

CHAPTER-III

PROCEDURE OF TEST CONSTRUCTION

CHAPTER · III

PROCEDURE OF TEST CONSTRUCTION

- 3.1 INTRODUCTION
- 3.2 MEANING OF PSYCHOLOGICAL TEST
- 3.3 CHARACTERISTICS OF GOOD TEST
- 3.4 STEPS IN TEST CONSTRUCTION
- 3.5 PLANNING THE TEST
- 3.6 PREPARATION OF FIRST DRAFT
- 3.7 ITEM EVALUATION
- 3.8 TRYING-OUT THE TEST
- 3.9 ITEM ANALYSIS
 - 3.9.1 ITEM DIFFICULTY
 - 3.9.2 INDICES OF ITEM VALIDITY
- 3.10 PREPARATION OF FINAL DRAFT
- 3.11 STANDARDIZATION
- 3.12 RELIABILITY
 - 3.12.1 MEANING
 - 3.12.2 TYPES
- 3.13 VALIDITY
 - 3.13.1 MEANING
 - 3.13.2 TYPES
- 3.14 CONCLUDING REMARKS

3.1 INTRODUCTION :

In the previous chapter researcher has taken a brief account of concepts of emotional intelligence , working of emotional brain and discussed the 10 competencies concerning the research topic.

This chapter deals with the test construction and standardization in detail.

3.2 MEANING OF PSYCHOLOGICAL TEST :

According to Best ,J.W. (1982,P.192), “As data gathering devices, psychological tests are among the most experimental and descriptive studies in education .”

A dictionary definition of the verb ‘ to test’ states that it means the subjection to conditions that show the real character of a person or thing in a certain particulars.

It has been also stated that a test is a series of questions or exercises or other means of measuring the skills, knowledge , intelligence or aptitude of an individual or group.

According to Freeman, F.S. (1971,P.46), “A dictionary of psychological terms defines a psychological tests thus ,A set of standardized or controlled occasions for response presented to an individual with design to elicit a representative sample of his behavior when meeting a given kind of environmental demand . It is now

when meeting a given kind of environmental demand... It is now common usage to include as a test any set of situations or occasions that elicit a characteristic way of acting , whether or not a task , and whether or not characteristics of the individuals best performance. Thus even a self-inventory or an attitude survey is called a test. ”

Freeman, F.S. (1971,P.46) defines ,“A psychological test as a standardized instrument designed to measure objectively one or more aspects of a total personality by means of samples of verbal or non-verbal responses or by means of other behaviour. ”

Anastasi, Anne (1982, P.22), “A psychological test is essentially an objective and standardized measure of a sample behaviour.”

Chronbach ,L.J. (1970 ,P .90) defines it in the following words , “A test is a systematic procedure for comparing the behaviour of two or more persons .”

Munn says that , “Test is an examination to reveal the relative standing of an individual in the group with respect to intelligence, personality , aptitude or achievement.”

Agarwal , R.N.(1964, P.75)

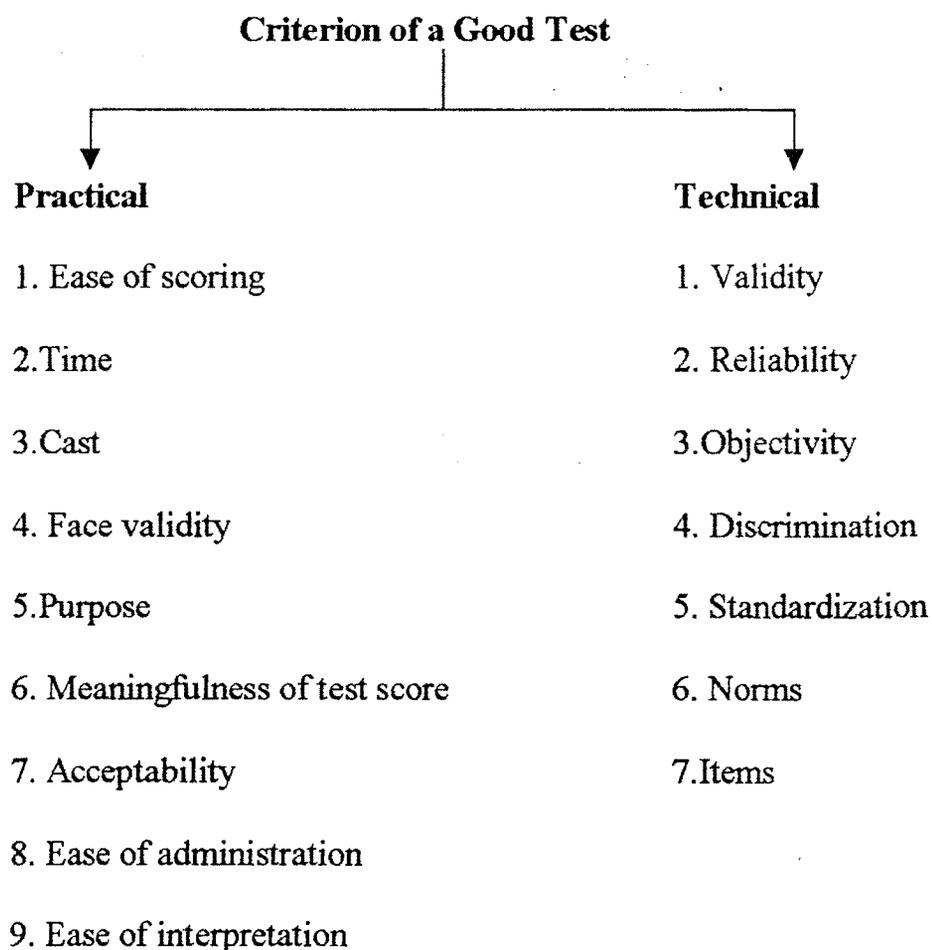
Hence , A psychological test is an instrument designed to describe the measure the sample of certain aspects of human behavior

