Quadrant-I (e-Text)

Details of Module and its structure

Module Detail	
Subject Name	**Education**
Course Name	Assessment for Learning
Course Code	EDU503
Module Name/Title	Validity: meaning, and different methods of estimating face, content, construct, predictive, criterion related and concurrent.
Module Code	AFL014
Pre-requisites	The learners should have 1) Knowledge on different assessment tools and techniques 2) Understanding about test and procedure of test construction 3) Idea about difference between different types of test items 4) Knowledge on various question formats to be used in tests
Objectives	After going through this lesson, the learners will be able to: 1. Explain the importance of determining the validity of a test. 2. Differentiate between different types of validity. 3. Estimate validity of test using different methods.
Keywords	Validity, face validity, content validity, construct validity, predictive validity, criterion validity, concurrent validity, estimation of validity

1. **Development Team**

Role	Name	Affiliation
Principal Investigator (PI)	Dr S K Bawa	Central University of Punjab
Subject Matter Expert (SME)	Ajithkumar.C	IGNOU

Contents

1.0 Introduction

Teachers regularly assess their student's achievement using various tools and techniques like achievement test, interview, rating scale, questionnaire, etc. This aspect or process or component related teaching learning-process is called assessment. Even we use terms like measurement and evaluation. In later part of this unit we may briefly discuss the differences among these terms. Assessment is an integral component of the instructional system and provide teachers an understanding and knowledge about the achievement of their students. Assessment also helps teachers to reorganise their teaching learning activities/learning experiences for better understanding of subject matter by the students. One of the common tools used to asses achievement of the children is achievement test or we simply call 'test'. Being teacher trainees, you must develop proper understanding about the concept test. What is a test? How it is different from terms like appraisal and examination? How tests are constructed? What are the qualities of a test? We may briefly discuss these aspects. But the focus of this module is one of the important feature/qualities of the test i.e. validity. Anyone developing tests needs to understand the meaning and importance of validity. Along with that, they should also know the ways of checking the validity. If the tests are valid, then we consider it as good test.

2.0 Assessment Basics

Before any teaching session is undertaken, the teacher decides the objective/s of the instructional process. Thereafter the teaching learning process is carried out and at the end of the process, teacher check the whether the objectives of the instructional process is achieved or not. The checking can be conducted at various stages like in-between the classroom process (formative assessment), at the end of the teaching of the teaching session (formative assessment) or after the completion of a complete unit/whole academic year (summative assessment). For checking the achievement of objectives teachers generally use test/achievement test. In tests, students are provided with questions to which they respond. Depending upon the answer students provide, they are awarded marks/grades and finally a judgement is made about the performance. Thus, the process of taking quantity of something is called measurement (eg: finding marks scored by students) and the process of assigning quality of something is called evaluation (making judgment like

students performed better than previous year in examination). We also use the term assessment and it describes the process of taking quantity and making inference. For example, finding marks scored by students and dividing them into low, average and high scoring groups. Assessment is of different types. One classification is formative and summative assessment. The second classification is assessment for learning, assessment of learning and assessment as learning. You have already understood these concepts in the previous module.

Now let us briefly discuss about self-appraisal, examination and test. In the context of education these terms are interchangeably used. Self-appraisal is the documentation of all the activities performed by an individual during specific time period. Students sometimes are asked to submit self-appraisal report. This helps teachers to evaluate their academic performance. Examination is nothing but basically refers to a test. We are familiar with examinations like quarterly, half-yearly and annual examination. Examination is organised to evaluate student's performance. What about test? Test is an instrument used to collect information as verbal or non-verbal responses. In the context of education (teaching -learning process) there are different types of tests like achievement test, paper pencil test, diagnostic test, remedial test, oral test and practical test, ect. Apart from that there are standardised and teacher made tests, criterion and norm referenced tests, achievement tests and diagnostic tests.

Achievement test is used to find out the achievement of students in learning. Generally objective and essay type questions are included in achievement test. And for constructing achievement test (especially if it is standardised test like state board examination/test) test makers follow certain steps. This step includes planning, preparing design, preparing blue print, writing test items, preparing scoring/marking scheme and preparing question wise analysis. If the test makers follow these steps, then such tests are called standardised test. A test is said to be standardised test (in other words a good test) when it stratifies certain qualities. Even teacher made tests are also said to be good test if it stratifies the same qualities. These qualities include validify, reliability, objectivity, comparability and discrimination power. In this module we will discuss in detail about validity, different types of validity and methods for estimating different validities.

