PART III

* Criterion
* Item Selection
* Final Run of the Test
CHAPTER VI

CRITERION

The present work is a departure from the usual procedure adopted by test constructors in so far as the validating criterion is concerned. In the present work, care has been sought from the beginning to develop a valid test. It is here that the present approach to test construction differs from the usual approach. The development of a good criterion for validating test items initially and subsequently validating the test as a whole forms an important aspect in the present work.

WHAT IS A CRITERION?

Criterion is that factor with which a procedure and/or instrument is compared to determine its validity. For example, if a medical aptitude test is to be used in selecting promising applicants for medical school, ultimate success in medical school will be the criterion.

IMPORTANCE OF THE CRITERION:

Any programme of research, specially in the field of test construction, implies the need to obtain satisfactory
criterion against which the test can be validated. Until some solution of criterion problem is reached, research in test construction can hardly proceed. The proper interpretation of a validity coefficient requires information regarding the specific criterion against which the test is validated.

It has been a sad experience of many test constructors to realize in the end that the test they had constructed measured something else than what they intended to measure. The main reasons for such a state of affairs are lack of reliable criteria and a proper technique for selecting the items. It is necessary that each item should be validated against some criterion before it is included in the test. This requires a predetermination of some criterion before the work of standardization of a test is undertaken. This demands that sufficient work should be undertaken to select a criterion which is both valid and reliable. According to Travers,

"It might be profitable to devote time to the study of the criterion than to proliferation of new tests which are somehow hoped will be more valid than previous ones."1

IMMEDIATE, INTERMEDIATE AND ULTIMATE CRITERIA:

The ultimate criterion is some appraisal of the man's lifetime success in his profession; e.g. the ultimate

criterion for the medical aptitude test would be the eventual achievement as a practising physician. Obviously, it would require a long time for such criterion data to mature. Moreover, it is doubtful whether a truly ultimate criterion is ever obtained in actual practice. Even if such an ultimate criterion is available, it will probably be subject to many uncontrolled factors which would render it relatively useless. As stated by Ross,

"Most 'ultimate' criterial measures are hard to get, all too unreliable, and of doubtful relevance." 2

Thus it is most advisable to employ the immediate or intermediate criteria which are nothing but various approximations of ultimate criteria. For example, the intermediate criteria for the medical aptitude test is success or failure in the medical school; while the immediate criteria for the same would be the performance record of the pupil at some stage of training in medical subjects.

**HOW TO SELECT A GOOD CRITERION?**

The immediate and intermediate criteria are the substitutes of ultimate criterion which are only partial and are never completely satisfactory. It is essential to choose the most satisfactory from amongst the measures that appear feasible to obtain. Thus the problem to be faced is to decide which of several criterion measures is most

---

satisfactory. To arrive at this decision, the four main qualities of a good criterion need attention.

These qualities are: (1) relevance, (2) freedom from bias, (3) reliability and (4) availability.

DIFFERENT CRITERION MEASURES:

There are many criterion measures that might be obtained and used for validating a test. The criterion measures which are widely used are: (1) Ratings, (2) performance records and (3) contrasted groups.

Ratings have been employed in the validation of almost every type of test. If ratings are obtained from trained raters under carefully controlled conditions, they can provide a valuable source of criterion data. It is generally desirable to secure independent ratings from more than one observer in order to rule out individual bias of the rater. Moreover the rater should have the opportunity to observe the individual in situations in which the particular trait is manifested.

In the development of special aptitude tests, a frequent type of criterion is based upon performance in a course of specialized training. For example, commercial aptitude test may be validated against performance records of the trainee in commercial subjects e.g. book-keeping, accountancy, stenography, typing etc. These performance
records may be the result of the academic examinations or achievement tests.

The method of contrasted groups generally involves a composite criterion which reflects the cumulative and uncontrolled selective influences of everyday life. In this method, the validity is checked by comparing the scores of two contrasted groups—one consisting of subjects with a high degree of the ability in question and the second consisting of subjects with a low degree of that ability. The contrasted groups may be selected with the use of ratings and/or performance records.

The measure of contrasted groups seems to be most reliable as it includes the use of both the previous measures with the elimination of individual bias. Hence the investigator has preferred to employ the last measure for the present work.

SELECTION OF CRITERION GROUPS:

In the present work, the criterion groups are selected by using: (1) the technique of rating by teachers and (2) achievements of the pupils in General Science in Class-rooms.

The major steps involved in the selection of the criterion groups are as under:

1. Rating on five-point-scale of pupils on 'participation in science clubs of schools' on the following
main traits by two senior science teachers:

(a) experimental bent and
(b) keenness of observation.

2. Studying the achievement of the pupils in General Science at the terminal examination.

3. Studying the performance of the pupils on a standardized achievement test in General Science.

4. Selecting two criterion groups - one group consisting of pupils who score consistently high in all the items mentioned above and a second group of pupils who score consistently low in all these items.

Step 1:

The investigator selected five secondary schools of Baroda District and requested two trained and experienced science teachers (of each school) who were in charge of science clubs to rate the pupils of the S.S.C. Class in their school on a five-point-scale on the following two main traits:

(a) Experimental bent and
(b) keenness of observation.

The rating scale developed for this is given in Appendix B (vide Page No. 174)
Step 2 and 3:

The investigator obtained the marks secured by the above pupils in General Science at the terminal examination. Then the investigator administered them "Achievement test in General Science for Std.X" published by orient Longmans, Bombay.

The 'Achievement Test in General Science for Std.X', which is used for the present work is a power test standardized on a sample of about 1000 pupils from different schools of Gujarat. The reliability of the test is reported sufficiently high.

Step 4:

The investigator, then, selected two criterion groups (contrasted groups) as follows:

(a) First group: 50 pupils securing consistently 'A' or 'B' grades in Step 1 by both the teachers and securing more than 50 per cent marks at step No. 2 and 3.

(b) Second group: 50 pupils securing consistently low grades in all steps.
SELECTED REFERENCES


Journals: