CHAPTER VII

STANDARDIZATION PROCEDURE

Standardization is the culmination of the process of the try-out. It is through this procedure that we confirm our selection and elimination statistically by administering the test to a representative sample of the population. In testing, standards are not fixed arbitrarily on a priori ground by a group of persons interested in this domain but are evolved out of the statistical devices. So far as the steps of standardization are concerned, they are as follows:

(i) Standardization of material.

(ii) Standardization of method.

(iii) Standardization of results.

The material and method have been standardized in the pilot test. The present chapter deals with the third step, the standardization of results which involves the following steps:

I Statistical analysis of the data;

II Determining reliability of the test;

III Determining validity of the test;

IV Fixing the norms.

I. Statistical Analysis of the Test Results

Before the test scores or rather quantitative data can be made comprehensible, interpretable and meaningful, it is always necessary to subject them to statistical treatment.

The number of items in the test is 170. The number of items in the first five subtests is 80 and one mark was given for each item. The number of items in the two substitution tables is 90 and the total score of these two subtests was reduced to one-fourth. Thus the maximum possible score on the test as a whole is 102.5. The highest score obtained on the test was 93 and the lowest one was 5. Thus the range of the scores is 88. This range within which all the scores were distributed was divided into ten class intervals, each interval being of ten units. The agewise distribution of scores is shown on the next page.

TABLE 33

Agewise Distribution of Scores (N = 6037)

	8	9	10	11	12	13	14	Total
90 - 99	0	0	0	0	0	0	3	3
80 - 89	0	0	2	10	10	75	103	200
70 - 79	5	4	29	20	98	151	240	547
60 - 69	47	40	80	77	215	183	201	843
50 - 59	52	108	140	201	160	250	100	1011
40 - 49	90	201	302	200	119	100	76	1088
30 - 39	259	267	235	134	50	60	45	1050
20 - 29	276	163	105	103	40	50	11	748
10 - 19	153	103	50	52	25	21	5	409
0 - 9	68	29	19	12	1	9	0	138
Total	950	915	962	809	718	899	784	6037

TABLE 34

Agewise Means and Standard Deviations

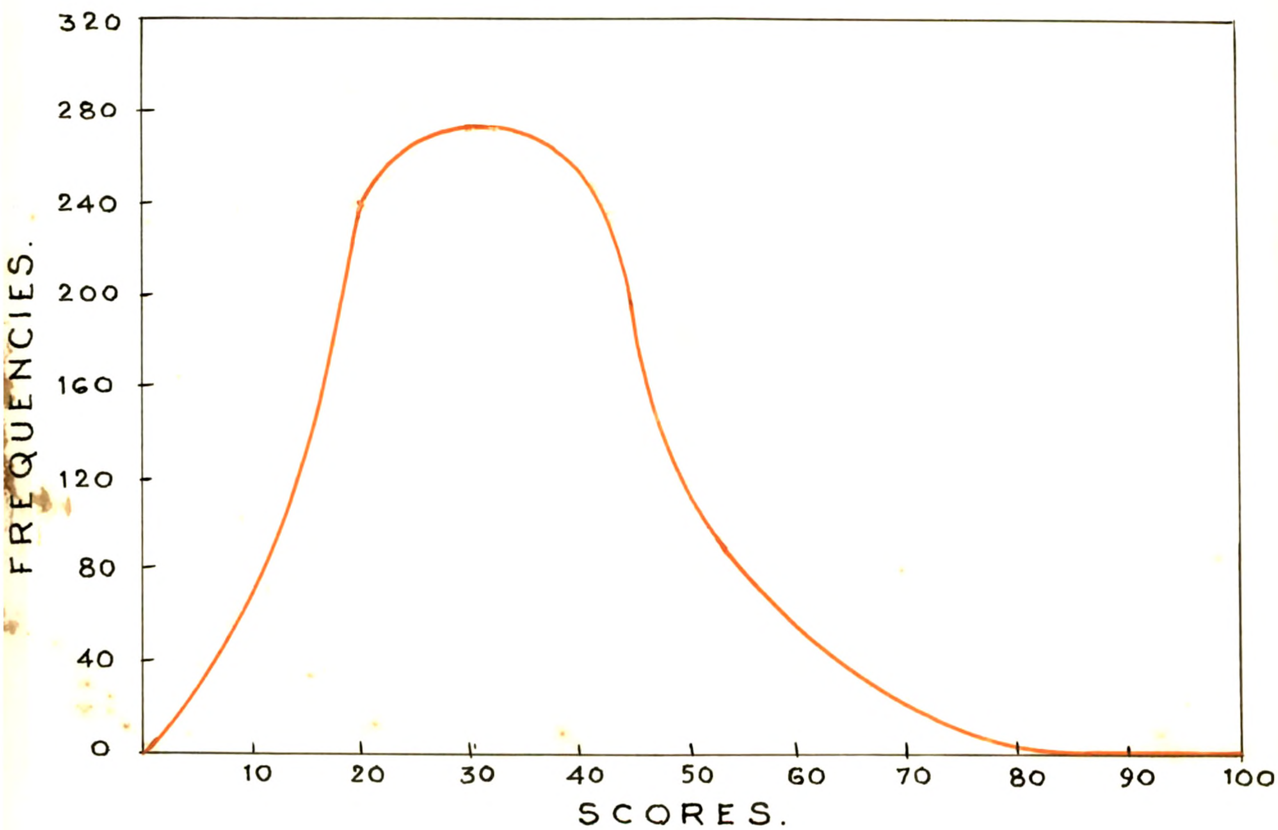
Age	N	Mean	S.D.
8	950	29.96	14.8
9	915	35.55	14.3
10	962	41.64	14.6
11	809	43.4	15.9
12	718	54.5	15.8
13	899	56.86	17.5
14	784	64.5	15.1

The nature of distribution of scores for different ages and of the group as a whole can be studied from the graphs given hereafter.