3.0 Meaning of Validity

Suppose a teacher wants her students to develop the ability to calculate the value of 'g' (acceleration due to gravity), in such a case, the teacher will first make her children understand the concept of acceleration due to gravity. Thereafter the mathematical formula for calculating 'g' will be taught. Then the teacher may provide different mathematical problems so that students will learn to compute the value of 'g'. After that the teacher would check whether the students have understood the procedure of calculating 'g'. For that, what all different types of tests the teacher can use? Let us assume two ways of assessing children. First, teacher asks question like; what is acceleration due to gravity? Second, teacher asks questions like: what is acceleration due to gravity followed by providing a mathematical problem to

calculate the value of 'g'. If we analyse both the question, one can easily understand that, the first question fails to measure what teacher is intended to measure. Because the teacher was expecting, students will answer the question by giving example computing the value of 'g'. But students end up giving the definition of acceleration due to gravity. While for the second questions students gave both definition and computed value of 'g' and that is what the teacher was intended to measure. Hence it is clear that, the first question failed to measure what its was supposed to measure. This happened because the question wasn't properly formulated.

Let us discuss one more example. A teacher develops a question like, what is the difference between passive and active voice? The teacher expects that, students will answer citing examples. But in reality, this may not be the case. Some students simply describe active and passive voice without example. In such a case, the intention of the teacher fails because the question was not valid. If the question was like, describe passive and active voice by giving examples. Then the students might have clearly described by giving examples. Thus, when constructing test item, it should be able to clearly identify/measure what the teacher is intending to measure. A test is good test if it measures what it is intended to measure and this feature of the test is called validity. So, a test is said to be valid if it measures what it is intended to measure. The concept of validity was formulated by Kelly (1927, p. 14) who stated that a test is valid if it measures what it claims to measure.

It is important that, the test maker should check the validity of the test for different reasons. The first and foremost, the test maker can ensure that, the test prepared shall definitely help him/her to find out what he/she was intended to measure. Secondly, validity is one of the criterions that determines, the test is a good test. Third, the test maker would be able make sure that, the test would give the feel that it will measure what he/she is going to measure. Fourth, validity helps to check the inclusion of all contents from which the test items are constructed. Finally, validity help to check the test scores with an external criterion so that future performance/behaviour of the examinees can be predicted.

In the previous module you have studied about the reliability of a test. Whenever a test is constructed one must check the reliability so that test maker can make sure that, the test would yield consistent result for repeated attempt by the examinees. Thus, reliability refers to the consistency of scores i.e. even if the student attempts the test multiple times, the score would remain same. At this point you must also understand the relationship between reliability and validity. The relationship is that; a reliable test may not be valid always but a valid test is always reliable. A reliable test which is not valid will not give consistent scores for repeated attempts but a valid test would definitely produce consistent scores.

4.0 Types of Validity and its Measurement

We discussed that, validity refers to the degree to which the test or other measuring device is truly measuring what was intended to measure. As you know, the objective of teaching is to develop behavioural changes that are

observable. The behavioural changes are grouped into three imaginative domains as suggested by Benjamin S. Bloom i.e. cognitive domain, affective domain and psychomotor domain and we call them Blooms Taxonomy of Educational objectives. Later Blooms Taxonomy was revised. In the Indian context the implementation of National Curriculum Framework in 2005 emphasised constructivist approach of teachings-learning which engages children in active learning environment for themselves to discover/construct knowledge by their won. Thus, whenever teacher develop a test for assessment, these factors must be taken into account so that, he/she can successfully assess the progress of learning.

Whatever may be the objective, the test maker must check the validity of the test in different dimensions. For example, if a test is constructed to check the psychomotor skills, the way questions are framed, contents covered, criteria considered and many other factors needs to be considered. This leads to different kinds of validity. The following are the major types of validity;

- Face validity
- Content validity
- Construct validity and
- Criterion validity (includes concurrent & predictive validity)

Among these, the content, construct, predictive and concurrent validity are commonly estimated during the construction of test. Let us discuss the different types of validity in detail and the ways/methods of estimating them.

