International Journal of Multidisciplinary Research and Growth Evaluation.

Validity of an instrument in educational measurements

Faith Bokolo [1*], Otuoku Karl [2]
[1-2] Department of Psychology, Guidance and Counselling, Isaac Jasper Boro College 0f Education, Sagbama, Bayelsa State, Nigeria

* Corresponding Author: **Faith Bokolo**

Article Info

ISSN (Online): 2582-7138
Impact Factor (RSIF): 8.04
Volume: 07
Issue: 02
Received: 14-01-2026
Accepted: 12-02-2026
Published: 03-03-2026
Page No: 62-64

Keywords: Validity, Instrument, Criterion-Related Validity, Educational Measurement

Abstract
A valid instrument is needed to evaluate learning in our schools. This paper throw light on researcher's interest in developing instrument in their field of studies. The need to implore experts in their work after developing tests items are explained vividly. Types and factors affecting validities which are important ingredients and like's problems are all examined. It is my conviction that researches will find this text useful in their quest for understanding validity and the administration of instruments.

Introduction

Validity are important factor when developing and testing any instrument for use in a study. Instrument is the general term that researchers use for a measurement device (e.g. survey, test, questionnaire, etc) Attention to these considerations helps to insure the quality of your measurement and the data collected for your study. Ikeotuonye, A.I. 2018).

Validity is refer to the degree to which an instrument accurately measures what it is intends to measure. An instrument that is a valid measure of a basic mathematics skills in a school may be a valid predictor of how well students might do in school, once they complete school. So we never say that an instrument is valid or not valid rather we say it is valid for a specific purpose with a specific group of people. It is rare for an instrument to measure 100% valid, that is why validity is measured in degree. Validity involves collection and analyzing data to assess the accuracy of an instrument.

Type of Validity

Four types of validity to be consider in this context are content, construct, criteria, and face validity

1. Content validity is the extent to which the items develop adequately measure the trait which the researchers wishes to measure. To get it right experts in the related field views is often a good step in development and validation of an instrument. Messick, S. (2016) [7]
2. Construct validity represents a psychological trait which exist in every individual in some measure. For instance, a measure of reasoning ability implies that there is a quality called reasoning ability which accounts for performance on the test. On this view, if a test is constructed to measure anxiety and those who score high on the test exhibit high anxiety, then the test has construct validity.
3. Criteria-related validity is a standard used for comparison and serves as direct measure of the trait performance to be used in correlation. Criteria relate validity is therefore the validity of a test that is based on its results similarity to some other external criteria that is usually another test that purports to measure the same behaviour. In this type of validity, two sets of

scores are involved, the predicator scores and criteria scores. For example, the validity of an achievement test in science can be established by correlating its scores with a test of scientific aptitude (that is a new test score correlated against a similarearlier test score) See G.w.Orluene's work regarding validity.

Face validity reflects the level to which well-informed persons (experienced teachers in the subject area and expert in test and measurement) deem an instrument satisfactory in measuring what it is designed to measure. It refers not to what the test actually measure, but what they appear superficially to measure.

Factors Affecting Validity of An Instrument

The following factor in the test itself can prevent the test items from functioning as desired and thereby lower the validity. Nwana.O. C (2007) [8].

