

CHAPTER V

PLAN OF THE INVESTIGATION

WHY SHOULD WE PLAN OUR WORK

Even an illiterate villager in this country is familiar with the word 'planning'. Everyday, he hears something of 'five-year planning' or 'family planning'. Not only that he knows the meaning of 'planning' but he also knows the tremendous importance of planning any type of activity.

We have already completed the 'Second Five Year Plan'. For over a decade, through the two Five Year Plans, India has endeavoured to harness natural resources and the energies of the people to the task of national development. Within this short period of only ten years, India has achieved marvellous successes in economic, social and many other fields of national life. This was made possible only because the projects had been previously planned. Thus the two plans have proved, without doubt, that through accurate planning one can reach one's goal speedily and correctly.

It is only for the last forty or fifty years that we have started 'planning' on national basis. But the teacher has been planning his class-room lessons right from the days of Pestalozzi and Herbert. 'Planning' is not new to him. He knows that he fails miserably in a class-room whenever he has

not planned his lesson properly. He also knows how successfully and easily he can teach a previously planned lesson unit.

As such a test constructor in Educational measurement or in Mental measurement should understand the importance of planning the research project and also that:

Sound methods of appraisal and research require careful planning, not only as a means of ensuring accurate results but of making it possible for others to repeat an investigation for the purpose of corroboration or refutation. ¹

As K. W. Vaughn² says,

Planning is an essential activity in all stages of a test construction project. Inattention to planning may result in a failure to meet production deadlines, or may necessitate the use of uneconomical procedures or of below-standard materials in order to meet those deadlines.

Thus to avoid waste of time, waste of money and possible frustration that may arise out of failure to achieve desired results, previous planning of the whole research project is very essential. Lack of proper planning may lead the investigator to innumerable and insurmountable difficulties in all stages afterwards. A need of planning a research

1 Barr, A.S., Davis, R.A., & Johnson P.O., 'Educational Research and Appraisal', J.B.Lippincott Company, Chicago - Philadelphia - New York, 1953, Chap.I, p.11.

2 Lindquist, E.F., Editor, 'Educational Measurement', American Council on Education, Washington, D.C., 1959, Ch. VI, p. 159.

project, is therefore, unquestionable.

What are, then, the operations in, and aspects of, test construction that demand planning? To answer this question, we shall again quote Vaughn.³ He says:

Test planning encompasses all of the many and varied operations that go into producing a test. Not only does it involve the preparation of an outline or 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 directions to examiner, to arrangements for try-out, to problems of test reproduction, to provision for expert review, to the provision of adequate equipment and facilities, to the procurement of personnel, and so forth.

The major concern of the chapter lies not with the details of how the various difficulties involved in test constructions are to be met, but rather with the need for anticipating these difficulties before they arise, of co-ordinating and tying together the various operations involved, and of ensuring smooth and efficient administration of the project as a whole.

**POINTS TO BE KEPT IN VIEW WHILE PLANNING
THE CONSTRUCTION OF A TEST**

A layman cannot and should not try his hand on construction of a test in mental measurement. If he dares to do so he fails miserably in the long run and even if he produces

³ Op.Cit., p. 159.

something, this is invalid and of no use. The man who thinks of constructing a test, must have a thorough knowledge of the principles of test construction. He must have thoroughly digested all the basic principles of test construction. For this, he must have studied all the literature available on techniques of test construction.

When one sets out to construct a test, there are several considerations that will determine one's approach and operations. It will depend upon one's philosophy of testing in general and of testing in the particular field and upon the use of the test to be produced. The test constructor, as such, will have to ask some questions to himself before he starts his work.

The first question he should ask, is, "Why am I constructing this test?" He should be crystal clear of the kind of test, whether it is an aptitude test or an achievement test. He must define the definite objectives he would achieve through the test. He must fix up the traits of human beings he wants to measure; and see whether it will be possible for him to measure only those traits through the test he is to construct.

The second question he should ask, is, "What is the purpose of the test?" He must be quite sure of the usefulness of the test. Why is he inclined to measure these traits of human beings? How will the results obtained be used?

The test constructor should have in his mind the section of the community for which the test is to be constructed. This will help him at a later stage in selecting a fairly representative sample for testing programme.

Any long run project of test construction will test the ability, patience and tenacity of the test constructor. It may happen that unforeseen difficulties may arise while the test construction is in progress and that much more time will be required than formerly stipulated. He should have a clear idea of the time that will be required to carry out the project successfully. He must be ready to work on the project for a fairly longer time than what he had stipulated in the beginning. He must also have fairly correct estimation of the cost that he will be required to bear for carrying out the whole project. It may happen that the project becomes much expensive at a longer run and that the investigator has to give up the whole project for lack of finances. This will be a very unhappy situation and it will lead to a huge waste of human labour, time and money. To avoid such a situation, the test constructor must ask himself whether he will be able to spend as much time and money as required for carrying out the project. If he is confident of this, he may start his work.

