CHAPTER IV

PLANNING AND PREPARING THE TEST

Careful planning should preceed the actual process of constructing a test. Planned procedures for test construction vary depending upon a large number of factors such as the nature of the test, the type of ability measured and the purpose of the test etc. Planning is an essential activity in all stages of a test construction project. Lack of proper planning may lead to countless difficulties at all stages afterwards. As K.W.Vaughn aptly puts it:

> 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 the examiner, to arrangements for tryout, 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.[1]

This chapter provides a brief orientation to a number of problems that are more thoroughly discussed in

---

1 Lindquist, E.F., Educational Measurement, Washington D.C., American Council on Education, 1959, p. 159.

later chapters. The major concern of the chapter is the need for anticipating some of the difficulties before they arise. This will help the process of co-ordination of various operations involved, for smooth and efficient, administration of the project as a whole.

Study of Tests Available

First of all, some of the standardized non-verbal tests, both Indian and foreign, were critically studied. Below is given the list of such tests:

(1) Lorge Thorndike Intelligence Tests: Verbal and non-verbal batteries: Form - A.

(2) Otis Group Intelligence Scale: Primary examination: Form - A.

(3) Non-language Multi-mental Test by E.L.Terman, WM. A. McCall, Irving Lorge: Form - A and B.

(4) The South End Group Test of Intelligence: George Harrap & Company.

(5) Kuhlmann - Anderson Tests Grade I, I.Semester.

(6) Chicago Non-verbal Examination.

(7) Madras Non-verbal Test: St.Christopher Training College, Madras, Age 9-13.

(8) A Non-verbal Group Test of Intelligence with special reference to the Mysore State. (Age 7-13) by M. G. Premalatha.

(9) Revised Beta Examination by C.E.Kellogg and N.W.Morton.

Choice of the Sub-tests

Analytical study of the above tests was made with a view to finding out the types of sub-tests, the types of items for different sub-tests and the number of items included in them. An outline of the present test was drawn on the basis of this analysis.

TABLE 13

Analysis of Some Non-verbal Tests of Intelligence for Selection of Sub-Tests

No.	Name of the test	Sub-tests included	No.of items in the sub-tests
1	Otis Group Intelligence Scale Form B	Classification	16
		Substitution	40
		Missing part	13
		Maze test	1
		Progressive series	7

No.	Name of the test	Sub-tests included	No.of items in the sub-tests
2	Non Langugge Multi-mental Test: Terman, McCall, Irving Lorge: Teacher's Training College, Columbia.	Classification	60
3	The South End Group Test of Intelligence: George Harrap & Company.	Figure classification	7
		Analogies	6
4	Kuhlmann-Anderson Tests Grade I, I Semester	Progressive series	8
		Missing parts	9
		Verbal meaning test	54
		Classification	15
		Figure completion	8
		Motor tests	11
5	Chicago Non-verbal Examination - Age 7 adult; Time 40 mts. - by Andrew W.Brown	Substitution	150
		Classification	46
		Block design	4
		Figure completion	24
		Progressive series	16
		Absurdities	6
6	Madras Non-verbal Test. St.Christopher Training College,Madras: Age 9-13; Time: 23 minutes	Similar designs	14
		One difference numbers	15
		Number series	14

No.	Name of the test	Sub-tests included	No. of items in the sub-tests
		Classification	10
		Absurdities	24
		Progressive series	19
7	Revised Beta Examination by C.E.Kellogg and N.W.Morton	Maize test	3
		Marking the shortest path	5
		Substitution	107
		Absurdities	26
		Block design	24
		Missing parts	24
		Similarities	32
8	A non-verbal Group Test of Intelligence with special reference to the Mysore State (Age 7-13) by M.G.Premalatha.	Similarities	14
		Classification	15
		Progressive series	15
		Analogies	15
		Absurdities	24
		Substitution	100

The analysis of the above tests revealed that different sub-tests were selected to test specific functions of intelligence. It can be seen that there are a number of

sub-tests that are common to most of the intelligence test, e.g. similarity, classification, analogy, progressive series, absurdity and substitution table. It was, therefore, decided to include in the present test those sub-tests that are more often used as a measure of intelligence. The following table gives the idea of selection of subtests for the present test.

