UNIT 6: VALIDITY

UNIT STRUCTURE

6.1 Learning Objectives

6.2 Introduction

6.3 Meaning and Definitions of Validity

6.4 Factors affecting Validity of a test

6.5 Types of Validity of a test

6.6 Methods of estimating Validity of a test

6.7 Let Us Sum Up

6.8 Further Reading

6.9 Answer to Check Your Progress

6.10 Model Questions

6.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- define Validity of a test
- understand the factors that affect Validity of a test
- explain the different types of Validity of a test
- explain the methods of estimating Validity of a test

6.2 INTRODUCTION

Psychological tests and other tools of measurement serve various purposes in education. Various decisions in educational theory and practice depend on the results or data collected using psychological tests and other tools of measurement. However, if the data collecting instruments give erroneous results, then it would lead to wrong decisions and would fail to serve its objective. Therefore, an instrument of data collection must be valid to give correct results about the object of measurement. An attempt has been made in this unit to give a clear understanding about the concept, factors, types and methods of validity of a test.

6.3 MEANING AND DEFINITIONS OF VALIDITY

Validity is the most important technical criterion of a good test and other types of measuring instrument. Validity of a test may be defined as the extent to which a test measures what it claims to measure. In other words, validity answers the question of whether the test measures what it intends to measure. The concept of Validity is so fundamental that it precedes the aspect of Reliability in importance. A test may be reliable but not valid. However, if a test is not valid, then its reliability has no meaning or importance. On the other hand, a valid test is usually reliable. The concept of validity is not absolute and hence validity of a test is always understood or determined with reference to the purpose of the test. Therefore a test may be valid for one purpose but not for others. Any test can be differentially valid, given the various uses to which it is put. For example, a test constructed to assess the capacity for logical thinking of adolescents may serve the stated purpose extremely well, but when the same test is given to students of a different age group, it may fail to give the desired results. Different psychologists have given various definitions of validity. A few of the important definitions are quoted below:

According to **L. J. Cronbach**, "Validity is the extent to which a test measures what it purports to measure."

F. S. Freeman defines validity as, "An index" that "shows the degree to which a test measures what it purports to measure when compared with accepted criterion."

Henry Garrett stated that, "The validity of a test or of any measuring instrument depends upon the fidelity with which it measures what it purports to measure."

Anastasi says, "The question of test validity concerns what the test measures and how well it does so."

In order to ascertain if a test actually measures what it aims to measure, two sets of measures of the variable aimed to measure are needed. One set of measure is obtained by the test itself and the other set is obtained by some well-established and acceptable instrument that

measures the variable. The latter is known as the Criterion. If the scores or measures obtained by the test correspond well with the criterion then the test may be accepted as valid. Thus, the validity of a test is the correspondence or correlation between the variable measured by the test and the variable objectively defined and measured. The level of correspondence is shown by the coefficient of correlation obtained from measures of the test and criterion which is also known as validity Coefficient. Hence, Monroe states that, "under the head of validity we inquire into the degree of constancy of the functional relation existing between the scores yielded by the test and the abilities specified as being measured in the statement of that function." Thus, Validity is capable of a statistical interpretation by means of a coefficient of correlation. This coefficient of correlation is estimated between the two series of scores: those of the tests and of the accepted criterion.

6.4 FACTORS AFFECTING VALIDITY OF A TEST

The validity of a test is affected by various factors. The most important factors that affect the validity of a test are as follows:

- **Factors inherent in the test itself:** The validity of a test largely depends on the items included in the test. In fact, validity of the test depends on how well each item included in the test could achieve the objectives of the test. If the items in a test fail to serve the purpose for which they were constructed, then the test will lack validity. Moreover, there are other factors related to the test that affect the validity the test. These factors include:
  - **Clarity of instruction:** Clarity of instruction about how to answer and record the items, if guessing is allowed etc. always raises the validity coefficient of a test. On the other hand, unclear instruction to the testee may reduce validity of a test.
  - **Length of the test:** The length of a test too affects its validity. A test is a collection of many items that represent the content or cover the objectives of the test. But if the test does not include

enough items, it may fail to represent the whole content or objectives affecting its validity adversely.