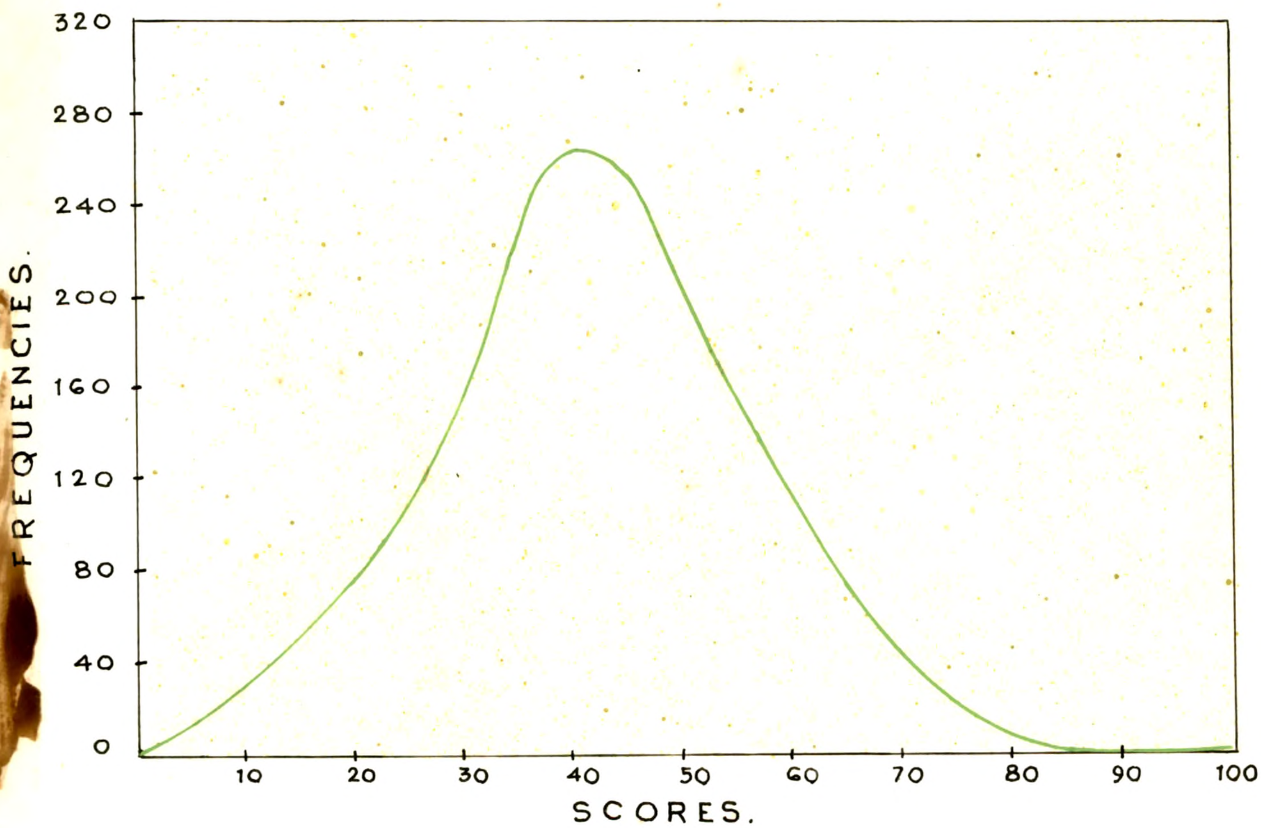
Some Statistics of the Total Distribution

1. Mean = 45.7
2. Median = 45.69
3. Standard deviation = 19.2
4. Quartile deviation = 14.47
5. Skewness (SK) = 0.0015
6. Kurtosis (Ku) = 0.292

GRAPH SHOWING THE DISTRIBUTION OF SCORES OF THE PUPILS OF 8 YEARS.

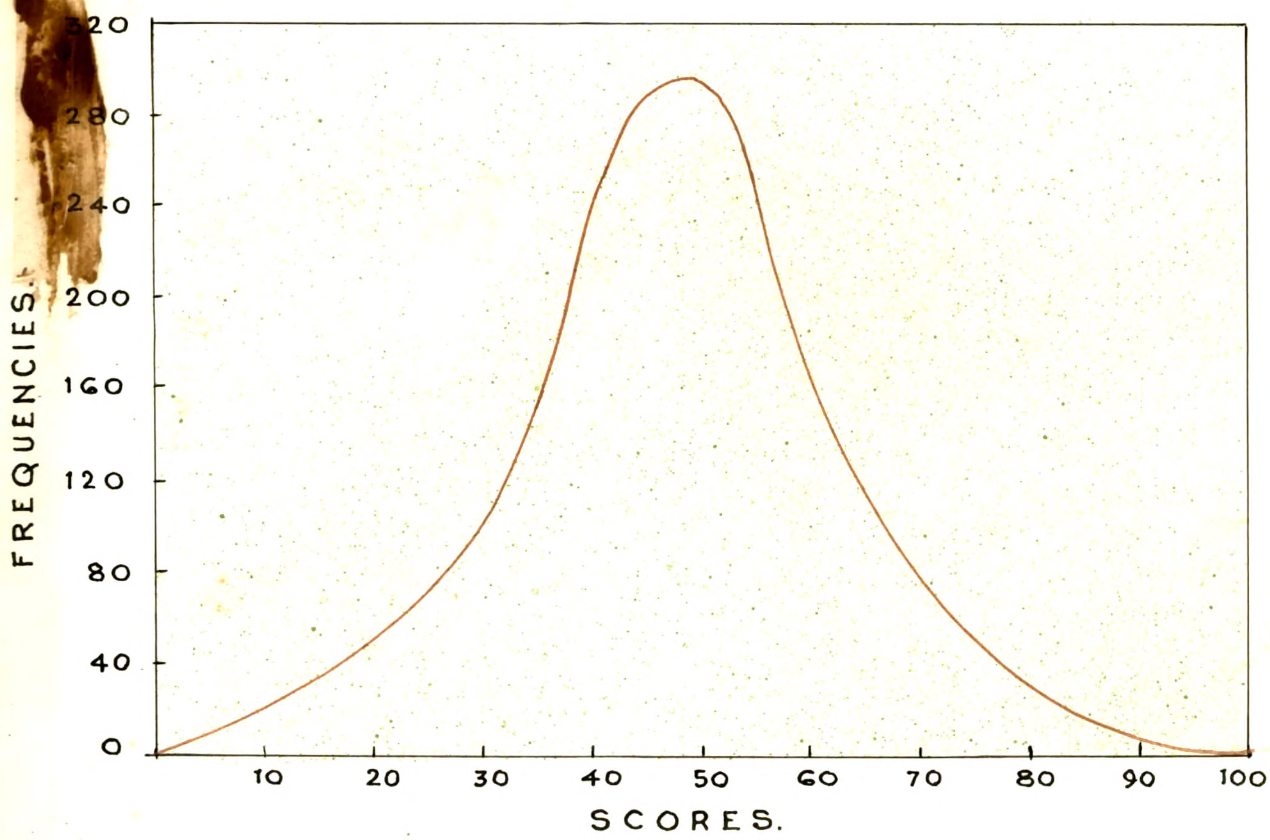


GRAPH SHOWING THE DISTRIBUTION OF SCORES OF THE PUPILS OF 9 YEARS.