5.0 Face Validity

Suppose a test maker develops some test times (or question paper) to check the understanding of students to add two numbers, then generally the question paper would contain questions related to addition of numbers. If, there are 25 questions, all the test items are generally questions related to addition. Thus, for students who glance through the question paper (test items) gives a general feel that, all the questions are addition related. Not only for the students, whoever glancing through the question paper get the same feel. Thus, the question paper gives a feel that, it 'appears to measure; the knowledge and comprehension of addition of two numbers. This feature of the test is called face validity. Face validity refers to the extent to which the test appears to measure what it is to be measured. If the test mentioned above has few questions related to subtraction, then someone reading the test may get confused that, whether the test is checking understanding of students in addition or subtraction. In such case we conclude that the test fails to have face validity.

Face validity describes the extent to which the test appears to measure what it is supposed to measure. If the external appearance of the test paper gives an impression that the test is going to measure the construct/variable/objective that the test is going to measure, then one can say that the test has face validity. Thus, a test designed to psychomotor skills must contain test items/questions that really check the psychomotor skills. Similarly, a test designed for checking the application of particular theory should only contain questions related to it. If that question paper consists test

items/questions belonging to the affective domain, then it may lack the face validity. The tests that lack face validity is a weak test/bad test.

Face validity is the weakest or sophisticated methods of validating a test. However, one has to determine the face validity of the test. This is one of the first validity every test maker should consider while developing test. If the test lack face validity, then of course it fails to achieve the objective of the test. To estimate/determine face validity, the test maker can make use of various methods. One of the effective methods is analysis of the questions. The analysis of the test items can be done by different persons. First, the individuals (usually respondents/students) may be asked to validate the test items. Or else any experts like test makers, teachers, parents, etc may be asked to analyse the questions for face validity. Even the items may be gat face validated form subject experts. But the validation done by subject experts is treated as checking for content validity (will be discussed next). While asking respondents/students/teachers/test makers, etc, they may also be alternatively provided with rating scale so that each question can be rated on particular defined criteria.

6.0 Content Validity

Content validity is the most important validity concerning a test. Content validity refers to the extent to which the test items represent adequate percentage of the content intended to measure a particular construct/variable /objective. In other words, the test must include all the content or represent the content relating to a particular variable or construct. For example, if a test on 'examination anxiety' is developed, the test should contain test items/questions relating to various dimensions that creates examination anxiety. The examination anxiety may be of internal or external reasons. Thus, test items representing both internal and external factors must be included the test. Similarly, a test designed to measure the intelligence of the students must contain items related to different aspects of intelligence. If the test consists of items covering different aspects of intelligence, then we call that test a content valid test. Thus, content validity describes the extent to which it covers all the content representing a construct to be measured is adequately covered. Content validity also called as Rational Validity or Logical Validity or Curricular Validity or Internal Validity or Intrinsic Validity.

Let us discuss the concept of content validity with respect to classroom assessment. Generally, the teaching is organised to bring behavioural changes among the students in three domains i.e. cognitive, affective and psychomotor domain. Suppose a teacher teaches the concept of inertia to his/her students, then the teacher might have definitely thought of making her children understand the concept of velocity and its practical applications in life. Thus, a test designed to measure these objectives must contain test items that check the concept of inertia and its practical applications in life. If it is satisfied, that test is content valid. If the teacher is developing test for half yearly examination, then surely the test must include items from various unit (chapters). The teacher must give adequate weightage to whatever chapters completed before conducting the half yearly examination. Not only weightage to the content but also, items must be planned in such a way that,

adequate percentage of items must be included belonging to cognitive, affective and psychomotor domain. Thus, there are various ways by which content validity can be ensured while developing test.

After developing a test, the first thing to be checked is the face validity. If the test has face validity, then the test maker must check the content validity. To determine the content validity, the best and most popular method is expert judgement. Subject experts may be asked to check the adequacy of content related to the construct/variable/objective. When the test is provided to the subject experts, they analyse the questions and check whether the contents for checking a particular construct is sufficiently provided and all contents have been covered or not. In the case of achievement test, they check for adequate representation of contents for assessing cognitive, affective and psychomotor skills. In addition, they may check items from various unit/chapters/sub units have been included in the test.

7.0 Construct Validity

Construct validity refers to the extent to which the test is said to measure the construct/psychological variable. Construct validity will help the test constructor to confirm the measurement of a construct more accurate. Thus, we call a test having construct validity if it accurately measures the construct/variable under consideration. For example, a test maker develops a test to measure the self-esteem of the individuals, then if the test has construct validity the test would definitely measure the self-esteem in an accurate way. Construct validity is also known as "Psychological Validity" or 'Trait Validity' or 'Logical Validity'. The idea of construct validity was put forwarded by terms (Cronbach & Meehl, 1955).

