CHAPTER V

PILOT TESTING

As described in the preceding chapter a set of test items was prepared, got criticized by experts and revised on the basis of the criticism. These revised items were tried out on a fairly representative sample. The test was, therefore, given to a small representative group of the population to be finally tested. This is known as pilot testing, an important step in the ladder of test construction procedure.

The objectives of the pilot testing are as under:

(1) To standardize the instructions to be given for the whole battery and also for each individual test.

(2) To fix the time-limit for each sub-test.

(3) To determine the difficulty value of each item.

(4) To determine the discriminating index of each item.

(5) To find out the range of applicability of the tests.

(6) To find out if any items needed any change in its structure and lay out.

Administration of the Pilot Test

In general, the administrative factors to be considered in the tryout are the same as those to be considered in the administration of the final test. The most important single principle to follow in planning the tryout is that it should be administered under as nearly as possible the same conditions as those under which the final test will be administered.

As regards the present test, the pilot test was administered to the pupils of 17 schools. Three districts of the State, viz. Surat, Broach and Baroda were selected for these purpose. The test was given to 1150 pupils. In selecting the sample for the pilot test the following principles were kept in view:

(i) Boys and girls were selected in almost equal numbers.

(ii) Basic and non-basic schools were proportionately selected.

(iii) Both rural and urban schools were selected keeping in mind the boys', girls' and mixed type of schools.

(iv) Primary schools and high schools with primary sections were included in the sample.

(v) The pupils were selected keeping in mind the different ages in the age-range.

(vi) Care was taken to select some schools wherein the medium of instruction was not Gujarati.

In sum, the schools of varied types were duly included with a view to making the sample as representative as possible. More details of the procedure of sample selection and of the educational set up of the Gujarat State are given in the next chapter.

The following table gives the details of the sample selected for the tryout.

TABLE 8

Details of the Sample of the Tryout

Sr. No.	Name of the School	Type of school	Pupils selected
1	Lokmanya Vidyalaya,Rander, Dist. Surat.	Mixed: High School	50
2	Baliadev Vidyamandir,Itola, Dist. Baroda.	Agriculture: High School	56
3	Jivan Sadhana, Baroda	Mixed: High School	65
4	Sharda Mandir, Varnama, Dist. Baroda.	Mixed: High School	56
5	Alembic Vidyalaya,Baroda	Mixed:High School	50
6	Navchetan Vidyalaya, Valia, Dist. Broach.	Mixed: High School (Primary section)	68
7	Urdu Primary School,Rander, Dist. Surat.	Girls: Primary School	45
8	Gujarati Main School, Rander, Dist. Surat.	Boys; Primary School	75
9	Gujarati School,Velachha, Dist. Surat.	Mixed: Basic Primary School	70
10	Surat Sudharai Shala No.3, Surat.	Boys: Primary School	110
11	V.S.Patel Vidyalaya,Kosamba, Dist. Surat.	Mixed: High School (Primary section)	153
12	Lok Vidyalaya, Kim, Dist. Surat.	Mixed: High School	31
13	Kumar Shala,Kim,Dist.Surat.	Mixed: Primary Non-Basic School	67
14	Jayshree Model High School, Baroda.	Mixed: Marathi Primary section.	97
15	Kumar Shala, Takarma, Dist. Surat.	Mixed: Basic Primary School	95

Sr. No.	Name of the school	Type of school	Pupils selected
16	Gujarati Kumar Shala, Kara, Dist. Broach	Boys: Basic Primary school	121
17	Buniyadi Shala, Arthan, Dist. Surat	Mixed: Basic Primary school	41

(1) Instructions for the Test

The directions to the testees in the tryout forms should be as nearly as possible identical with those that are to be used with the final form of the test. Directions for the pilot test were prepared on the basis of the preliminary testing. Both general instructions and directions for the subtests were prepared and tried out. The directions for the subtests included the explanation of the sample item for the subtests as well. These directions were not got printed as the investigator wanted nobody else to administer the pilot test. It was found that the sample item of the absurdity test was rather confusing. In the final test, therefore, the sample item was replaced by two new sample items. Similarly, in the substitution form 1, 3 sample items were given instead of two. Moreover, the position of these sample items in the substitution test was also changed in order to make it more clear and understandable.