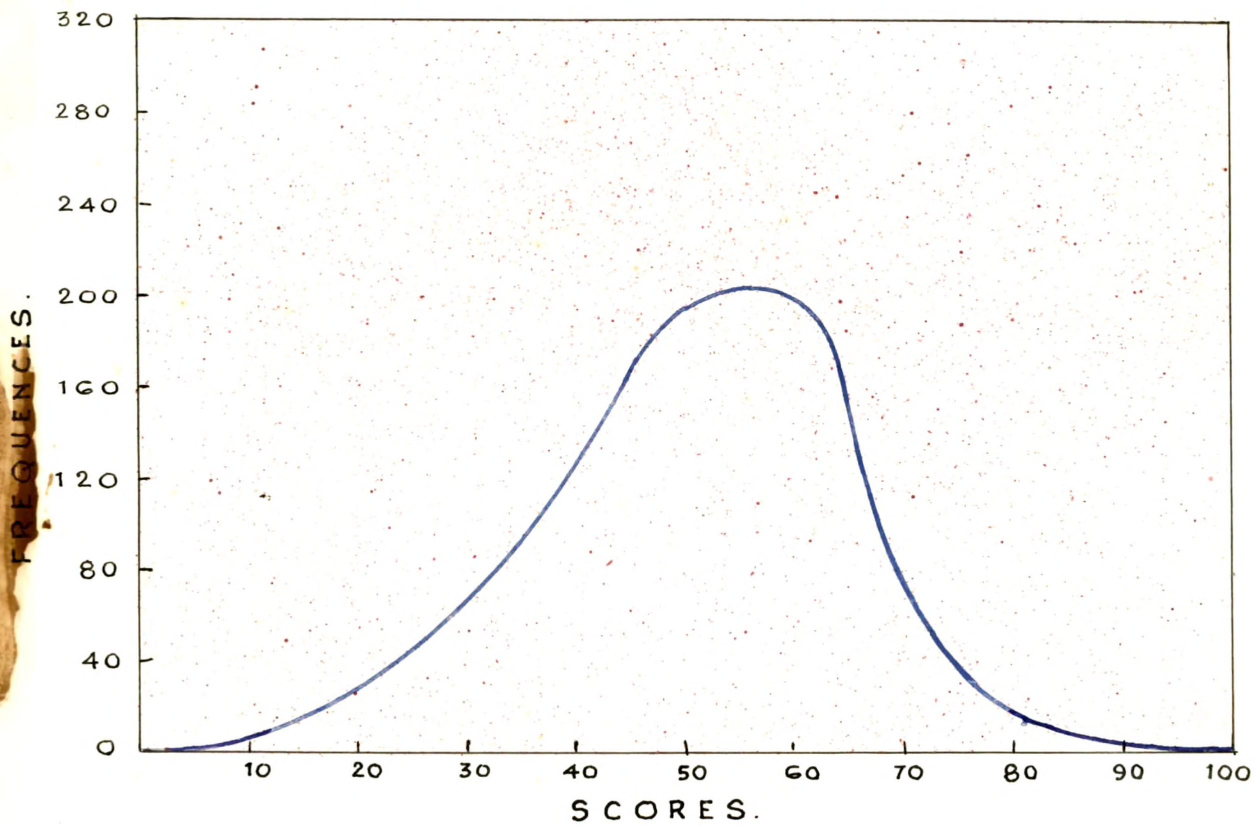


151-C

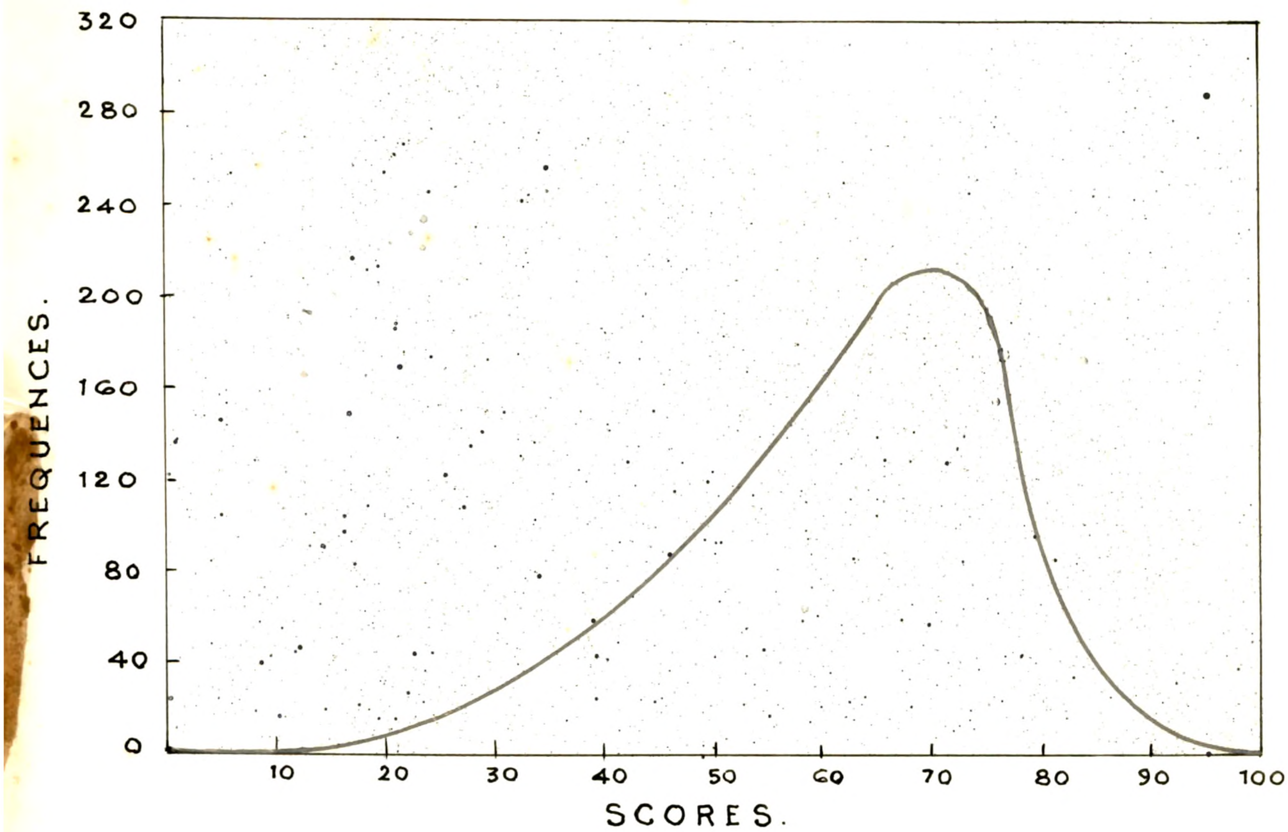
GRAPH SHOWING THE DISTRIBUTION OF SCORES OF THE PUPILS OF 10 YEARS.