TABLE 4

Selection of Subtests in the Present Test

No.	Ability	Subtests which can test the ability	Subtests selected in the present test *
1	Reasoning	1.Classification of similar objects	*
		2.Analogies	*
		3.Similarities	*
		4.Figure arrangement	-
		5.Progressive Matrices	-
2	Perceptual	6.Absurdities	*
		7.Perceptual speed tests	-
		8.Missing parts	*
		9.Maize tests	-
		10.Progressive series	*
3	Memory	11.Substitution	*
		12.Digit span	-
4	Number	13.No series	-
		14.Progressive series	*
5	Motor tests	15.Speed tests	-
6	Spatial	16.Quantity	-
		17.Space tests	-
		18.Figure completion	-
		19.Figure classification	-

Thus it was decided to include seven subtests of

the following type:

No.	Type	Approximate number of items
1	Similarity	16
2	Classification	16
3	Analogy	16
4	Absurdity	16
5	Progressive series	16
6	Substitution (2 forms)	100
	Total	180

Choice of Age Range

The test was to be standardized for the children of the age group 7-13. This age group normally covers the school going children from third to eighth standard. It is not possible to go below the age of seven because this test will require the ability to write the name, the name of the school and to understand the oral instructions. Moreover, group testing will not be suitable to younger age groups than the one under the present investigation. This age range includes at least the first year of the secondary stage. This will certainly add to the utility of the test as it will be useful at the very start of a pupil's secondary school career.

Sample

The test was to be standardized for the Gujarat State. This would naturally include the three main regions of the State viz. Gujarat, Saurashtra and Kutch. The total area of the State is 72,137 square miles. Administratively the State is divided into seventeen districts viz.Ahmedabad, Sabarkantha, Banaskantha, Mehsana, Kaira, Panchmahals, Baroda, Broach, Surat,Dangs, Rajkot, Jamnagar, Surendranagar, Bhavnagar, Junagadh, Amreli and Kutch. The population of the State according to 1961 census is 206.12 lacs approximately. This population lives in 249 town groups and 18,777 villages of 17 districts of the State.

The following table shows the total number of children in the age group 6-11 and 11-14 during the year 1961-62.

TABLE 5

Number of Children by Age Groups

Age Groups	Estimated Number of children			Percentage to population
	Boys	Girls	Total	
6 - 11	1358000	1279000	2637000	12.80
11 - 14	724000	681000	1405000	6.82

The following tables show the enrolment of

children in the age groups 6-11 and 11-14 in all educational institutions during the year 1961-62.

TABLE 6

Enrolment of Children in the Age Group 6-11

Estimated population of children in the age group of 6-11	No.of children enrolled in schools			Total
	Nursery schools	Classes I - V	Classes VI - XI	
Boys 1358000	16494	1236568	12245	1255307
Girls 1279000	13481	650361	6671	670513
Total 2637000	29975	1876929	18916	1925820

TABLE 7

Enrolment of Children in the Age Group 11-14

Estimated population of children in the age group 11-14	No.of children enrolled in schools			Total
	Standard VI	Standard VI-VIII	Standard IX - XI	
Boys 724000	94759	141836	8885	245430
Girls 681000	42984	61727	4478	109189
Total 1405000	137743	203563	13363	354669

More detailed statistics of this sort will be

discussed in a later chapter. The study of the above tables spotlights the following facts:

(i) There will in all 40,42,000 children within the age range 6-14. This is 19.72 per cent of the general population of the State.

(ii) There were 22,80,489 children within the age range of 6-14. That means about 56 per cent of the children of this age group go to schools.

(iii) Enrolment of girls is comparatively less.

Looking to the vast number of the school-going children in the new Gujarat State, it was decided to take fairly representative sample of the children of the age group 7-13. The following points were kept in mind while selecting the sample:

(i) Only the school-going children were selected because the birth-dates of the other children would not be so dependable.