(2) Timing the Test

Full time was given to the testees to answer all the items of the test. The testees were instructed to raise their hands as soon as they completed each subtest. The time taken by the quickest pupil in each subtest was noted. The idea behind fixing the time-limit for such a speed test was that only a few testees might be able to complete the test. The quickest pupil in one subtest might not be equally quick in another subtest. So, it should be rather impossible for any testee to finish all the subtests in the prescribed time. In the present test the time-limit was fixed by increasing the shortest time recorded, by about 30 seconds.

The table below shows the least time recorded and the time-limit fixed thereof for each test.

TABLE 9

Time-Limits for the Tests

Name of the subtest	Least time noted in minutes	Time-limit fixed finally in minut-es
Similarity	1.5	2.0
Classification	1.5	2.0
Analogy	2.5	3.0
Absurdity	2.0	2.5
Progressive series	2.5	3.0
Substitution 1	3.0	3.5
Substitution 2	0.75	1.0

Thus the actual time to be given for the final run of the test was fixed as 17 minutes. General instructions of the test and the directions for each subtest required about 20 minutes. So, it would be possible to administer the test within a normal school period of 35 to 40 minutes.

(3) Difficulty Values of the Items

After completing the administration of the pilot test, all the answer papers were scored. The item analysis of all the items of the first five subtests was then completed. The main object of the item analysis was to determine the difficulty value and the discriminating index of each item. This was very important to discard those items which had both very low or very high difficulty values and very low discriminating indices.

The test items in each subtest were arbitrarily arranged in the tryout. The main purpose of the tryout of a test is to find out the difficulty values of the items, so that they can be rearranged in the order of difficulty.

Many ways of expressing the difficulty level of an item have been proposed. The most obvious of these is the percentage of tryout group that marks it correctly. According to this method, an indication of the difficulty

level of an item is given by the percentage of individuals in a tryout group who can respond to the item correctly. The smaller the percentage, the more difficult the item and vice versa.

The second method which is in vogue at present is to prepare an item analysis chart. The first step in preparing such a chart is to arrange the test booklets according to total scores so that the test booklet with highest score remains at the top. The next step is to count off the top 27 per cent of test booklets laying aside the middle 46 per cent. Then the number in the top group which passes each item on the test is found out. Similarly, the number in the bottom group is also found out. These two numbers are converted into percentages. The difficulty value can then be found out using the following formula:

$$D = \frac{U + L}{2}$$

Where: D = Difficulty value.

U = Percentage of testees scoring the item in the top group.

L = Percentage of testees scoring the item in the bottom group.

Some doubts are expressed about the reliability

of difficulty value computed by this method as it involves the elimination of the middle 46 per cent answer books. F. Davis has investigated the problem and has concluded that the loss of reliability incurred by estimating indices from only 54 per cent of the sample is not sufficient to be of practical consequence when the two criterion groups employed at least 100 testees. The same investigator further says:

> Experimental evidence has shown that the difficulty indices of this sort are extremely reliable when they are based on samples as large as 400.[1]

In the present case the test was given to about 1150 pupils out of which 1110 test booklets were selected for the purpose. This number was selected with a view to avoiding unnecessary calculation as the 27 per cent of 1110 would be 300. The forty answer scripts that were discarded were selected at random. The difficulty levels of the test-items when full time is given to pupils, are considerably lower than those when limited time is given to them. In a way, the difficulty values have no meaning except that they may be made use of in re-arranging the items. The new order was given to the items of the first five tests. The question of re-arranging the items in the

---

1 Lindquist, E.F., Educational Measurement, Washington D.C., American Council on Education, 1950, p. 283.

two substitution forms did not arise at all. The table below gives the details of difficulty values of all the items of the first five sub-tests along with the new number of the items rearranged. It should be noted here that the lower the difficulty value of an item the higher is the difficulty level of the item.