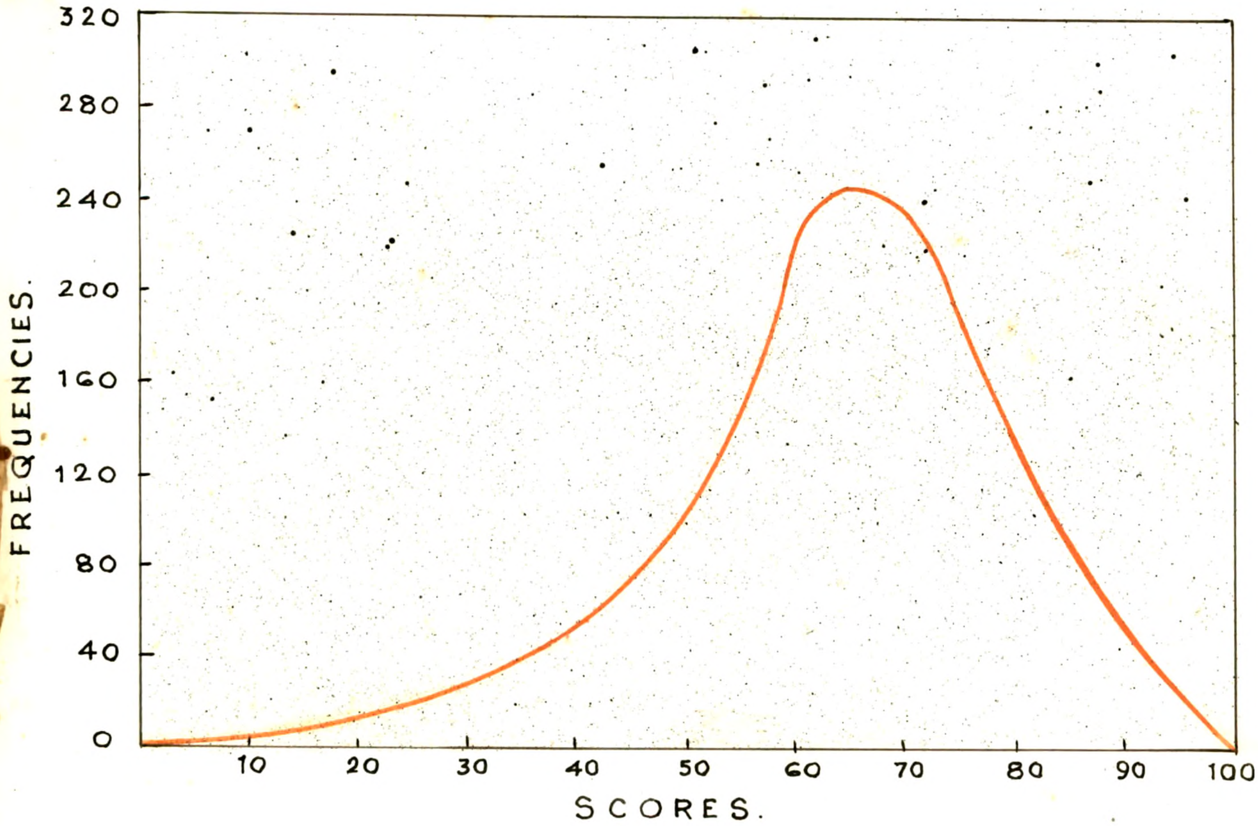
GRAPH SHOWING THE DISTRIBUTION OF SCORES OF THE PUPILS OF 11 YEARS.