WHAT THE TEST IS GOING TO MEASURE

The present test is designed to measure a person's aptitude for teaching. The first requisite of a good test is

that it must measure what it is meant to measure. This is possible only when the test constructor is quite clear of the term aptitude. He must define the term clearly and stick to it consistently throughout the work. The meaning of the term aptitude has been thoroughly discussed in chapter III. The same definition of the term 'aptitude' is accepted as far as the present work is concerned. To repeat, the definition of the term aptitude, accepted for the present treatise is, "Aptitude is a condition symptomatic of a person's relative fitness, of which one essential aspect is his readiness to acquire proficiency - his potential ability - and another is his readiness to develop an interest in exercising that ability."

DEFINING THE PURPOSE OF THE TEST

For any type of test, there should be available a general statement of the test purpose. Such a general definition of objectives should specify what is to be measured, who is to be tested, and what uses are to be made of the test scores.

The present test is to measure an aptitude for teaching in secondary schools. This test is to be administered to the university graduates who aspire to be teachers in secondary schools in Maharashtra and Gujarat divisions of the old Bombay State. No doubt the sample for standardization of the test was selected from Maharashtra and Gujarat but the test was constructed on such a basis as it could be administered to

university graduates in any State in India. The resulting scores are to be used in predicting the aptitude for teaching in the secondary schools. Thus the test will be useful in forecasting teaching ability and so in selecting the personnel for training in the teaching profession. The school authorities in this country have to employ untrained teachers in their schools as there is a great dearth of trained teachers. This test will be useful to them in selecting the right persons with an aptitude for teaching.

As the test is designed to predict an aptitude for teaching, it can be used as a tool for vocational guidance also. If the test is administered to any college student and if it is found that he has an aptitude for teaching, he can be advised to select the teaching profession as his career.

Thus this test is to serve a double purpose of vocational selection and vocational guidance.

There is an advantage in defining the purpose of the test clearly. We are constructing an aptitude test. We shall need satisfactory criterion data for the validation of the test score. As such the general statement of the purpose of the test is to predict a particular criterion, and the final check on the test is its relationship with this criterion.

THE NATURE OF THE TEST

There can be two types of tests - either a power test

or a speed test. The technique of standardisation varies according as the test is a speed test or a power test. So it was necessary to decide in the beginning whether the test under construction was to be a power test or a speed test.

In the speed tests, various items are fairly uniform in difficulty and the score is dependent primarily upon speed. On the other hand, in the power tests, the different items are of varying difficulty and they are arranged in order of uniformly increasing difficulty but with no time limit. The main object of such a test is to find out how much difficult task the subject may successfully do if given sufficient time.

Diagnostic and aptitude tests are most generally power tests as no useful purpose is served by making them speed tests. The test under construction is an aptitude test and so it is a power test. There is no time limit set for administering the test.

After considering the nature, purpose etc. of a test under construction, we shall now turn to the actual planning of the test.

AN OUTLINE PLAN OF THE CONSTRUCTION AND STANDARDISATION OF "AN APTITUDE TEST FOR SECONDARY SCHOOL TEACHERS" IS GIVEN IN THE FOLLOWING PAGES

A CRITICAL STUDY OF THE EXISTING LITERATURE

Three types of literature were studied:

- (1) Literature on test construction.
- (2) Literature on "Aptitude testing" in general.
- & (3) Literature on "Aptitude test for teachers" in particular.

For this purpose all available books, magazines, journals and reviews were tapped and studied thoroughly.

JOB ANALYSIS

This consists of two parts.:

- (1) Job description.
- (2) Job analysis.

Job description: This we have discussed at length in chapter II.

Job analysis: We are primarily interested in job analysis as a source of hypothesis for selection of tests and as a source of insight about criterion measures.

The job analyst should interpret the job in terms of attributes necessary in a worker. What qualities are called for if the worker is to be successful? A comprehensive answer to this question provides us with a set of hypothesis and a series of proposals for test construction.

The following sources were tapped for job analysis.:

- (1) Experience of the Training College Teachers &

Education Department personnel.

- (2) Experience of the Secondary School Principals.
- (3) Interviews with an interrogations of successful teachers.
- (4) Interviews with and interrogations of unsuccessful teachers.
- (5) Direct experience of the test constructor.

NATURE OF SUB-TESTS

For deciding the nature of sub-tests, the following sources were used:

- (1) Results of the job analysis.
- (2) The existing pool of testing instruments.

As it is already discussed in chapter IV, no "aptitude test for teachers" is available in India. No foreign test also can be used in toto as the conditions - both cultural and educational are different. All the sub-tests were, therefore, constructed anew to suit the conditions in India. While constructing the sub-tests, the foreign material was used to guide the selection of test items.