TABLE 10

Difficulty Values of the Items
(Irrespective of Age)

Subtest	Item No.	U	L	Difficulty Value	New order
Similarity	1	90	47	68.5	4
	2	90	33	61.5	10
	3	97	49	73.0	1
	4	96	50	73.0	2
	5	96	48	72.0	3
	6	96	23	59.5	11
	7	90	37	63.5	8
	8	94	33	63.5	9
	9	93	41	67.0	5
	10	98	29	63.5	7
	11	88	21	54.5	12
	12	52	15	33.5	14
	13	82	17	49.5	13
	14	96	33	64.5	6

Subtest	Item No.	U	L	Difficulty value	New order
Classification	1	82	50	66	3
	2	97	61	79	1
	3	79	32	55.5	8
	4	81	33	57	6
	5	82	32	57	7
	6	90	50	70	2
	7	76	50	63	4
	8	77	20	48.5	9
	9	56	14	35	10
	10	33	3	18	13
	11	31	7	19	12
	12	92	27	59.5	5
	13	39	8	23.5	11
	14	24	10	17	14
Analogy	1	75	35	55	3
	2	95	23	59	1
	3	84	28	56	2
	4	77	29	53	4
	5	45	22	33.5	12
	6	67	25	46	8
	7	67	25	46	7
	8	69	16	42.5	10

Subtest	Item No.	U	L	Difficulty value	New order
	9	81	20	50.5	5
	10	59	10	34.5	11
	11	84	14	49.0	6
	12	46	11	28.5	13
	13	30	15	22.5	15
	14	36	8	22	16
	15	75	14	44.5	9
	16	38	11	24.5	14
	17	28	10	19.0	17
Absurdity	1	88	35	61.5	3ø
	2	97	40	68.5	2
	3	92	47	69.5	1
	4	64	13	38.5	11
	5	75	24	49.5	7
	6	54	13	33.5	15
	7	70	30	50	6ø
	8	81	38	59.5	4
	9	76	14	45	9
	10	57	12	34.5	14
	11	82	15	48.5	8
	12	42	8	25.0	18
	13	45	8	26.5	17

Subtest	Item No.	U	L	Difficulty value	New order
	14	63	13	38.0	13
	15	64	12	38.0	12
	16	84	22	53.0	5
	17	48	8	28.0	16
	18	66	12	39	10
Progressive series	1	97	49	73	1
	2	37	3	20	16
	3	98	26	62	5
	4	98	47	72.5	2
	5	97	33	65.0	3
	6	95	30	62.5	4
	7	79	37	58.0	6
	8	84	22	55.0	8
	9	50	23	36.5	10
	10	72	29	50.5	9
	11	88	23	55.5	7
	12	48	23	35.5	12
	13	70	2	36.0	11
	14	66	3	34.5	13
	15	45	2	23.5	15
	16	50	6	28	14
	17	30	4	17	17

It will be seen from the above table that the difficulty values of the items range from 17 to 79. It is found that there is no hard and fast criterion for selecting or rejecting an item from the view point of its difficulty value. Different test makers use different limits of higher and lower difficulty levels. This, many times, depends upon the nature and purpose of the test. In the present case, the test is meant for the age range of 6 years and hence the items of all difficulty levels have been selected. So the difficulty values of the items, here, were useful only in so far as they provided a base for rearrangement of items in the order of difficulty.