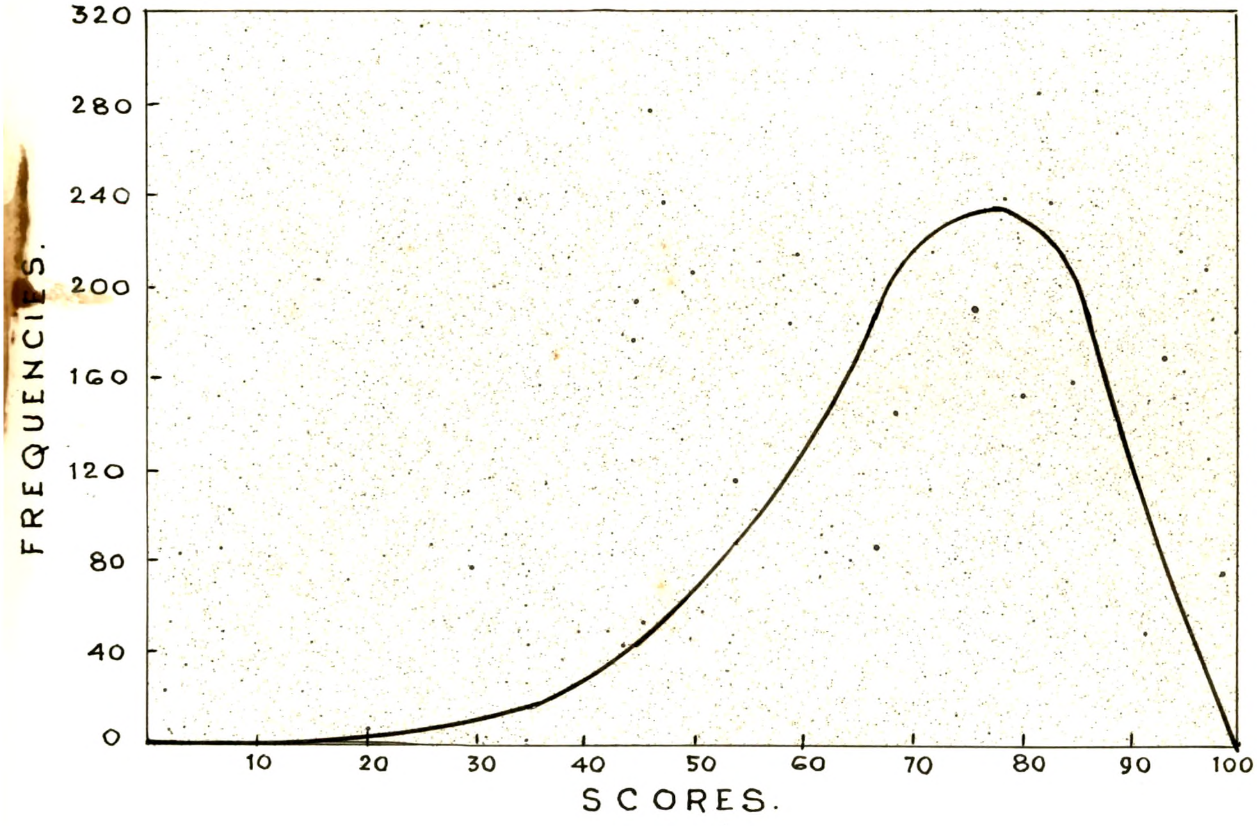
GRAPH SHOWING THE DISTRIBUTION OF SCORES OF THE PUPILS OF 12 YEARS.



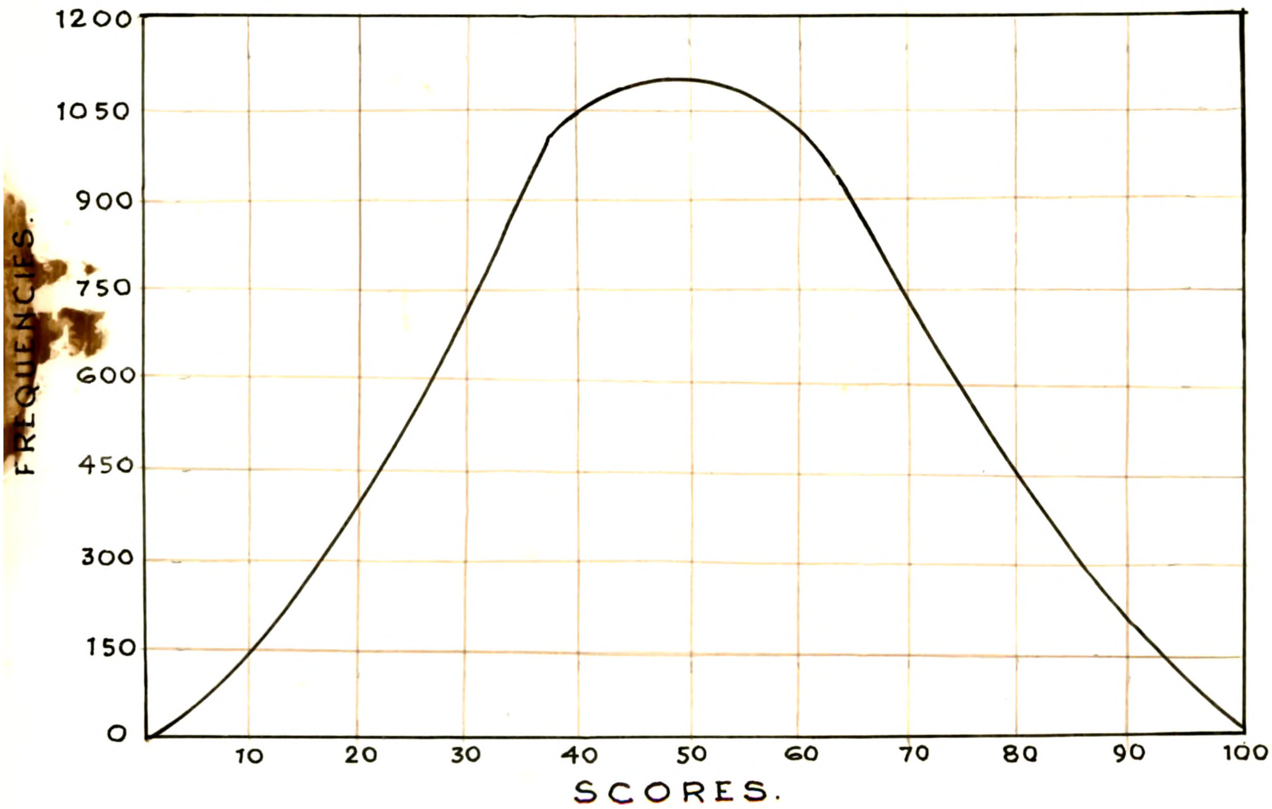
GRAPH SHOWING THE DISTRIBUTION OF SCORES OF THE PUPILS OF 13 YEARS.