In the beginning, a number of sub-tests were constructed and tried out on a small sample on experimental basis. Then, a fixed number of sub-tests were selected and

used for the first try-out.

WRITING THE ITEMS FOR EACH SUB-TEST

Following steps were followed:

- (1) Similar tests that are available were studied critically.
- (2) Literature on the topic was studied.
- (3) The forms of the test items were decided.
- (4) A number of items for each sub-test were prepared.
- (5) These items were discussed critically with experts for content validity. The items on which there was no consensus of opinion about the content validity were eliminated.
- (6) The correct answers were discussed with the experts and finalised. They were rediscussed and subjected to change when it was found after first tryout that there was great divergence between the answers decided and the testees' responses.
- (7) The mode of scoring each item was finalised.
- (8) Instructions to be given to the testees were framed.
- (9) Maximum time to answer all the items, also, was

thought of.

THE CRITERION AND CRITERION GROUPS

The test is designed to forecast teaching ability. The predictive validity of the test score, therefore, was to be found out. For this some external criterion was necessary. This was the total score obtained by the testee in the university B.Ed. examination. If the test is to be highly valid, the criterion score must be highly reliable and valid.

To judge the validity of the university examination results, the opinions of the secondary school principals were sought. Their opinions were collected mainly on the question, 'Do the trained teachers work more skillfully and successfully than the untrained ones?'

The task of selecting a reliable and valid criterion is a much difficult one. If it is done successfully, half the battle of constructing a good valid test is won at the first lap.

From the sample selected for the first tryout two criterion groups for validating each item of each sub-test were selected. The outline of the procedure for selecting the criterion groups was as follows.:

- (1) Opinions of the college teachers about the probable success of a pupil-teacher as a school teacher were collected. For this internal

marking of the different college teachers on different aspects of activities of the pupil-teacher was obtained.

- (2) Total marks obtained by each pupil-teacher at the university examination were obtained.
- (3) The product-moment coefficient of correlation, 'r' between these two sets of marks was found out. The 'r' was significantly high. It proved to some extent, the validity of the university examination results.
- (4) Then, from the whole group, following three groups were formed.

Group I: Students getting HIGH score on internal marking and HIGH score on university examination.

Group II: Students getting HIGH score on internal marking and LOW score on university examination or vice-versa.

Group III: Students getting LOW score on internal marking and LOW score on university examination.

Group II was naturally discarded.

Then the principals of the schools, where these teachers from groups I and III were serving, were contacted and their opinion about the work of these teachers was sought. The

teachers about whom the principal's opinion was different from the forecast of the university marks, were discarded from the group. This procedure tested also the validity of the criterion.

In this way, the two criterion groups were selected finally.

TRYOUT TESTING

After the tryout test was ready, it was administered to a sample of 153 candidates seeking admission to the B.Ed. class in the Faculty of Education and Psychology, M.S. University of Baroda, Baroda. The data obtained were used in validating the test items against the two criterion groups formed as above. The items, that were not valid, were discarded from the test and the pilot test form was prepared.

PILOT TESTING AND ITEM ANALYSIS

The pilot test was administered to the student-teachers seeking admission to the B.Ed. degree course in the Faculty of Education & Psychology, Baroda, S.T.T. College, Bombay Tilak College of Education, Poona and S.N.D.T. B.T. College, Baroda.

Then, for item analysis, the following three procedures were adopted.:

(1) Item validation: This was followed at the first

tryout stage. Each item was validated against the two criterion groups. The responses given to each item in each sub-test by each group were studied to see whether the high group gave correct responses in greater number. The difference between the correct responses given for each item by each group was found out and its significance was studied. Only those items which showed a significant difference between the high and the low groups were selected. Others showing negative difference or insignificant difference were rejected.

- (2) Internal consistency: The two extreme groups were selected. The upper and the lower 27 per cent of the sample constituted the two extreme groups. The total score on a test was used as the criterion variable for evaluating each individual item in the test. The item-test correlation coefficients were computed using Flanagan Table.
- (3) Item difficulty: This, also, was calculated from the two extreme 27 per cent groups. The following formula was applied.

$$\text{Difficulty value} = \frac{U + L}{2}$$

Where

U = percentage of correct responses to an item from the upper 27 per cent group.

L = percentage of correct responses to the same item from the lower 27 per cent group.

These percentages were corrected for guessing.

These three procedures helped the investigator in screening out the defective items. The valid and correct items were finally selected to form the final test.

THE FINAL TEST

The final test is a printed one and in a small booklet form. A separate answer sheet is prepared. It also is a printed one.