Furthermore, the difficulty value of the items should go on increasing (i.e. the items should be found easier) as we move towards the higher ages. It is beyond doubt that the total score on the test will go high as we move towards the higher ages. Similarly, the number of students responding to an item correctly should increase with the age. The table below gives the agewise difficulty values of items of the first five subtests:

TABLE 11

Agewise Difficulty Values of the Items of the First Five Subtests

Subtests	Item No.	7 years	8 years	9 years	10 years	11 years	12 years	13 years	14 years
Similarity	1	61	71.5	82	82	89	90	90	93
	2	11	21.5	30	43.5	48	80	84.5	85
	3	44	61	81	83.5	85.5	85	87	89.5
	4	61.5	80	82	82	85	83	83.5	88
	5	63	83.5	86	87	88	88	89	91
	6	10	24	46	48	48.5	53	55.5	58
	7	23.5	35	37.5	42	44	46.5	65	67
	8	44	49	53	62	65	68	72	86
	9	74	77.5	78.5	82	85.5	87	91	96
	10	27	39.5	61	84	86	88	93	95.5
	11	14	15.5	17	24	28.5	57	58	58.5
	12	15	17	18.5	19	19	29	51	53
	13	16	18.5	20	19.5	27	38	54	71

Subtests	Item No.	7 Years	8 Years	9 Years	10 Years	11 Years	12 Years	13 Years	14 Years
	14	45.5	63	70.5	81	86	88.5	92	97
Classification	1	43	69	72	98.5	82	86	89.5	95
	2	80.5	83	84.5	87	90	91	93	96
	3	45	47.5	49	57	53	53	57.5	62
	4	65	71	77	82	85	85.5	89	93
	5	41	51	51.5	67	73	84	88	91
	6	71	80	84	87	91	93	96.5	98
	7	71.5	80.5	86	88	93	95	96.5	97
	8	0	22	27	31	37	38.5	44	48
	9	21.5	39	40	39	39	39.5	48	80
	10	5	18	27	14	51	56.5	67	73.5
	11	4.5	16	18	19	19.5	27	34	41
	12	61	66.5	81	85.5	89	91	93.5	96
	13	4	45	54	55	56	56	78	86
	14	1.5	17	17	18.5	31	76	89	93

Subtests	Item No.	7 Years	8 Years	9 Years	10 Years	11 Years	12 Years	13 Years	14 Years
Analogy	1	21	47	61	70	81	83	88	89
	2	23	69	82	83	83	87	93	96
	3	22	71	80	81.5	87	89	92	95.5
	4	22.5	28	68	71	86	88	89.5	96
	5	6	11	22.5	26	44	50	52	57
	6	7.5	14.5	49	69	77.5	78	78	79
	7	6.5	10	61	63	70	73	77.5	83
	8	7	11.5	42	48	61.5	70	73	78
	9	21	29	60.5	68.5	82	85	90.5	96
	10	5.5	9.5	23	46	63.5	69	80	85
	11	28	31	62	70.5	85	86	90	93
	12	4	14	27	29.5	41	48.5	51	57
	13	8.5	14	16	17	27.5	31	34	37
	14	8.5	16.5	16.5	18	18.5	19	19	19.5
	15	7	21	49	62	68	69	73	77.5

Subtests	Item No.	7 Years	8 Years	9 Years	10 Years	11 Years	12 Years	13 Years	14 Years
	16	9	15	15.5	16	41	42	72	78
	17	8	12	13	15	16	17	18	17
Absurdity	1	29	65	80	86	90	91.5	93	96
	2	41	49	82	83	87	90	96	98
	3	44	65	70.5	78	86	92	96	99
	4	10.5	22	32	40.5	47	54	64	72
	5	33.5	40.5	51	61	80.5	87	91	94
	6	6.5	9.5	12.5	24	28	33	34	51
	7	21.5	41	49	67	80	80	82	84
	8	21	60.5	70	83.5	86	87.5	90.5	93
	9	26	31	43	51	63	65.5	72	80
	10	4.5	7.5	22	26	31	33	44	50.5
	11	21	30	49.5	60	70	88	82.5	88
	12	21.5	33	39	51	54.5	69	73	78
	13	3	9	14	14.5	16	17.5	18	23

