education and other areas of human endeavour. Instruments used in carrying out research and evaluation are expected to measure the characteristics and purpose they are designed to measure to give meaning to research and evaluation studies. Various instruments are employed in carrying out research and evaluation studies. These instruments include tests, interviews, questionnaire, checklist, rating scales, projective techniques, observation, sociogram, etc. These instruments have to be

validated before they can be used in research and evaluation studies. Therefore, research and evaluation instruments are the tools, used in gathering data and analysis of such data in research and evaluation studies.

4. Validation of Research and Evaluation Instruments

Nwankwo (2010) explained that in determining the content validity of an instrument, the first phase is enhancing the content coverage of the instrument. This is achieved by giving the instrument to few experts in the discipline of study for input. The instrument is accompanied by instructions to the experts on what to do. The experts are requested to improve the content coverage of the instrument by adding many more appropriate items as possible. The inputs, new ideas of the experts are integrated into the instrument.

The second phase is identifying the relevant items for the instruments. At the second stage, copies of the instrument are given to other experts different from those experts used to improve on the content coverage in the first phase. These experts are requested to indicate the suitability or relevance of the items by rating each of them on a scale, such as not relevant (1 point), somewhat relevant (2 points), relevant (3 points), and very relevant (4 points). The responses of the experts are analyzed using the mean. After the analysis, items rated accepted as relevant are included in the instrument, while items rated not relevant are discarded. At the end of the second phase, the resulting instrument could be said to have content validity.

Nwankwo (2010) further explained, that construct validity of an instrument is achieved by: correlating the scores of the instrument with those from already existing one, or standardized instrument on the same construct to be developed or validated; and determining the internal consistency of the instrument (a measure of homogeneity of items of an instrument) using the split-half and Cronbach alpha methods.

According to Anastasi (1986), the validation process begins with the following:

- Formulation of detailed trait or construct definitions derived from psychological theory, prior research, or systematic observation and analysis of the relevant behavior domain;
- Preparation of items of the instrument to fit the construct definitions;
- Empirical item analysis with the selection of the most effective (valid) items from the initial item pools;
- Other appropriate internal analysis may be carried out including factor analysis of item clusters or subsets; and
- Validation and cross-validation of various scores and interpretation scores through statistical analysis against external, real-life criteria.

The (AERA, APA & NCME, 1999, p. 37) outlined four phases in developing or validating an instrument in the standard for psychological and educational testing as follows:

- Phase 1: Delineation of the purpose of the test (instrument), scope of construct or extent of the domain to be measured;
- Phase 2: Development and evaluation of the test specifications;
- Phase 3: Development, field testing, evaluation and selection of the items and scoring guides and procedures; and
- Phase 4: Assembly and evaluating the test for operational use.

Evergreen, Gullickson, Mann and Welch (2011) outlined four instrument validation steps as follows:

- Clearly defining the object of measurement. Operational definition of abstract concepts in concrete terms;
- Definition of the relevant elements of this object to determine its "domain content", review of literature, consulting stakeholders and gathering of preliminary qualitative data through focus groups interviews or surveys;
- develop measures (e.g. instrument question) for each element; and
- Review and pilot test the instrument

In validating research and evaluation instrument, therefore, the following steps could be adopted or adapted:

1. Identify and define the construct derived from psychological theories, prior research, systematic observation, and analysis of relevant behaviour domain;
2. Prepare the items of the instrument to fit the construct definitions.
3. Give draft copies of the instrument items to experts in the area of identified construct (psychologists, measurement and evaluation specialists, psychometricians, subject experts, etc.). This is to enhance and improve the content coverage of the instrument.
4. Give the draft copies of the instrument to other experts different from those experts used to improve the content coverage in step 3. This is to determine the suitability and relevance of the items using a 4-point scale of very relevant, relevant, somewhat relevant and not relevant.
5. Select the most relevant or effective items from the initial item pools using the mean statistics.
6. Correlate the scores of the instrument with the scores of already existing standardized instruments on the same construct, and determine the internal consistency of the items. Split-half and Cronbach alpha methods can be used.
7. Other appropriate internal analysis such as factor analysis of item clusters or subsets can be carried out.

5. Conclusion

The validation of research and evaluation instrument is a process whereby instruments used for research and evaluation studies in education and other disciplines are validated. The guidelines for validated research and evaluation instruments were surveyed.

It is, therefore, necessary that research and evaluation instruments be validated using any of the steps or guidelines highlighted in this paper. This will go a long way in producing research and studies that are valid and useful for applicability. Researchers can adopt or adapt any of the guidelines or steps outlined in the paper in validity research and evaluation instruments.

6. References

- i. Adeleke, J.O. (2010). The basics of research and evaluation tools. Lagos: Somerest ventures.
- ii. American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Evaluation (NCME) (1999). Standards for Educational and Psychological Testing. Washington, DC. American Educational Research Association.
- iii. American Psychological Association (1954). Technical recommendations for psychological test and diagnostic techniques. *Psychology Bulletin*, 51, 201-238.
- iv. Anastagi, A. (1986). Evolving concepts of test validation. *Annual Reviews Inc.* 37, pp 1-15.
- v. Anyanwu, A. (2000). *Research methodology in business and social science*. Owerri: Canum publishers Nig. Ltd.
- vi. Asuru, V.A. (2014). *Measurement and evaluation in education and psychology*. Port Harcourt: Pearl Publishers International Ltd.
- vii. Evergreen, S., Gullickson, C. M. & Wayne, W. (2011). Developing and validating instruments. Retrieved from evalu-ate.org/
- viii. LongJohn, I.T. (2014). Qualities of a good test. In M.O.N Obagah and D.C. Inko-Tariah (Eds). *Educational measurement and evaluation*. Port Harcourt: Rodi printing and publishing company.
- ix. Nwankwo, O.C. (2010). *A practical guide to research writing*. Port Harcourt: Golden publishers.
- x. Nwaocha, P.C. (2004). *Measurement and evaluation in the field of education*. Owerri: Versatile publishers.
- xi. Zumbo, B.D., E.K.H Chan (2014). *Validity and validation in social behavioural, and health sciences, Social indicators research series 54*. Switzerland: Springer International Publishing.