(ii) Boys and girls were selected in equal numbers.

(iii) Urban and rural parts of the Gujarat State got appropriate representation in the sample.

(iv) Three different regions of the State have pecularities of their own and due representation was given to them.

(v) Geographical consideration was also taken into account while selecting the schools.

(vi) Basic schools were proportionately included in the sample.

(vii) Non-Gujarati schools were also included in the sample as the children in these schools could understand the oral directions given in Gujarati.

(viii) The sample was selected keeping in view the following four criteria:

(a) Big cities:	Population more than 1,00,000.
(b) Small cities:	Population more than 25,000 but less than 1,00,000.
(c) Towns:	Population more than 5,000 but less than 25,000.
(d) Villages:	Population less than 5,000.

Some Other Points

(1) It was decided to prepare a test which does not require more than a school period on the whole. It is found that such speed tests normally require shorter time than the power tests. Considering the time that would be required to give directions, it was thought of constructing a test the actual working of which would not require more than 20 minutes. Such a time limit would facilitate the work of a tester and practically the school routine would not be disturbed to a great extent.

(2) It was decided to give separate directions for different subtests in Gujarati. These directions were to be orally read by the tester.

(3) It was decided not to give a separate answer sheet for the test. Children of this age group might find it complicated.

(4) It was decided to keep the size of the blocks of the individual items sufficiently large. The purpose of group testing would not be served unless the pictures are not of a fair size. This would also be helpful in giving the directions. The testees would look at and listen to the tester. The tester would explain to them with a test booklet in his hand.

Scoring

It was decided to score the test in a simple way by adding together the correct responses in each test. No special method of scoring was adopted for any test except the substitution tests in which the scores were reduced to one fourth. Some test-makers adopt a different method of scoring for items having only two options in the answers to eliminate the effect of random guessing. But in the present tests, nowhere two options have been provided, so, it was not necessary to use any method of eliminating the guessing effect. A formula has been evolved by statisticians to reduce the guessing effect even when there are as many as five options, but the use of the formula introduces complications in scoring which, in fact, should be as simple a process as possible. Moreover, the formula reduces the scores of even those testees who have generally answered the tests. So, in the present test the scoring was done in the simple manner by adding together the correct responses.

The composite score of all the tests is used by many test-makers for further standardization. Otis, for example, does not give different weightages to different tests in all his scales, but scores all tests as equal. Such uniform weightage is, however, not justifiable even though it gives satisfactory results, since different tests have different characteristics and different degrees of

difficulty. This method is found defective for the following reasons:

(1) Spearman found that the tests differ in their saturation with 'g'. Weightage should, in his opinion, be given to a test in proportion to its saturation with 'g' so as to make the whole battery the best measure of 'g'.

(2) The tests differ in the degree of difficulty. To give equal weightage to an easy test and a difficult one would lower down the value of the total score as a measure of intelligence.

(3) The number of items in the tests are fixed arbitrarily and when the raw scores of these tests are added together, naturally, a test with a greater number of items influences the composite score of the battery more than a test with less items.

(4) The variabilities (SDs) of different tests are different which the uniform weight method does not take into account.

Various methods have been in vogue to-day for removing the defects mentioned above. Some testers determine the weightages of different tests of their batteries

arbitrarily, taking into account the relative importance of the tests as measures of intelligence. Spearman evolved a method of fixing weightages of different test-scores in different tests to make a battery the best estimates of 'g'. This method involves a lot of calculation but, it has e its own merits, as it enhances the value of the battery as a measure of intelligence. Desai has adopted an altogether different method of determining weights of the tests. In his battery the weights are assigned to the tests in proportion to the time-limits fixed for them.

With regard to the present test, it was decided not to use any correction formula for guessing. As discussed in the previous paragraphs, a test with a greater number of items influences the composite score of the battery more than a test with less items. In the present test the total number of items on similarity, classification, analogy, absurdities and progressive series would be about 80 while that on substitution would be about 100. So in this case it was very likely that the scores of two substitution tests would influence the total score of the battery very heavily. In order to do away with this influence it was decided to reduce the scores of the substitution tests by one fourth.