Subtests	Item No.	7 Years	8 Years	9 Years	10 Years	11 Years	12 Years	13 Years	14 Years
	14	4	20.5	25	29	32.5	41	46	46.5
	15	4.5	21	30	31.5	41	49	49.5	60.5
	16	29.5	41	52	53	58	62	63	73
	17	11	12	13.5	15	16	17	17.5	28
	18	3	9	10	11.5	13	15	16	16.5
Progressive series	1	82	84.5	87	90	92	93	96	97
	2	9	11	13	14.5	15	21	26	31
	3	21	43	51	63	81	86	89	90
	4	72.5	81	80	82	87	93	93	94.5
	5	43	62	67	84	88	92	95	97
	6	19.5	45	69	80	86	90	93	94.5
	7	28	51	55.5	67.5	82	91	93	95
	8	27.5	29.5	51	54	85	87	90.5	93
	9	8	22	29	40.5	67	69.5	83.5	87.5
	10	6	26	31	49	80.5	80.5	86	89

Subtests	Item No.	7 Years	8 Years	9 Years	10 Years	11 Years	12 Years	13 Years	14 Years
	11	20	31.5	45	51	81	90.5	92	93
	12	6.5	9.5	21	28	51	65.5	88	91
	13	4.5	7	20	31	46	46.5	54	81
	14	4	8	22	33.5	49	54	61	69
	15	6	10	11.5	13	22	27	31.5	44
	16	3.5	5	6.5	23	26.5	44	50.5	61
	17	7	9	10	12	15.5	16	17	17.5

(4) The Discriminating Index of Each Item

The discriminating index of an item is determined by the extent to which the given item discriminates among testees who differ sharply in the function (or functions) measured by the test as a whole. A number of methods have been devised for determining the discriminative power of an item. But biserial correlation is usually regarded as the standard procedure in item analysis. Biserial r gives the correlation of an item with total score on the test, or with scores in some independent criterion. The adequacy of other methods is judged by the degree to which they are able to yield results which approximate those obtained by biserial correlation.

The method used in the present case is that of forming extreme groups for computing the validity of an item. The procedure is the same as that of finding out the difficulty value. The U (percentage of testees scoring the item in upper group) and L (percentage of testees scoring the item in the lower group) values of an item are used once again for finding out the item validity. Entering Flanagan's table of normalized biserial coefficients, with the per cent of successes in the two groups, one can read the biserial form the intersecting column and row in the

body of the table.[1]

The table below gives the biserial r of the items i.e. the discriminative power of the items - their consistency with total score on the test.

TABLE 12

Biserial Coefficient of Test Items
(First Five Subtests)

Item No.	Biserial coefficients				
	Similarity	Classification	Analogy	Absurdity	Progressive series
1	0.505	0.36	0.41	0.56	0.65
2	0.605	0.575	0.74	0.70	0.56
3	0.65	0.48	0.57	0,54	0.79
4	0.61	0.49	0.48	0.54	0.695
5	0.62	0.51	0.26	0.51	0.73
6	0.755	0.48	0.43	0.465	0.70
7	0.575	0.28	0.43	0.40	0.437
8	0.665	0.565	0.54	0.455	0.61
9	0.60	0.47	0.605	0.62	0.295
10	0.775	0.527	0.55	0.50	0.43
11	0.665	0.385	0.68	0.66	0.65
12	0.415	0.67	0.43	0.45	0.276
13	0.64	0.43	0.205	0.48	0.77
14	0.705	0.23	0.40	0.53	0.725
15			0.61	0.55	0.65
16			0.36	0.61	0.585
17			0.28	0.51	0.46
18				0.57	

1 Garrett,H.E., Statistics in Psychology and Education, Bombay, Allied Pacific Private Ltd., p. 366.

It will be seen from the above table that no item has the biserial coefficient less than 0.23. According to R.L.Thorndike, "Correlation coefficient of 0.25 represents an outstanding validity."[3]

Looking to the item validity of all the items of the first five subtests, it was decided to retain all the 80 items in the final run.