Let us also discuss an example in the teaching learning context. The achievement test constructed to assess children's ability to 'critique the positive and negative aspects described in a poem in English', must contain test items helping children analyse and critically evaluate it. Thus, the test items must be constructed in such way that, it should provide a chance to critically analyse it. In such a test, items checking the knowledge of the children won't suffice but the item should of the nature that, it would help children to relate the response to social circumstances and there by evaluate the societal situation. Thus, before constructing the test items, the test maker should be clear about, what is going to measured? What all items shall accurately measure what I am going to assess? Are the items sufficient to check the construct/variable I am intended to measure?

To estimate the construct validity of the test, there are no single methods instead multiple methods are being practiced. The process of determining the construct validity is considered as construct validation. In construct validation, the purpose for which the test is constructed by test maker is calculated and a decision is made such that the test is fit for measuring what the test maker was intended to measure. To determine construct validity, either correlational method (convergent/divergent) or factor analysis is used. In convergent validation, the test is correlated with another test measuring the same construct/variable. If the correlation is high, then the test is said to

have convergent validity. While in divergent validity, the test is correlated with another test measuring different construct/variable. If the correlation is low, then the test is said to have divergent validity. Factor analysis is also used to determine construct validity and is a complex statistical procedure. Factor analysis as name indicates analyses the test items and determines the accuracy of the items in measuring the construct/variable.

8.0 Criterion Validity

There are instances where, the scores obtained by an individual is correlated with the already/existing scores of another test measuring the same construct/variable (this is considered as external criteria). For example, the scores in aptitude test (constructed by test maker) is correlated with a standardized test of aptitude. This method of comparing scores leads to criterion validity. Criterion validity refers to the extent which the individual's score in a test are correlated with score of other variables (known as criteria). It is not necessary that, the score is correlated with the same criterion but different criterions can also be used. For example, the student's performance in achievement test is correlated with scores in test anxiety. In such case negative correlation is a good proof that, the test is valid in terms of criterion. Thus, here achievement test is the test and test anxiety are the criterion.

The comparison of one score with an external score(criterion) is done in two ways and this leads to concurrent validity and predictive validity. The word concurrent means 'same time. Thus, in concurrent validity the score of and individual is correlated with the external criteria at the same time. Thus, the test scores are correlated with criterion scores at the same time. So, Concurrent validity refers to the extent to which the test scores correspond to already established or accepted performance, known as criterion. For example, the test scores of intelligence test (developed by test maker) is correlated with standardized (external criterion) scores of Wechsler Adult Intelligence Scale or Stanford-Binet Intelligence Test. If the correlation is highly positive (coefficient of correlation), then we say the test is concurrently valid. This type of validity is also known as "External Validity" or "Functional Validity". In the case of predictive validity, the test are scores are correlated with criterion score at later point of time. This is used to predict the future performance. Thus, the predictive validity refers to the degree to which a test accurately predicts a criterion that will occur in the future. For example, the scores obtained in entrance test, banking recruitment test, clerical tests, etc. predict the future performance of individuals in such areas. This type of validity is sometimes referred to as 'Empirical validity' or 'Statistical validity' as our evaluation is primarily empirical and statistical.

We have discussed that criterion validity is the extent to which test scores are correlated with an external criterion. Thus, to determine estimate the criterion validity (and concurrent validity or predictive validity) the correlation method is used. The test score is correlated with external criterion at the same time in concurrent validity while the test scores are correlated with an external score at a later point of time.

9.0 Summary

Assessment of student progress is carried in different ways. This may be during the classroom teaching-learning process or after completion of a unit/chapter or after completion of few units/chapters. One of the common tools used for assessing students is achievement test. Achievement tests majorly fall into two categories i.e. teacher made or standardised test. Usually standardised test follows certain procedures for standardisation. A standardised test is said to be a good test. Thus, it is mandatory to check various parameters of the test. One among is the validity of the test. Validity refers to the extent to which a test is intended to measure. Validity depends on different factors. In this module we have discussed about those factors and different types of validity. Also, the different ways of determining validity have been discussed. Teachers must consider these factors and determine the validity of the test constructed for assessing the progress of the children in learning.

EDU503 Assessment for Learning Module: AFL014

Quadrant-III - (Learn More / Web Resources / Supplementary Materials):

Books, articles, research papers, journals, case studies etc.:

Aggarwal, J.C. (2005). *Essentials of examination system: Evaluation, tests, and measurement.* New Delhi: Vikas Publishing House Pvt Ltd.