and test may be used to compare the behavior of two or more persons at a particular time or one or more persons at different times.

3.3 CHARACTERISTICS OF GOOD TEST :

According to Asthana, B. and Agarwal, R.N.(1991,P.75) characteristics of a good test can be classified into two following categories.

1. Practical Criteria
2. Technical Criteria



3.4 STEPS IN TEST CONSTRUCTION :

Construction of the test has to be methodical . Asthana , B. and Agrawal ,R.N. (1991, P.40) given some important steps in test construction are as follows.

1. Planning the test- Objectivity
2. First draft of the test –Item selection
3. Trying –out the test
4. Final drafting
5. Standardization.

3.5 PLANNING THE TEST :

The first step in the development of a standardized test is the preparation of a plan. The plan will vary, depending upon the type of test that an author is preparing. But every plan should contain a statement of the purpose or purposes of the test and some outline of the content of the test.

The purpose of the plan is to present the rationale for the test and to guide the preparation and evaluation of items to be used in the test. The purpose also serves as a general guideline for a potential user as he considers the quality of a test and how well the purpose was achieved.

3.6 PREPARATION OF FIRST DRAFT :

After the preparation of a test plan, the next step is writing and item evaluation. Item writing is the preparation of the test itself.

If poor items are prepared or if the items are not related to test purpose , we can not meet the test objectives .According to Asthana , B. and Agrawal ,R.N. (1991, P.42) , In item writing the following suggestions are taken into consideration.

1. The number of items in the preliminary draft should be more than that in the final draft.
2. The items should be clearly phrased so that their content and not their form, determines the response.
3. The test should be comprehensive enough.
4. No item should be such that it could be replied by referring to any other item, or a group of items.
5. Each item should carry equal marks.
6. The wording of items should be such that the whole content determines the answer and not a part of it.

So in item writing every precaution has to be taken to ensure that items are valid , appropriate and unambiguous.

3.7 ITEM EVALUATION :

Item evaluation is the process of judging the adequacy of test items to fulfill the designated purpose of a test.

According to Womer, F.B.(1968, P.75)

It takes two forms –

1. Subject judgement
2. Statistical judgement

The subjective judgements may be made by test specialist or by subject matter specialist or by both.

The test specialists look for ambiguities of wording or for special clues that lead to the answer for item characteristics unrelated to the skill or ability or other attributes that the item is supposed to measure that may influence the way an examinee answers a question. Subject matter evaluators also look for potential ambiguities and also judge whether an item seems to be measuring a knowledge or skill that is related to the purpose of the test.

Statistical evaluation of the test items are aimed at the same goal as judgmental evaluations by specialists at reducing ambiguities and maximizing measurement of the attribute that the test is designed to measure . In statistical evaluation the difficulty level and

discrimination power of items are calculated and this is done through the process of item analysis. After evaluation, the first draft of the test is prepared.

3.8 TRYING OUT THE TEST :

In this step, first draft of the test should be administered on a sample who are representative . Conditions of administration should be normal. By using the scoring key scoring should be done. Then the next step , item analysis , is to be taken .

3.9 ITEM ANALYSIS :

The item analysis of a test gives two kinds of information. It provides index of item difficulty and an index of validity.

3.9.1 ITEM DIFFICULTY :

Item difficulty should be determined for administrative as well as statistical reasons. It may , for example , be good policy to put a few easy items at the beginning of a test as a means of relieving a test taker's anxiety. There are also technical reasons for concern over item difficulty. Item difficulty level influences the shape of total score distribution and at least theoretically , validity and reliability.

3.9.2 INDICES OF ITEM VALIDITY :

The typical item analysis of a test of ability yields two kinds of information. It provides an index of item difficulty and an index of validity.

Where the term validity is used in a very broad sense.

The first major objective of an item analysis is to obtain objective information concerning the items we wrote for the test.

The information is valuable for several reasons.

- 1) It provides the opportunity to check up the test writers subjective judgement in selecting the items to compose the test.
- 2) By experience of such checking the test writer learns to improve in his art.
- 3) He learns how examinees react to items in general and to the items of each test in particular.
- 4) In multiple -choice tests, he learns which distracters (wrong answers) or misleads are not functioning as shown by their relative unpopularity.
- 5) He gains new insights into the kind of item that does best in this kind of test and thinks of new hypothesis concerning the nature of the ability being measured.
- 6) He learns where and how items need to be rewritten.