(5) Fixing the Age Range

Since the tests were prepared for the age range 7 to 13 roughly, it was necessary to determine the exact lower limit of the grade-range to which the tests could be applied. 7 year pupils are generally found in standards II and III. First of all the tests were given to pupils of standard III. It was found that only 50 per cent of them could follow the instructions and worked out the tests satisfactorily. The tests, then, were tried in standard II. Here it was found that practically none could follow the instructions properly. Most of them had no practice of holding a pencil. So, the test was as good as meaningless for the pupils of standard II. In rural areas the test was not properly understood by some pupils of standard III.

Moreover, the difficulty value of each item was

3 Thorndike, R.L., Personnel Selection, New York, John Wiley & Sons, Inc., p. 245.

found out separately for different ages. The items were then grouped according to their difficulty values for different ages as shown in the table on the next page. It should be remembered here that the higher the difficulty value of an item the easier is the item.

TABLE 13

Agewise Distribution of Difficulty Values of Items
(N = 370)

Similarity (1)

Age	No.of pupils	0-19	20-39	40-59	60-79	80-100
7	46	2,6,11,12,13	7,10	3,8,14	1,4,5,9	-
8	46	11,12,13	2,6	9,8,10	1,3,9,14	4,5
9	46	11,12	2,13	6,7,8	9,10,14	1,3,4,5
10	46	12,13	11	2,6,7	8	1,3,4,5,9,10,14
11	46	12	11,13	2,6,7	8	1,3,4,5,9,10,14
12	46	-	12,13	6,7,11	8	1,2,3,4,5,9,10,14
13	46	-	-	6,11,12,13	7,8	1,2,3,4,5,9,10,14
14	46	-	-	6,11,12	7,13	1,2,3,4,5,8,9,10,14

Classification (2)

Age	No.of pupils	0-19	20-39	40-59	60-79	80-100
7	46	8,10,11,13,14	9	1,3,5	4,7,12	2,6
8	46	10,11,14	8,9	3,5,13	1,4,12	2,6,7
9	46	11,14	8,10	3,5,9,13	1,4	2,6,7,12
10	46	11,14	8,9	3,10,13	1,5	2,4,6,7,12
11	46	11	8,9,14	3,10,13	5	1,2,4,6,7,12
12	46	-	8,9,11	3,10,13	14	1,2,4,5,6,7,12
13	46	-	11	3,8,9	10,13	1,2,4,5,6,7,12,14
14	48	-	-	8,11	3,10	1,2,4,5,6,7,9,12,13,14

Analogy (3)

Age	No.of pupils	0-19	20-39	40-59	60-79	80-100
7	46	5,6,7,8,10,12,13, 14,15,16,17	1,2,3,4,9, 11	-	-	-
8	46	5,6,7,8,10,12,13, 14,16,17	4,9,11,15	1	2,3	-
9	46	13,14,16,17	5,10,12	6,8,15	1,4,7,9,11	2,3
10	46	13,14,16,17	5,12	8,10	1,4,5,6,7,9, 11,15	2,3
11	46	14,17	13	5,12,16	6,7,8,10,15	1,2,3,4,9,11
12	46	14,17	13	5,12,16	6,7,8,10,15	1,2,3,4,9,11
13	46	14,17	13	5,12	6,7,8,15,16	1,2,3,4,9,10,11
14	48	14,17	13	5,12	8,15,16	1,2,3,4,6,7,9, 10,11

Absurdity (4)

Age	No.of pupils	0-19	20-39	40-59	60-79	80-100
7	46	4,6,10,13,14,15 17,18	1,5,7,8,9,11, 12,16	2,3	-	-
8	46	6,10,13,17,18	4,9,11,12,14, 15	2,5,7,16	1,3,8	-
9	46	6,13,17,18	4,10,12,14,15	5,7,9,11,16	3,8	1,2
10	46	13,17,18	6,10,14,15	4,9,12,16	3,5,7,11	1,2,8,16
11	46	13,17,18	6,10,14	4,12,15,16	9,11	1,2,3,5,7,8
12	46	13,17,18	6,10	4,14,15	9,11,12,16	1,2,3,5,7,8
13	46	13,18	6,17	10,14,15	4,9,12,15,16	1,2,3,5,7,8,11
14	48	18	13,17	6,10,14	4,12,16	1,2,3,5,7,8,9,11