- **Language used by test items:** The language of test items should always be clear, simple and unambiguous. Complicated vocabulary, complex structure of sentence, ambiguous words and phrases make it difficult for the takers of the test to understand the items which lowers the validity of the test.
- **Difficulty level of test item:** Difficulty level of items included in the test may also affect the validity of the test. Generally it has been observed that when the difficulty level of the items such in a test is homogeneous, validity of the test is low; on the other hand, when difficulty level in a test is heterogeneous, validity of the test is high.
- **Item discrimination power:** Like the level of difficulty, the power of discrimination of test items also affects the validity of a test. Tests with items having higher discrimination tend to possess higher validity and tests with items having low discrimination are of lower validity.
- **Defective construction of items:** Defective item construction lowers the validity of the test. Items that provide clues to answer, or items that carry answers to some other question, or such other defects in item construction lower the validity of the test.
- **Inappropriate items:** Sometimes items are constructed to measure one trait or variable, and in practice they measure some other trait or variable. Such inappropriate items lower the validity of the test.
- **Faulty arrangement of items:** Items in a test must be arranged logically and in an ascending order of difficulty. If items are not arranged maintaining a logical sequence, that may affect the validity of the test adversely.
- **Guessing:** If the items and the test is constructed and arranged in such ways that it allows the students to guess the correct answers then the validity of the test will surely be doubtful.

- **The Criterion of validation:** In order to estimate validity of a test the scores obtained in the test have to be compared with some acceptable measure or criterion. Various indices of validity may be obtained when different criterion is used to validate the test.
- **Method of estimating validity:** There are various methods of estimating the validity of a test. Different estimates of validity are obtained when different methods are used to estimate validity of the test.
- **Nature of the group:** Validity of a test also depends on the nature of the sampling group on which the test is administered for estimating validity. Validity is always specific to a particular group of individuals.
- **Factors related to test-takers' response:** The psycho-physical condition of the test-takers, their emotional and mental conditions besides cultural factors including socio-economic condition, social class, language, religion, educational background, sex etc. influence the scores of individuals and thereby affect the validity of a test.
- **Factors related to administration and scoring of the test:** Sometimes administration and scoring procedure of a test may interfere with its validity. Factors such as unclear direction, inappropriate time, cheating, and unfair help to candidates, unreliable and subjective scoring and adverse psycho-physical conditions affect the scores of individuals on the test ultimately affecting its validity.

The factors discussed above and many other similar factors may affect the validity of a test.

6.5 TYPES OF VALIDITY OF A TEST

Validity of a test is closely associated with the purpose of the test. It is for this reason that the validity of a test is always estimated with reference to the purpose of the test. Every test has its own distinctive purpose. On the basis of the different purposes of different tests validity is estimated

using various methods. Thus each test has its own different type of validity. Different test-experts and Psychometricians have classified validity into different types and named them somewhat differently. Out of the many different types, the following are the most popular and widely used types of validity.

- Face validity
- Content validity
- Criterion validity
- Construct validity
- Factorial validity

Face validity: Face validity is the simplest type of validity of a test. However, face validity is not validity in actual psychometric meaning. It is a superficial or extra quality of a test and it refers only to what the test 'appears to measure' and not to what it actually measures. Basically, face validity is not one of validity in the technical sense but from a practical point of view it is important for establishing rapport and to increase acceptability of the test among stakeholders. The items of the test should not appear to be inappropriate and irrelevant. When the test items appear irrelevant or inappropriate, the result is always poor regardless of the actual validity of the test. Thus, if the test consists of items that look to measure the variable that the test intends to measure, then the test has face validity. For example, if items in a test constructed for selecting candidates for recruitment of teachers use words and phrases that relate to classroom and teaching, then it may be said that the test has face validity. But at the same time it cannot be said that the items in the test could actually measure the skills required for teaching. It is noteworthy that even though a test possesses face validity, it may not be truly valid. Hence, while assessing the validity of a test, only face validity is not sufficient. It is generally considered only at the preliminary stage of validating a test. Moreover, when there is an urgent need to estimate the validity of a test within very limited time, then face validity may be used. Face validity, in a sense, ensures relevance to the context by using appropriate words and phrases in the items.