Construction of Items

Constructing good test items is a creative art.

Construction of each item presents new situations and new problems. We cannot have a science of test construction. Just as there can be no hard and fast rules for writing a good story or a good poem, so there can be no set rules that will guarantee the production of a good test item. An item is best defined as "a single task or question that usually cannot be broken down into smaller unit."[2] E.Lindquist defined an item as "A scoring unit in an exercise."[3]

Non-verbal tests are not always culture free. The objects included in such tests should be such that the children can easily recognize them. Moreover, the objects should yield to pictorial presentation. Thus while constructing the items care was taken to include only those objects which are familiar to the children of Gujarat. In the beginning, the items were constructed in the verbal form for classification, progressive series, analogy and absurdity. Items for similarity and substitution were straight-away prepared in the form of pictures.

Pictorial presentation of all the items was done by an experienced artist. The artist had the practice of drawing pictures for text-books also. Furthermore he worked

---

2 Bean, K.L., Construction of Educational and Personal Tests, New York, McGraw Hill Book Inc., p. 15.

3 Lindquist, E.F., Educational Measurement, Washington D.C., American Council on Education, p. 185.

for many years as a drawing teacher in a school. These drawings were prepared in black ink on white drawing sheets.

Processing the Items

These drawing sheets were individually shown to 20 children all within the age range of 7-13. This was done only for one purpose and that was to see whether every object could be recognized by them. On presenting each of the sheets to the individual child he was asked to name the objects therein. Whenever there was any difficulty in recognition the child was asked the reason of his difficulty. The remarks were noted down. There were 203 drawings to be recognized by them. In the light of the difficulties experienced by pupils some drawings were got modified and the modified drawings were tried out again on the same children to assure the recognizability of the object. Such trying out was not needed for two subtests on substitution. For each subtest one item was selected for the illustration. This selection was done keeping the child of seven in mind.

After necessary corrections and modifications, drawings for individual items were got prepared for five subtests viz. similarity, classification, analogy, absurdity and progressive series. The drawings of individual items were separately prepared so as to facilitate the changing of order of items in a particular subtest. These drawings

were also prepared in black ink so that blocks for individual items could be prepared directly from such slips of drawing paper. Such slips were pasted on a cardboard roughly in order of difficulty. The item selected for illustration was kept first and some blank space was left in order to make it distinguishable from the rest of the items.

The next step was to see how every item as a whole worked. This was done by individual testing of 50 children who fall in the age range 7-13. Each child was seated comfortably in front of the cardboard. After giving some general encouraging hints, oral directions were given for a subtest. Then the item for illustration was explained. The child had to give the response with the help of a pointer. Here the investigator had to take the help of another person whose task was to note down the number of an item which was responded correctly by the child. In this manner, the item analysis chart was prepared with a view to finding out how each item worked. The chart did give some idea of the difficulty value of an individual item. Necessary corrections, omissions and modifications were made at the end of this tryout. The items, after discussions with experts were finalised. Some items on which the opinions differed were revised before the blocks of the designs were got prepared. In fine, this was all that was needed for the preparation of the pilot form of the test.

REFERENCES

1. Anastasi, A., Psychological Testing, MacMillan and Co., New York, 1956.

2. Bean, K.L., Construction of Educational and Personnel Tests, McGraw Hill Book Inc., New York, 1953.

3. Cronbach Lee, J., Essentials of Psychological Testing, Harper Brother, Publishers, New York, 1960.

4. Jordan, A.M., Measurement in Education, McGraw Hill Book Company, New York, 1953.

5. Lindquist, E.F., Educational Measurement, American Council on Education, Washington.

6. Ross, C.C., Measurement in To-days Schools, Prentice Hall, Inc., New York.

7. Thorndike and Hagen, Measurement and Evaluation in Psychology and Education, John Wiley and Sons, Inc., New York, 1955.

8. Wechsler, D., Measurement of Intelligence, The Williams and Wilkins Company, 1944.