Anastasi, A. (1976). *Psychological testing*. New York: Macmillan Publishing Co.

CBSE. (2010). *Continuous and comprehensive evaluation manual for teachers.* New Delhi: CBSE.

Check, J. and Schutt, R.K. (2012). *Research methods in education.* California: Sage Publication.

Cook, W.W. (1958). *Educational measurement.* Washington DC: American Council of Education.

Cozby, P.C. (2001). *Measurement concepts. methods in behavioral research (7th ed.).* California: Mayfield Publishing Company.

Cronbach, L. J. (1971). *Test validation. In R. L. Thorndike (Ed.). Educational measurement (2nd ed.).* Washington, D. C.: American Council on Education.

Cronbach, L. J., and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin,* 52, 281-302.

Eble, R.L. and Frisbie, D.A. (1991). *Essentials of educational measurement.* New Delhi: Prentice Hall of India Pvt Ltd.

Hathaway, S. R., & McKinley, J. C. (1943). *Manual for the minnesota multiphasic personality inventory.* New York: Psychological Corporation.

Kelley, T. L. (1927). *Interpretation of educational measurements.* New York: Macmillan.

Koul, L. (1997). *Methodology of educational research.* New Delhi: Vikas Publications

Linn, R.L. (1968). *Educational measurement.* USA: Macmillan.

Linn, R.L. and Miller, D.M. (2005). *Measurement and assessment in teaching.* NJ: Pearson.

Nandra,I.D.S.(2005).*Learning resources and assessment of learning.* Patiala: 21st century Publications.

National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC: Authors.

NCERT. (2005). *National curriculum framework, 2005.* New Delhi: NCERT

Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement,* 22(4), 287-293.

Singh, A.K. (2002). *Tests, measurements and research methods in behavioural sciences.* Patna: Bharti Bhawan.

Links to web sites giving additional readings, Wikipedia, blogs, open source content etc.:

https://www.simplypsychology.org/validity.html
https://opentextbc.ca/researchmethods/chapter/reliability-and-validity-of-measurement/
https://explorable.com/types-of-validity
http://edcan.org.au/edcan-learning-resources/using-edcan-resources/implementation-resources/assessment/reliability-validity
https://study.com/academy/lesson/validity-in-assessments-content-construct-predictive-validity.html
https://fcit.usf.edu/assessment/basic/basicc.html
https://assessment.tki.org.nz/Using-evidence-for-learning/Working-with-data/Concepts/Reliability-and-validity
https://www.ascl.org.uk/news-and-views/blogs_detail.html?shorturl=the-four-pillars-of-great-assessment-validity
https://www.slideshare.net/MAHIJUSTICE/validity-its-types-measurement-factors
https://www.slideshare.net/Azia1980/validity-65888959
https://slideplayer.com/slide/5827748/
https://slideplayer.com/slide/5827748/
http://www.umich.edu/~psycours/381/masuda/381notes/381pptCH5.ppt

Questions

1) The commonly tool used for measuring progress in learning of students includes the following
   a) Achievement Test
   b) Interview
   c) Rating scale
   d) All the above
2) A test is said to be valid if it measures what it is intended to measure
   a) True
   b) False
3) The extent to which the test appears to measure what it is to be measured is called
   a) Content validity
   b) Face validity
   c) Construct validity
   d) None of the above
4) The best method to determine face validity is distribute the test to---------- for analysing the test items.
   a) Students
   b) Teachers
   c) Parents
   d) All the above
5) The extent to which the test items represent adequate percentage of the content intended to measure a particular construct/variable /objective is called content validity
   a) True
   b) False
6) The best possible method of estimating content validity is
   a) Student judgement
   b) Parent judgement
   c) Expert judgement
   d) None of the above
7) Construct validity is also known as
   a) Psychological Validity
   b) Trait Validity
   c) Logical Validity
   d) All the above
8) To estimate construct only one single method is available
   a) True
   b) False
9) In criterion validity, the test items are correlated with an external criterion
   a) True
   b) False

10) The extent to which score of an individual is correlated with the external criteria at the same time is called

a) Content validity
b) Concurrent validity
c) Predictive validity
d) Face validity

Answers: 1-a, 2-a, 3-b, 4-d, 5-a, 6-c, 7-d, 8-b, 9-a, 10-b