Content validity: Content validity is also known as curricular validity, Rational or Logical validity, internal validity etc. Content validity is particularly applicable in the tests of educational achievement. In the field of education, learners are guided through a curriculum in order to achieve the goals and objectives of education. At the end of education, if the items included in the test to measure the learner's progress can represent the curriculum correctly and can assess the extent of achievement of the objectives, then the test is said to possess content or curricular validity.

Content validity is the extent to which the items of a test match the instructional objectives. In other words content validity is the process of determining whether the sample used for the test is representative of the area or aspect to be measured. It is non-statistical type of validity. Thus, content validity may be defined as the extent to which a test measures a representative sample of the subject matter or content and the behavioural changes being measured. However, it goes without saying that the content validity of a test must always be examined with reference to a particular objective to be measured. The measure of content validity is presented subjectively after careful examination of the content of the test with the objectives of the course of instruction.

Content validity is mostly relevant for tests used to evaluate achievement, education, occupation, and the like. At the same time, content validity is not appropriate for test of aptitude, personality, intelligence etc.

Criterion Validity: Criterion Validity is also known as Empirical Validity or Statistical Validity. In criterion validity or criterion-related validity, the scores obtained by a test are compared and correlated with some criterion. A criterion is a standard with which scores on a test are compared or evaluated. Thus, criterion validity is the extent to which the test scores correspond with the criterion. It is a quantitative or statistical procedure of calculating the coefficient of correlation between scores on the test and the criterion. The validity of the test is also determined by its ability to predict performance on the criterion. There are different types of criterion related validity based on different criteria that are used as the standard for judgment.

They are as follows:

Predictive Validity: Predictive Validity refers to the extent to which a test can predict something that it should theoretically be able to predict. The predictive validity of a test is determined on the basis of an established criterion. A test has predictive validity, if scores on it predict future performance of test-takers in a particular field. In order to calculate predictive validity the test is administered and scores are obtained. In the next stage, the scores of the same individuals signifying their actual performance in the particular area (criterion) are compared with their previous scores on the test. If the two sets of scores correlate closely then the test is said to possess predictive validity. For example, scores obtained by students in an entrance examination are compared and correlated with their actual scores in that course or curriculum (for which they were selected by the entrance examination) in order to estimate predictive validity of the test. Predictive validity is substantial for entrance exams, vocational tests, and aptitude tests.

Concurrent Validity: Concurrent Validity is the extent to which a test correlates with a previously validated test for the same construct. In order to estimate concurrent validity of a newly constructed test, the scores obtained by students in the (newly constructed) test are compared and correlated with their scores on a previously validated and standardized test. For example, scores on a newly constructed test are compared with Stanford-Binet in order to estimate validity of the new test. When a new test is validated against some established and standardized test, the latter is known as the criterion of validation. When a test constructor makes a new test and wants to make it sure that the test accurately measures the variable for which the test was constructed, he has to follow the procedure. In order to do this he has to find out established tests for measuring the same variable, administer both the tests, and compare the results of the tests with each other. If the results closely correlate, then the new test is having concurrent validity. Criterion validity is generally used for validation of intelligence, multiple aptitude and tests of personality.

Construct Validity: In psychology, a construct is a concept developed in the mind to explain some phenomenon. Psychological tests are constructed to measure different constructs. While making a test for measuring a particular construct, the test maker must clearly define the construct in clear terms what he wants to measure. Construct validity is the extent to which a test measures the theoretical construct as defined by the test maker. According to Anastasi, construct validity is the extent to which the test may be said to measure a theoretical construct or trait or psychological variable.

Factorial Validity: A test is usually constructed to measure different mental construct or factors. For measuring different such factors or constructs, different units of items are included in a test. Factorial validity refers to the coefficient of correlation between the scores on units of items constituting a factor and the scores on the total test. Tests are generally composed of different factors. The correlation of the test with each factor constituting the test is called factorial validity of the test.



CHECK YOUR PROGRESS

Q 1: What is validity?

.....

Q 2: What is face validity?

.....

Q 3: Write any two factors that affects the validity of a test.

.....

Q 4: What are the two different types of criterion-related validity?

.....

Q 5: What is Construct Validity?

.....