SAMPLE FOR STANDARDISING THE TEST

The test was administered to the newly admitted trainees in the different training colleges of Maharashtra and Gujarat. This gave a sample of about six to seven hundred pupil-teachers. It was the most representative of the whole population as pupil-teachers from all the corners of the two divisions of the Bombay State came for training in these colleges.

ADMINISTERING THE FINAL TEST

By this time the instructions to be given to the

testees had been finalised. These were as few as possible, precise and concise.

There was no time limit. But the testees were advised to work as fast as possible.

The investigator himself carried out the field work and the test was administered at all centres under his direct supervision.

STATISTICAL ANALYSIS OF THE TEST DATA

The statistical analysis was applied to the data for the whole sample as well as to the data for each unit. Each training college formed a separate unit.

Different data were classified as under.:

- (1) Classification of the total test score of the whole sample.
- (2) Classifications of the total test scores of the sample for each unit.
- (3) Classifications of the total score of each sub-test of the whole sample.
- (4) Classifications of the total score of each sub-test of the sample for each unit.

Then, the means, medians and SD's of the total test

score and scores of each sub-test for the whole sample and for different unit-samples were calculated.

The means were calculated by the 'Assumed Mean' method i.e. by 'Short Method'.

The medians were calculated by the following method.:

$$\text{Mdn} = \mathcal{L} + \left(\frac{\frac{N}{2} - F}{f_m} \right) \times i$$

Where

\mathcal{L} = lower limit of the class-interval upon which the median lies.

$\frac{N}{2}$ = One half the total number of scores.

F = sum of the scores on all intervals below \mathcal{L} .

f_m = frequency (numbers of scores) within the interval upon which the median falls.

i = length of the class interval.

The SD's were calculated also by the 'Short Method'.

SE's of each mean, median and SD were calculated. Standard formulas in the books were used. Significance of each statistic was tested.

A. The nature of the distributions of the total test score and of total scores on each sub-test for the whole sample were judged by finding out skewness and kurtosis of each distribution.

Skewness was calculated with the help of the following two formulas:

$$1. \quad Sk = \frac{3 (\text{mean} - \text{median})}{\sigma}$$

$$2. \quad Sk = \frac{P_{90} + P_{10}}{2} - P_{50}$$

(The measures of skewness obtained by both the formulas were different because the two measures of skewness were computed from different reference values in the distribution, and hence were not directly comparable).

The SE of the skewness was calculated with the help of the following formula:

$$\sigma_{Sk} = \frac{.5185D}{\sqrt{N}}, \quad \text{where } D = (P_{90} - P_{10})$$

Significance of σ_{Sk} also was determined.

The following formula was used to calculate kurtosis.

$$Ku = \frac{Q}{(P_{90} - P_{10})}$$

The SE of kurtosis was calculated by the following formula:

$$\sigma_{Ku} = \frac{0.28}{\sqrt{N}}$$

The significance of σ_{Ku} was tested.

B. The nature of the distributions of the total test score of the whole sample as well as each sub-test score of the whole sample was determined by graphical representation of the scores.

The following graphs were drawn.:

- (1) Frequency polygon.
- (2) Histogram.
- (3) Smoothed frequency polygon.
- (4) Superimposition of an ideal normal curve on the obtained distribution.

C. The nature of the distribution of the test score was also tested by applying the chi-square test.

NORMS

Standard score norms, T-score norms and percentile norms were calculated. The percentile curve was drawn.

Letter grades were assigned to the raw scores and standard scores. A graphical representation of the same was given.

RELIABILITY

Reliability of the test was determined by three methods. The following techniques were used.:

- (1) Split-half method.

(2) The method of Rational Equivalence or 'Kuder-Richardson' method (formula-20).

(3) Method of Analysis of Variance (Hoyt's method).

A sample of 200 testees was selected out of the whole sample for determining the reliability of the test by split-half method, while the whole sample (N = 530) was used for the remaining two methods.

All the three methods given above were used to determine the reliability of the whole test score. The reliabilities of each sub-test score were determined only by split-half method. The reliabilities of all unit samples also were estimated by split-half method.

VALIDITY

Validity of the test was established by the following procedures.:

- (1) Selecting only valid items in the test.
- (2) Validating the test score against the criterion score (Predictive validity).
- (3) By factor analysis of the test.

Validity coefficients were calculated for total sample as well as for each unit sample.

FACTOR ANALYSIS

A factorial analysis of the test was carried out on the whole sample. For factor-analysis, Thurstone's Centroid method was used.

MULTIPLE CORRELATION AND MULTIPLE REGRESSION EQUATION

'The Wherry-Doolittle Test Selection Method' was applied for calculating Multiple, R and for deriving Multiple Regression Equation. This equation is useful in predicting the criterion score, i.e. with the help of the Multiple Regression Equation, a person's Aptitude for teaching can be predicted.