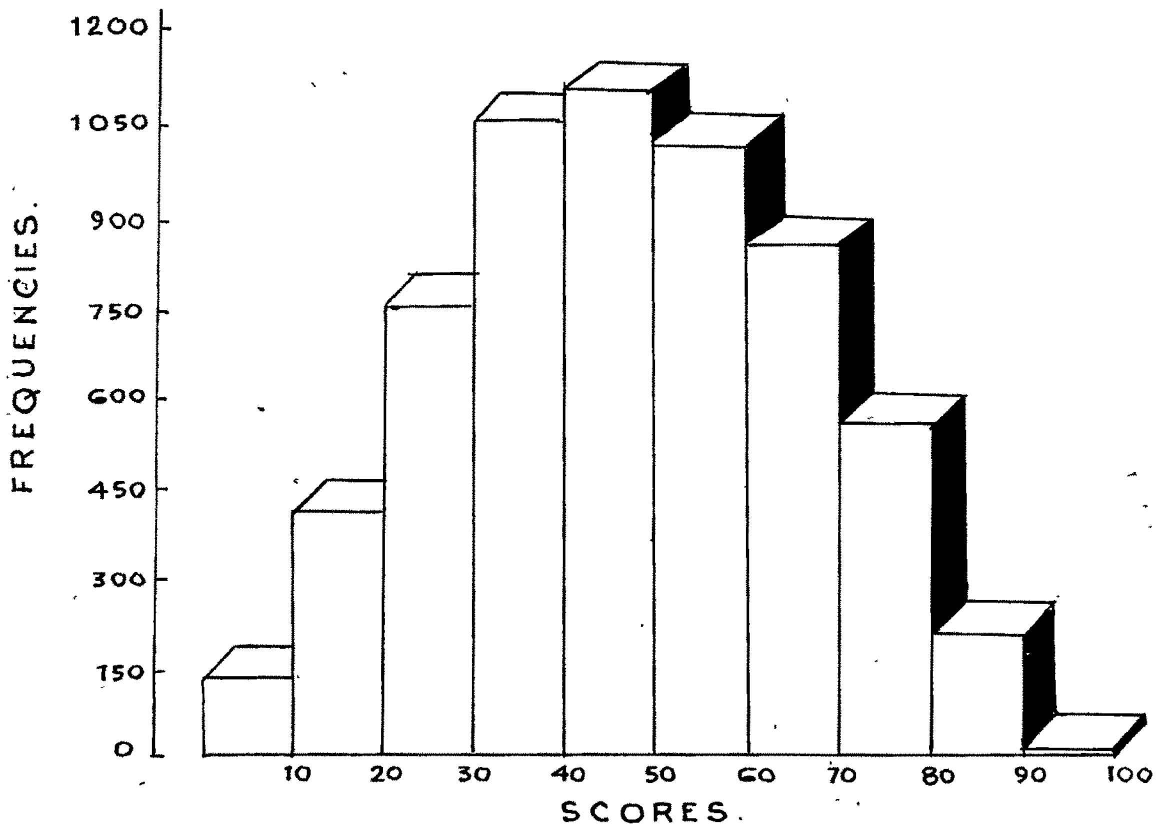
GRAPH SHOWING THE DISTRIBUTION OF SCORES OF THE PUPILS OF 14 YEARS.



GRAPH SHOWING THE DISTRIBUTION OF SCORES OF THE PUPILS OF THE ENTIRE AGE GROUP (8-14.)



HISTOGRAM SHOWING THE DISTRIBUTION OF SCORES OF THE PUPILS OF THE ENTIRE AGE GROUP (8-14.)



Reliability of the Statistics

(i) Reliability of the Mean

$$SE_{Mean} = \frac{\sigma}{\sqrt{N}}$$

where σ is the standard deviation of the distribution and N is the number of children included in the sample.

$$= \frac{19.2}{\sqrt{6037}}$$

$$= .247$$

∴ The 'true' mean lies between

45.7 ± 2.58 x .247 (99 per cent level of confidence)

i.e. between 46.34 and 45.06.

This narrow range within which the 'true' mean lies shows that the obtained mean is highly reliable.

(ii) Reliability of the Median

$$SE_{MDN} = \frac{1.253\sigma}{\sqrt{N}}$$

$$= 1.253 \times 0.247$$

$$= 0.31$$

The 'true' median lies between

45.69 ± 2.58 x 0.31 (99 per cent level of confidence)

i.e. between 46.49 and 44.89.

This narrow range within which the 'true' median lies shows that the obtained median is highly reliable.

(iii) Reliability of the Standard Deviation

$$SE_{S.D.} = \frac{0.71\sigma}{\sqrt{N}}$$

$$= 0.71 \times 0.247$$

$$= 0.17$$

The 'true' standard deviation lies between

$19.2 \pm 2.58 \times 0.17$

i.e. between 19.64 and 18.76

This narrow range within which the true standard deviation lies shows that the obtained standard deviation is highly reliable.

(iv) Reliability of the Skewness

To find this we have to determine the critical ratio 't' as measured by the formula:

$$t = \frac{SK}{\sigma SK}$$

where σSK is the standard error of the skewness.

Now σSK is obtained by using the formula:

$$\sigma SK = \frac{.5185\ (P_{90} - P_{10})}{\sqrt{N}}$$

$$= \frac{.5185 \ (69.76 - 20.26)}{77.6}$$

$$= .29$$

Now $\frac{SK}{\sigma SK} = \frac{0.0015}{.29}$

$$= .0051$$

From the table of testing significance, it can be seen that this value is not at all significant. Hence 0.0015 represents no real deviation of the frequency distribution from normality.

(v) Reliability of the Kurtosis

The Kurtosis of the obtained distribution is 0.292 and that of the normal curve is 0.263. Hence the significance of this difference is to be tested. The difference D between these values is 0.292 - 0.263 = 0.029.

Standard error of the Kurtosis ' Ku' is obtained by the formula:

$$\sigma Ku = \frac{.27779}{\sqrt{N}}$$

So $\sigma Ku = \frac{.27779}{77.6}$

$$= .0035$$

$$\text{Now} \quad t = \frac{D}{\sigma_{Ku}}$$

$$= \frac{0.029}{0.0035}$$

$$= 8.28$$

This value is significant for N = 6037 so the conclusion is that the curve is significantly platykurtic in nature.

The next step was to see whether the test discriminates pupils from year to year. This can be seen from the table and the graph given hereafter.