The most common use of the item analysis data is in the selection of best items to compose the final test form.

Types of indices of item validity :

According to Gilford ,J.P.(1954, P.417) , there are numerous indices and procedures for determining item validity. The more common ones fall roughly into four groups.

1. One approach uses a measure of precision ,in line with the theory that the probability of passing an item is an ogive function of ability.
2. Second approach stresses the numbers of discrimination of the desired sort that the item is capable of making.

It emphasizes the extent to which the item predicts segregation of examinees in to those with high versus those with low criterion scores.

3. The third approach, which co-relates the item with the criterion score in some way, is probably the most popular one.
4. The fourth approach is by way of analysis of variance.

3.10 PREPARATION OF FINAL DRAFT :

After doing item analysis, items with good discriminating value may be taken into the final draft and other items may be eliminated. Time of test is decided on the basis of administration.

3.11 STANDARDIZATION :

Standardization in testing has two clearly distinct meanings. In its first sense the term refers to standardization of procedure. Instructions are developed so that they will be as clear as possible to the population to be tested. Definite time limit may be established , sequence of the items is standardized and all other factors which may influence the test performance are kept from varying to an excessive degree. This kind of standardization is vital in any testing programme. Without it, the tests can not hope to provide reliable scores. the second meaning refers to a standardized interpretation of scores – the development of so-called norms groups . Norms may be of many types such as age norms , sex norms, grade norms etc.

3.12 RELIABILITY :

3.12.1 MEANING :

A Measurement procedure is reliable to the extent , to which it provides constant results on repeated measurements. Test reliability tells that to what extent individual differences of scores can be assigned to chance errors.

In the words of Womer, F.B. (1968, P.30) , “Test reliability may be defined as the degree to which errors of measurement influence test scores.”

Guion ,R.M. (1965, P.30) reliability as “the extent to which a set of measurements is free from random – error variance ”

In words of Anastasi ,Anne (1982 , P.102), “ Reliability refers to the consistency of scores obtained by the same individuals when re-examined with the same test on different occasions or with different sets of equivalent items or under other variable examining conditions.”

In other words of Garrett, H.E.(1967,P.103) “ The reliability of the test or of any measuring instrument depends upon the consistency with which it gauges the ability to whom it is applied.

Guilford, J.P.(1954, P.104),“reliability is the proportion of the true variance in obtained test scores.”

Reliability of a test is generally expressed in terms of reliability coefficient .

3.12.2 TYPES OF RELIABILITY COEFFICIENTS :

Generally four methods are used for this purpose. These methods are,

1. Test-Retest Method
2. Parallel Test Method

3. Split –Half Method
4. Method of Rational Equivalence.

Split –Half Method :

The most widely used procedure for estimating reliability from single testing is the Split –Half Method. In this method , the test is first divided into two comparable halves and the correlation is found between the two scores obtained for each individual on the two equivalent halves. For splitting the test into two equivalent halves , the odd-even split is quite common in use .

This method is used , when it is not feasible to construct parallel forms of the test nor advisable to repeat the test itself.

3.13 VALIDITY :

3.13.1 MEANING :

Validity of a test means its truthfulness. If a test measures what it intends to measure, then it is said to be valid. The validity of a test is determined by measuring the extent to which it matches with a given criterion.

Gullicsen ,H. (1967 ,P.88) has defined validity as “ the correlation of the test with some criterion.”

Cronbach ,L.J. (1970, P.122) says that the validity may be determined by showing , “that a test covers to be measured , or it

may be established inductively by naming the traits represented in the items at hand.”

Freeman ,F.S. (1971 ,P.88) puts it , “ The first necessary condition of a valid test is that it has an adequate degree of reliability .If the reliability coefficient of a test is ‘zero ’, it can not correlates poorly even with itself can not correlate well with a measure of another Variable.

According to Anastasi, Anne (1982 ,P.131) “ The question of test validity concerns what the test measures and how well it does so ”

3.13.2 TYPES OF VALIDITY :

According to Agarwal ,A.N (1964 ,P.40) ,Cronbach has elaborated two chief, main major methods of determining validity

I) Logical II) Empirical

Freeman suggests four types of validity

1. Operational 2. Functional 3. Factorial 4. Face.

Jordan has categorized validity into I) Internal and II) External

According to , Asthana , B. and Agarwal ,R.N. (1991, P.94) ,types of validity are as follows,

TABLE 1

Types of validity

No	Types of validity	Meaning	Method
1	Curricular or content validity	The test should represent the curriculum or content	Content Analysis and on its basis proper sampling and weightage of items.
2	Logical validity	The test should be concerned with relevant items	Purpose of the test should be kept in mind. Say, it intends to measure units or problem solving.
3	Face validity	The test should appear to measure what it intends to measure.	Judge the items.
4	Concerned validity	The test scores should give an estimate of some present performance.	The test should be given and its score obtained. Its correlation should be calculated with some other performance.

3.14 CONCLUDING REMARKS :

In this chapter the researcher has described the procedure of test construction in detail.

In the next chapter, the researcher has taken a brief review of related literature.