6.6 METHODS OF ESTIMATING VALIDITY OF A TEST

Validity of a newly constructed test may be estimated using various methods. Following are some of the common methods used to estimate validity of a test:

Correlation Coefficient method: Test constructors generally validate new tests by correlating the scores of the new test with the results obtained from the best accepted test, which is known as the criterion of validation. The coefficient of correlation gives the extent of validity index of the test. A high coefficient of correlation between the results of the two tests indicates high validity for the new test. When coefficient of correlation is used to describe validity, they are called validity coefficients. According to Safrit and Wood, in general validity coefficients of 0.90 or higher are desirable but values above 0.80 are acceptable. They also say that, for tests used to make predictions lower validity co-efficient (0.50 to 0.60) are acceptable.

For the purpose of validating new tests using the correlation coefficient method, Pearson's method of correlation is most widely and popularly used. However, the technique of correlation depends on the nature of data obtained on the test as well as on criterion. According to Berg and Latin (2008), following methods of correlation are commonly used for validity of newly constructed tests:

Validity type	**Statistics (method of correlation)**
Face validity	None
Content validity	None
Concurrent and predictive validity	✓ Pearson Correlation ✓ Spearmen Correlation ✓ Multiple Correlation
Construct validity	Independent t

Source: Berg and Latin (2008)

The table presented above clarifies one important thing about the Correlation Coefficient method that this method can be applied only in the

cases of concurrent and predictive validity.

Cross-Validation Method: Cross Validation is the process of checking the validity of a test for a second time on a second sample. Whatever method is used for validating the test, cross validation is very important. It is a process in which a test is administered to a different sample of individuals drawn from the same population for the purpose of verifying the results obtained from the original process of validation. In other words, cross validation means trying out the previously developed and refined test on a completely new group for a second examination of its validity.

The common types of cross validation methods include:

Holdout validation: Though a method of cross validation, in this method the data never are crossed over. Observations or scores are chosen randomly from the preliminary administration to form a sample of the validation data, and the remaining observations are retained as the training data. Normally, less than one third of the initial sample is used for validation data.

K-fold cross-validation: In this method, the original sample is divided into K subsamples. Out of the K subsamples, one subsample is used as the validation data, and the remaining sub-samples are used as training data. The cross-validation process is then repeated K times (the folds), with each of the K subsamples used once as the validation data. The results from all the folds are then averaged or combined to produce a single estimation.

Leave-one-out cross-validation: This method uses a single observation from the original sample as the validation data, and the other observations as the training data. This process is repeated and each observation in the sample is used once as the validation data.

Expectancy Table Method: An expectancy table is a two-way table showing the relationship between two tests. According to Helmstadter, an expectancy table provides the probabilities indicating that students achieving a given test score will behave in a certain way in some second situation. This method can be used to help in predicting the performance level on another test; in other words, an expectancy table can be used to estimate

predictive validity.

An expectancy table helps in estimating the chances of an individual's success in some future activity from his present performances on a test. It is a table of numbers from which the likelihood or probability that a particular individual or group of individuals will attain a specified level of success can be predicted.

An expectancy table is a simple and straightforward method of estimating predictive validity of a test. This table may help in answering questions like: what are the probabilities of a student securing 1st Division in Secondary School Leaving Examination completing all of the six semesters of Degree Course? Or what are the probabilities of a student passing in the 3rd Division of successfully completing the Degree Course? An expectancy table may help in answering such questions. The following illustration may show how an expectancy table works:

Table: Expectancy table

Divisions in S.L.E	Percentage of students Completing Semesters					
	1	2	3	4	5	6
1st	99	95	91	89	89	85
2nd	93	90	85	83	78	75
3rd	80	75	71	67	61	57

The table above shows that the probability of a student securing 1st Division in Secondary School Leaving Examination of completing all the six semesters of Degree Course is 85%. Similarly probabilities of a student passing in the 3rd Division of successfully completing the Degree Course are 57% only.

It is important to note that the format of an expectancy table is not uniform. The format and the arrangement of data depend on what the user wants to accomplish or predict. However, all expectancy tables provide estimation of the probabilities that a particular level of performance may be predicted if the previous test score is known. Thus an expectancy table estimates the validity of the test scores in terms of their predictive efficiency in the form of probabilities.

Method of Factor Analysis: Factor analysis is a statistical method used to study inter-correlation of items in a test to establish its validity. Factor analysis is a technique that is used to reduce a large number of items in a test into fewer numbers of factors. The correlation of the test with each factor constituting the test is calculated to find out the contribution of each of the factors to the total performance of the test and also to find out if all the factors are internally consistent and assessing the same variables as intended by the test.