1. Length of the test – A test usually represents a sample of many questions. If the test is too short to become a representative one, then validity will be affected accordingly. Homogeneous lengthening of a test increases boith validity and reliability.
2. Unclear Direction- If directions regarding how to respoind to the items, whether it is permissible to guess and how to record the answer, are not clear to the pupil, then validity will tend to reduce.
3. Reading vocabulary and sentence structures which are too difficult – The complicated vocabulary and sentence structure meant for the pupils taking the test may fail in measuring the aspects of pupil performance; thus lowering the validity.
4. Inappropriate level of difficulty of the test items – when the test items have an inappropriate level of difficulty, it will affect the validity of the tool. For example, in criteria – referenced tests, failure to match the difficulty specified by the learning outcome will lower the validity.
5. Poorly constructed tested items – the test items which provides unintentional clues to the answer will tend to measure the pupils' alertness in detecting clues as well as the aspects of pupil performance which ultimately affect the validity
6. Ambiguity – ambiguity in statements in the test items leads to misinterpretation, differing interpretation and confusion; sometimes it may confuse the better student more than the poorer ones resulting in the discrimination of items in a negative direction. As a consequence, the validity of the test is lowered.
7. Test items inappropriate for the outcomes being measured -Many times what happens is that we try to measure certain complex types of achievement, understanding, thinking skills, etc with test forms that are appropriate only for measuring factual knowledge. This affects the results and leads to a distortion of the validity.
8. Improper arrangement of items - items in the test are generally arranged in order of difficulty with the easiest items first. If the difficult items are placed early in the test, it may make the students spend too much of their time on these and fail to reach other items which they could answer easily. Also, such an improper arrangement may influence the validity by having a negative effect on pupil motivation.
9. Identifiable pattern of answers – when the students identify the systematic pattern of correct answer (e.g. T.T.F.F or ABCD, ABCD) they can cleverly guess the answer and this will affect the validity.

Instrument And Instrumentation

To help distinguish between instrument and instrumental consider that the instrument is the course of action (the process developing, testing and using the device).

Instruments fall into two broad categories, researchers-completed and subject-completed, distinguished by those instrument that researchers administer versus those that are completed by participants. Researchers chose which type of instrument, or instruments, to use based on the research question. Onunkwo, G.I.N. (2002) [9].

Examples are listed below

Researcher-completed instruments	Subject-completed instruments
Rating scale	Questionnaires
Interview schedule/guides	Self-checklists.
Tally sheets	Attitude scale
Flow charts	Personality inventories
Performance checklist	Achievement/aptitude tests
Time-and-motion logs	Projective devices
Observation forms	Socio-metrics devices

Usability

Usability refers to the case with which an instrument can be administered, interpreted by the participant, and scored/interpreted by the researcher. Example of usability problems include:

1. Students are asked to rate a lesson immediately after class, but there are only a few minutes before the next class begins (problem with administration).
2. Students are asked to keep self-checklists of their, after school activities but the directions are complicated and the item descriptions confusing (problem with interpretation).
3. Teacher are asked about their attitudes regarding school policy, but some questions are worded poorly which results in low completions rates (problem with scoring/interpretation).

References

1. Anastasi A. Psychological testing. New York: Macmillan Publishing Co.; 1976.
2. Anastasi A, Urbina S. Psychological testing. Singapore: Pearson Prentice Hall; 1997.
3. Bracken BA, editor. The psycho-educational assessment of preschool children. Boston: Allyn and Bacon; 2000.

4. Bond TG, Fox CM. Applying the Rasch model: Fundamental measurement in the human sciences. Mahwah (NJ): Lawrence Erlbaum Associates; 2001.
5. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments. Am J Med. 2006;119(2):166.e7-166.e16.
6. Ikeotonye AI. The reliability and construct validity of the differential aptitude tests in Kaduna State, Nigeria. BENSU J Educ. 2018;1(1):13-20.
7. Messick S. Validity. ETS Res Rep Ser. 2016;2016:1-208. doi:10.1002/j.2330.
8. Nwana OC. Educational measurement and evaluation. Owerri: Bomaway Publishers Inc.; 2007.
9. Onunkwo GIN. Fundamentals of educational measurement and evaluation. Owerri: Cape Publishers International; 2002.
10. Orluwene GW. Fundamentals of testing and non-testing tools in educational psychology. Owerri: Cape Publishers International; 2012.

How to Cite This Article
Bokolo F, Karl O. Validity of an instrument in educational measurements. International Journal of Multidisciplinary Research and Growth Evaluation. 2026;7(2):62–64.