PART II

* Test Design
* Test Construction
CHAPTER NO. IV
TEST DESIGN

Planning is an essential activity in all stages of a test construction project. Planned procedures for test construction vary, depending upon a number of factors like the nature and type of a test, its purpose, its use, the type of trait or ability to be measured, etc. Inadequate planning in the initial stage may a time leads to a number of difficulties later on. K.W. Vaughn of the Cooperative Test Service, while referring to test-design, writes,

"Planning encompasses all the many and varied operations that go into producing a test. Not only does it involve the preparation of an outline or a table specifying the content or operations to be covered by the test, but it must also involve careful attention to item difficulty, to types of items, to arrangement for tryouts."

In designing the present test, the investigator considered all the factors and varied operations that are

associated with the preparation of a good test. The design of the present test has been laid down bearing in mind the criticism levelled against the existing tests of scientific aptitude.

CARDINAL FACTORS OF A TEST DESIGN:

Some of the cardinal points that are essential in the design of a test construction project are:

(1) What is it that the test is designed to measure?

(2) Why is the test planned? - Purpose of the test.

(3) What is to be the nature or the type of the test?

(4) How are the sub-tests to be selected?

(5) How is the sample of standardisation to be decided?

(6) How will the individual test item be selected?
   a. Against what external criterion the test will be validated?
   b. How will the item validity be determined?
      What techniques will be used for finding the internal consistency and the difficulty level of item.

(7) How will the test be standardized? What types of norms will the test have? How will the reliability and validity of the test be estimated?

These are the major issues involved in designing a good test. Each one of these is discussed below:
1. What is to be measured?

The first requisite for a good test is to have a clarity about the trait or the ability to be measured. The present test is meant to be a valid and reliable tool to measure pupils' aptitude for science. The meaning of 'Scientific aptitude' has been discussed and clarified in Ch.II. The definition of the term 'Aptitude' has already been accepted as the one given by W.V. Bingham.

2. Purpose of the test:

The second important factor of a test design is to clarify the purpose of the test. The main purpose of the present test is to provide a valid tool to measure the aptitude for science of secondary school leaving pupils of Gujarat. Based on this test, one can predict with a reasonable degree of confidence the pupils' ability to pursue science courses in higher studies. The test is aimed to be used for selection of pupils at the University entrance stage and for guidance to pupils passing the Secondary School Certificate Examination at the Secondary School stage.

3. Type of the test:

The third important factor of a test design is to decide the type of the test. Is it going to be a speed test or a power test? According to Anastasi, "No useful purpose is served by making aptitude test a speed test." 2

Freeman also supports this, "The principle criticism of most of the aptitude tests is that they place too much weight on rate of work, thus adversely affecting their predictive value." The present test is meant to measure the scientific aptitude of pupils and hence it is to be a power test and liberal time will be given to the pupils.

4. Selecting the sub-tests:

The next important consideration in test design is to decide the procedure to be adopted for the selection of the sub-tests. As the present test is to be a test of aptitude the investigator must be clear about the concept of "aptitude" and has to decide the type of the sub-tests to be included in the test battery on the basis of the previous work done in the field. The review of the past work has not been able to give any useful evidence about the existence of some factors in scientific aptitude. Hence the selection of the sub-tests to be included in the present test will be based upon a new independent approach involving the following steps:

(a) Study of the sub-tests included in the test batteries constructed in the past.

(b) Analysis of the abilities of successful scientists.

(c) Ranking of these abilities by experts.

d) Selecting the abilities which have high ranks and

e) Basing the sub-tests on these abilities.

5. Sample:

The sample selected for the standardization of a test must be sufficiently large and thoroughly representative of the parent population. For the present work, the sample will be selected from the population of the S.S.C. Class pupils of the Secondary Schools of Gujarat on the basis of the following criteria of stratifications:

1. Rural and urban
2. Boys and girls and
3. Different regions.

The test will be administered to about 400 pupils for the tryout while it will be administered to about 1500 pupils during the final run.

6. Item Selection:

Item selection is the most indispensable procedure for constructing a good test. An elaborate procedure of item selection will be adopted for the present work. The following will be the chief steps for item selection:
(a) Fixing the criterion for item validation:
Two contrasted criterion groups will be obtained on the basis of (i) teacher's ratings for 'participation in science clubs of schools' and (ii) studying achievement of pupils in science subjects. One group will consist of pupils with high criterion score while the second group will consist of pupils with low criterion score.

(b) Item validation:
Each item will be validated against these two criterion groups. Only the items which will discriminate significantly between the two criterion groups will be included in the present test battery. The chi-square test will be applied to test the significance.

(c) Internal consistency:
The well-known technique of finding out the item-total test correlation will be used for determining the internal consistency of the test. In the present work, the two extreme groups will consist of the upper and the lower 27 per cent of the sample and Flanagan's table will be used for determining the internal consistency of the items.
(d) Item difficulty:

The item difficulty will be computed from the two extreme groups by applying the following formula:

\[ D = \frac{U + L}{2} \]

Where, 

- \( D \) = Difficulty value
- \( U \) = Percentage of correct responses to an item from the upper 27 percent group.
- \( L \) = Percentage of correct responses to an item from the lower 27 percent group.

These percentages will be corrected for guessing.

7. Standardization of the test:

The test design must clearly mention the procedures for determining the norms and the methods to establish the reliability and validity of the tests.

(a) Fixing the norms:

The following types of norms will be fixed for the present test:

(i) Percentile scores
(ii) Z - Scores
(iii) T - Scores and
(iv) Letter grades.
(b) Reliability:

The reliability of the present test will be determined by the following techniques:

(i) Split-half method,
(ii) Test-retest method,
(iii) Use of K - R formula and
(iv) Analysis of variance technique.

(c) Validity:

The validity of the following types will be established for the present test:

(i) Content validity
(ii) Concurrent validity,
(iii) Construct validity and
(iv) Predictive validity.

The test will also be cross-validated against new independent samples.

(d) Factor analysis:

A factorial analysis of the test will be carried out on a sample of about 100 S.S.C.Class pupils by applying Thurstone's Centroid method.

A full picture of the detailed procedure followed in the construction and standardization of the present test is given in the following chapters.
SELECTED REFERENCES