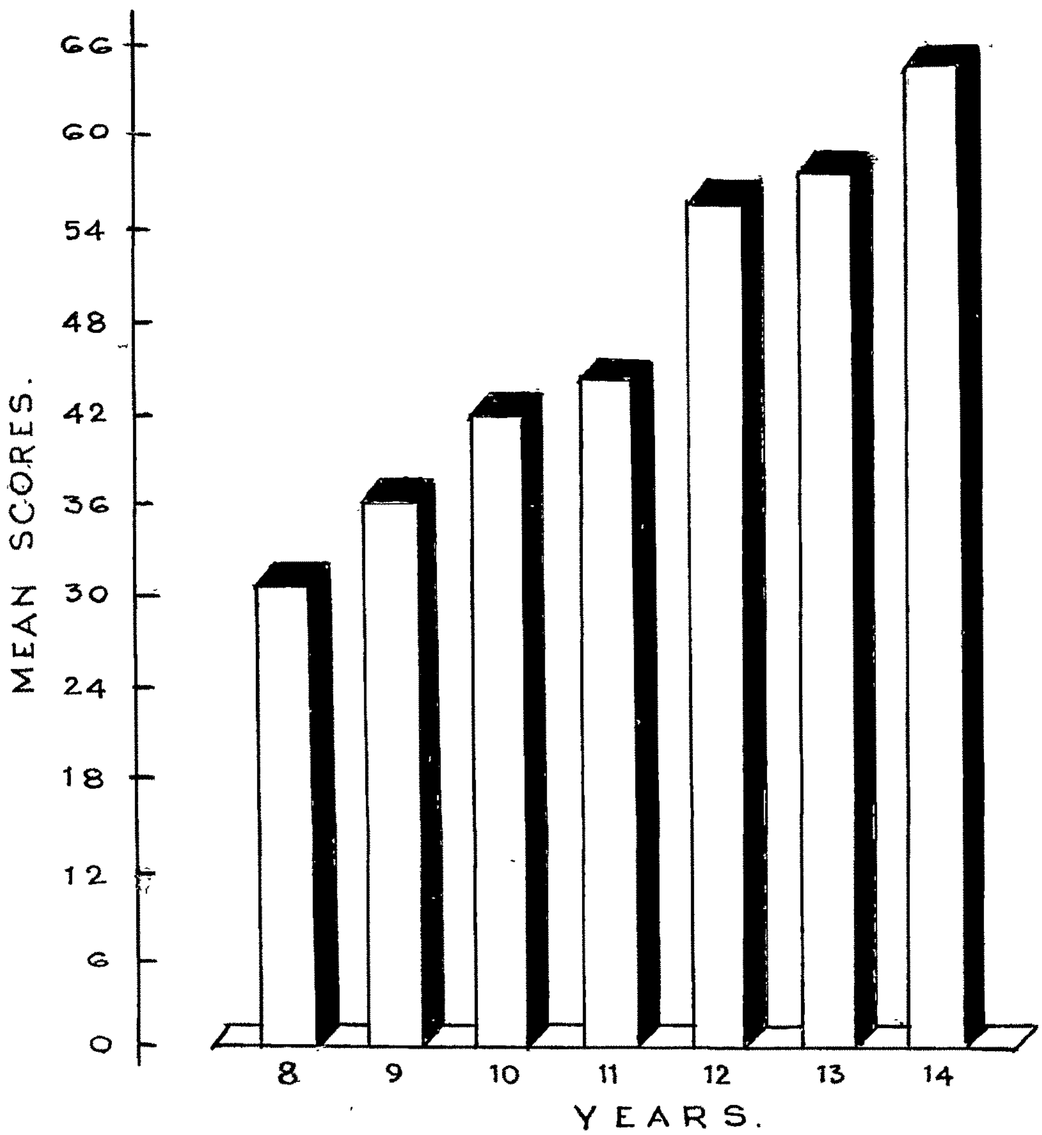
TABLE 35

Significance of Difference Between Means of Different Ages

Between ages	Age	Mean	S.D.	Difference in Means	N	SED	Critical ratio	Significance
	8 years	29.96	14.8		950			
8-9				5.59		0.67	8.34	Significant at 0.01 level
	9 years	35.55	14.3		915			
	9 years	35.55	14.3		915			
9-10				6.09		0.66	9.23	Significant at 0.01 level
	10 years	41.64	14.6		962			
	10 years	41.64	14.6		962			
10-11				1.76		0.72	2.44	Significant at 0.05 level
	11 years	43.4	15.9		809			
	11 years	43.4	15.9		809			
11-12				11.1		0.8	13.9	Significant at 0.01 level
	12 years	54.5	15.8		718			

Between ages	Age	Mean	S.D.	Difference in Means	N	SED	Critical ratio	Significance
	12 years	54.5	15.8		718			
12-13				2.36		0.82	2.68	Significant at 0.01 level
	13 years	56.86	17.5		899			
	13 years	56.86	17.5		899			
13-14				7.64		0.63	12.13	Significant at 0.01 level
	14 years	64.5	15.1		784			

GRAPH SHOWING THE MEAN SCORES OF ALL THE AGES(8-14)



From the table and the graph, it is observed that there is a steady rise in the mean scores as we go from lower age to higher one. It is to be noted that there is not much difference in standard deviations of the scores of different ages. The critical ratios for different pairs of ages were referred to the table for testing significance. The table shows that the differences between the means of different ages were significant. In fine, it can be concluded from the above table that the test discriminates children from year to year.

The next step was to see whether the test discriminates children even half yearly. The following table gives the means and standard deviations for every half year. The results were treated for testing the significance of difference between the half yearly means once again.

TABLE 36

Significance of Differente Between Half-Yearly Means of Different Ages

Age	Half year in months	Mean	S.D.	N	Difference between Means	SE_D	Critical ratio	Significance
8 years	7- 6 to 7-11	27.62	14.21	435	2.47	0.98	2.52	Significant at 0.05 level
	7-12 to 8- 5	30.09	14.53	415				
9 years	8- 6 to 8-11	34.33	14.51	437	1.59	0.94	1.69	Not significant
	8-12 to 9- 5	35.92	14.6	478				
10 years	9- 6 to 9-11	39.69	15.19	452	2.61	1.01	2.58	Significant at 0.01 level
	9-12 to 10- 5	42.3	14.89	510				
11 years	10- 6 to 10-11	42.52	15.3	428	2.04	0.34	5.97	Significant at 0.01 level
	10-12 to 11- 5	44.56	15.2	381				

Age	Half-year in months	Mean	S.D.	N	Difference between Means	SE_D	Critical ratio	Significance
	11- 6 to 11-11	54.3	15.9	339				
12 years					0.81	1.19	0.68	Not significant
	11-12 to 12- 5	55.11	16.1	379				
	12- 6 to 12-11	55.93	16.3	403				
13 years					2.28	1.09	2.09	Significant at 0.05 level
	12-12 to 13- 5	58.21	16.2	496				
	13- 6 to 13-11	60.43	15.2	360				
14 years					4.45	1.08	4.1	Significant at 0.01 level
	13-12 to 14- 5	64.88	15.1	424				

From the table, it can be observed that the differences between half yearly means of different ages are significant except in cases of 9 and 12 years. This speaks for how well the test discriminates children even half yearly.