Progressive Series (5)

Age	No. of pupils	0-19	20-39	40-59	60-79	80-100
7	46	2,9,10,12,13,14,15, 16,17	3,6,7,8,11	5	4	1
8	46	2,12,13,14,15,16,17	8,9,10,11	3,6,7	5	1,4
9	46	2,15,16,17	9,10,12,13,14	3,7,8,11	5,6	1,4
10	46	2,15,17	12,13,14,16	8,9,10,11	3,5,7	1,4,6
11	46	2,17	15,16	12,13,14	9	1,3,4,5,6,7,8, 10,11
12	46	17	2,15	13,14,16	9,12	1,3,4,5,6,7,8, 10,11
13	46	17	2,15	13,16	14	1,3,4,5,6,7,8, 10,11,12
14	48	17	2	15	14,16	1,3,4,5,6,7,8,9, 10,11,12,13

The study of the above table spotlighted the following facts:

(1) The items on similarity worked well with all the children of the age group 7 to 14.

(2) The items on classification also worked well with the entire group.

(3) The items on analogy, absurdity and progressive series were found to be very difficult by the children of seven years. These subtests were practically useless for the children of seven years.

(4) The children of 13 and 14 years scored high in the first two tests.

From what has been discussed above, it was decided to keep the lower age-limit as 8 years. In this investigation a pupil who is between 7 years and 6 months and 8 years and 5 months is considered to have the chronological age of 8 years. Since the tests were satisfactorily answered by pupils of 8 years, they were easily applicable to the higher age groups. ~~It was found that~~ The test was also tried out on the pupils of 15 years (i.e. 14-6 to 15-5). It was found that they scored very high. Finally, it was decided to keep the age range of the test from

8 years to 14 years. It is to be remembered here that the test did not work well with the pupils of standard II even if they fell within this age range. It was found that the pupils below standard III found it very difficult to use the pencil. This is more true for the pupils of the rural areas.

(6) Change in the Items

As this is a non-verbal test, the question of changing the wording of items did not arise. With regard to the general layout, it was found that there was no difficulty in following any item of the test. The blocks were very clear and there was no need of modifying any of the items of the test. There was no such question for the two substitution forms and the items therein were practically retained as they were.

So, the pilot test

(i) gave the investigator an experience of administering these tests;

(ii) enabled him to fix the new age-range for the test;

(iii) led him to finalise the instructions to be given for the whole battery and also for the individual subtest;

(iv) showed the difficulty values and discriminating indices of the test items which paved the way to item selection;

(v) gave the clue for timing the tests;

(vi) enabled him to make necessary changes in the sample items; and above all,

(vii) prepared him to avoid the mistakes made in the tryout itself.

These are the things necessary for the final run of the test and it is here that the pilot test proves to be an important step in the ladder of test construction procedure.

REFERENCES

1. Anastasi, Anne, Psychological Testing, MacMillan Co., New York, 1955.

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

3. Garrett, H.E., Statistics in Psychology and Education, Allied Pacific Private Ltd., Bombay, 1961.

4. Guilford, J.P., Fundamental Statistics in Psychology and Education, McGraw Hill Book Co., New York, 1956.

5. Lindquist, E.F., Educational Measurement, American Council on Education, Washington D.C., 1950.

6. Lindquist, E.F., Statistical Analysis in Educational Research, Houghton Mifflin Company, Boston,1940.

7. Micheels, W.J., and Karnes, Measuring Educational Achievement, McGraw Hill Book Inc., New York.

8. Ross, C.C., Measurement in To-day's Schools, Prentic Hall, New York, 1954.

9. Thorndike, R.L., Personnel Selection - Test and Measurement Techniques, John Wiley and Sons,Inc., New York.