Besides the methods discussed above, other methods such as item analysis, expert's judgment and opinion etc. are also used to validate newly constructed tests.



6.7 LET US SUM UP

After going through the unit, it can be summarized as---

- Validity refers to the truthfulness of a test. Validity of a test is defined as the extent to which a test measures what it claims to measure. The concept of validity is so fundamental that it precedes the aspect of reliability in importance. A test may be reliable but not valid. However, if a test is not valid, then its reliability has no meaning or importance. On the other hand, a valid test is usually reliable. The concept of validity is not absolute and hence validity of a test is always understood or determined with reference to the purpose of the test. Therefore a test may be valid for one purpose but not for others.
- Validity of a test is affected by various factors such as: factors inherent in the test itself including difficulty and discrimination level of items, length of the test, quality of items, clarity of instruction, guessing etc.; the criterion of validation; methods of estimating validity; nature of the group; factors related to test-takers' response; factors related to administration and scoring of the test and the like.
- Different test-experts and Psychometricians have classified validity into different types; the most popular and widely used types of validity

are: Face validity, Content validity, Criterion validity, Construct validity and Factorial validity.

- There are various methods of estimating the validity of a test. The important methods include: Correlation Coefficient method, Cross-validation method, Expectancy table method, Factor Analysis and the like.



6.8 FURTHER READING

1) Anastasi, Anne (1982): Psychological testing, Edition 5, illustrated. Macmillan, The University of Michigan
2) Berg, Kris E. and Latin, Richard Wayne (2008): Essentials of Research Methods in Health, Physical Education, Exercise Science, and Recreation, Lippincott Williams & Wilkins
3) Ebel, R.L. and Frisbel, D.A. (1979): Essential of Educational Measurement, 3rd Edition, Prentice Hall of India, New Delhi
4) Gupta, S.P. (2009): Modern Measurement and Evaluations with Statistics, Sharda Pustak Bhavan, Allahabad
5) Kelley, T. L. (1927): Interpretation of educational measurements, Macmillan, New York
6) Newton, Paul & Shaw, Stuart (2014): Validity in Educational and Psychological Assessment, Sage Publications, London.
7) Safrit, Margaret J. And Wood, Terry M. (1995): Introduction to Measurement in Physical Education and Exercise Science, Mosby, Missouri
8) Stanley & Hopkins (1966): Education and Psychological Measurement and Evaluation, Prentice Hall Pvt. Ltd. Eaglewood Cliff, New Jersey



6.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Validity is the most important technical criteria of a good test or other measuring instrument. Validity refers to the truthfulness of a test. Every test is constructed with a particular purpose of measuring a mental trait or criterion. Validity of a test is concerned with the purpose

of the test and it is the extent to which the test measures what it purports to measure.

Ans to Q No 2: Face validity is the simplest type of validity of a test. Face validity is not validity in actual psychometric meaning. It is a superficial or extra quality of a test and it refers only to what the test 'appears to measure' and not to what it actually measures.

Ans to Q No 3: Two factors that affect the validity of a test are: (i) the validity of a test largely depends on the items included in the test. In fact, validity of the test depends on how well each item included in the test could achieve the objectives of the test; (ii) Validity of a test also depends on the nature of the sampling group on which the test is administered for estimating validity. Validity is always specific to a particular group of individuals.

Ans to Q No 4: The two different types of criterion related validity are: (i) predictive validity, and (ii) Concurrent validity.

Ans to Q No 5: Construct validity is the extent to which a test measures the theoretical construct as defined by the test maker. According to Anastasi, construct validity is the extent to which the test may be said to measure a theoretical construct or trait or psychological variable.



6.10 MODEL QUESTIONS

Short Questions (Answer each question in around 150 words)

Q 1: What is the importance of validity in test construction?

Q 2: Define content validity of a test.

Q 3: What is construct validity of a test?

Long Questions. (Answer each question in around 300-500 words)

Q 1: Discuss the factors that affect the validity of a test.

Q 2: What do you mean by criterion validity? Discuss the different types of criterion related validity.

Q 3: Explain the different methods of estimating validity of a test.

*** ***** ***