II. Reliability of the Test

A test is said to be reliable if repeated measurements would give us more or less similar results. Scores achieved on unreliable tests are neither stable nor trustworthy. Stability and trustworthiness depend upon the degree to which the score is an index of "true ability" - is free of chance error. The Stanford-Binet I.Q., for example, is known to be a dependable measure. Hence, if a child's I.Q. is reported to be 110 by a competent examiner, we feel confident that this "score" is a good estimate of the child's ability to handle tasks like those represented by the test. The correlation of the test with itself is called the reliability coefficient of the test.

There are four methods in common use for computing the reliability coefficient of a test. These are:

(i) Test-retest method.

(ii) Split-half method.

(iii) Method of Rational Equivalence.

(iv) Alternate or parallel forms methods.

All these methods furnish estimates of the reproducibility of test scores; sometimes one method and sometimes another will provide the better measure. With regard to the present test the first three methods were used for computing the self-correlation of the test.

(i) <u>Test-Retest Method</u>

To apply this method, 400 pupils of two schools were tested again after an interval of about four months from the date of the first testing. The two sets of scores in the two trials were then correlated and the coefficient of reliability was found to be 0.94.

TABLE 37

Reliability by Test - Retest Method N = (400)

Test scores of the second trial	Test scores of the first trial: 0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99	Total
90-99										2	2
80-89							1	2	9		12
70-79						7	11	22			40
60-69			1			23	43				67
50-59				3	31	66					100
40-49			5	6	83						94
30-39			7	39							46
20-29		6	19								25
10-19		10									10
0- 9	4										4
Total	4	16	32	48	114	96	55	24	9	2	400

$\therefore r = .94 \pm .0057$ (.99 confidence interval)

(ii) Split-half Method

To apply this method, the tests were split up into two halves, one containing only odd items and the other containing only even items. The test booklets of 400 pupils were selected at random for this purpose. From the reliability of the half test, the self correlation of the whole test was then estimated by the Spearman-Brown-Prophecy formula.

TABLE 38

Scattergram to Determine Reliability by the Split-half Method

Y variable even scores	X variable odd scores: 5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	Total
45-49									1	1
40-44									3	3
35-39							11	14	10	35
30-34					1	8	15	16		40
25-29				2	5	39	20			66
20-24			5	4	23	84				116
15-19		2	15	32	57					106
10-14		2	20	10						32
5- 9	1									1
Total	1	4	40	48	86	131	46	30	14	400

$r = 0.86$

After applying Spearman-Brown Prophecy formula $r = 0.92 \pm .019$

(.99 confidence interval)

(iii) Method of Rational Equivalence

The method stresses the inter-correlations of the items in the test and the correlations of the items with the test as a whole. The formula for determining test reliability is:

$$r_{11} = \frac{n}{(n-1)} \times \frac{\sigma^2 t - \Sigma pq}{\sigma^2 t}$$

in which r_{11} = reliability coefficient of the whole test.

n = number of items in the test.

σt = the S.D. of the test.

p = proportion of the group answering a test item correctly.

q = (1-p) = the proportion of the group answering a test item incorrectly.

TABLE 39

Reliability by Rational Equivalence Methods
(N = 400)

Subtest	Item No.	A	P=A/400	Q = 1-p	PQ
Similarity	1	300	.75	.25	.1875
	2	279	.697	.303	.2112
	3	271	.677	.323	.2187
	4	275	.687	.313	.2150
	5	260	.65	.35	.2275

Subtest	Item No.	A	P=A/400	Q=1-p	PQ
	6	269	.673	.327	.3201
	7	289	.722	.278	.2007
	8	242	.585	.415	.2428
	9	235	.587	.413	.2424
	10	235	.587	.413	.2424
	11	194	.485	.515	.2498
	12	181	.452	.548	.2477
	13	179	.447	.553	.2472
	14	150	.375	.625	.2344
Classification	1	311	.777	.223	.1733
	2	300	.75	.25	.1875
	3	291	.727	.273	.1985
	4	265	.662	.338	.2229
	5	274	.685	.315	.2158
	6	236	.59	.41	.2419
	7	193	.482	.518	.2497
	8	215	.537	.463	.2486
	9	221	.552	.448	.2473
	10	180	.45	.55	.2475
	11	115	.287	.713	.2046
	12	125	.312	.688	.2147
	13	100	.25	.75	.1875
	14	85	.212	.788	.1704

Subtest	Item No.	A	P=A/400	Q=1-p	PQ
Analogy	1	280	.7	.3	.21
	2	249	.622	.378	.2413
	3	230	.575	.425	.2444
	4	198	.495	.505	.2500
	5	189	.472	.528	.2492
	6	157	.377	.623	.2349
	7	121	.302	.698	.2108
	8	97	.242	.758	.1834
	9	91	.227	.773	.1755
	10	102	.255	.745	.1900
	11	80	.2	.8	.16
	12	75	.187	.813	.1520
	13	74	.185	.815	.1508
	14	74	.185	.815	.1508
	15	68	.17	.83	.1411
	16	71	.177	.823	.1457
	17	62	.155	.845	.1310
Absurdity	1	301	.752	.248	.1865
	2	305	.762	.238	.1813
	3	286	.715	.285	.2038
	4	281	.702	.298	.2092
	5	283	.707	.293	.2071
	6	280	.7	.3	.21

Subtest	Item No.	A	P=A/400	Q=1-p	PQ
	7	259	.647	.353	.2284
	8	268	.67	.33	.2211
	9	235	.587	.413	.2424
	10	223	.557	.443	.2467
	11	225	.562	.438	.2462
	12	198	.495	.505	.2500
	13	201	.502	.498	.2500
	14	161	.402	.598	.2405
	15	150	.375	.625	.2344
	16	153	.382	.618	.2361
	17	149	.372	.628	.2336
	18	131	.327	.673	.2201
Progressive series	1	280	.7	.3	.21
	2	284	.71	.29	.2059
	3	281	.702	.298	.2092
	4	263	.657	.343	.2253
	5	259	.647	.353	.2284
	6	235	.587	.413	.2424
	7	230	.575	.425	.2444
	8	235	.587	.413	.2424
	9	210	.525	.476	.2494
	10	207	.517	.483	.2497
	11	198	.495	.505	.2500

Subtest	Item No.	A	P=A/400	Q=1-p	PQ
	12	183	.457	.543	.2478
	13	187	.467	.533	.2489
	14	151	.377	.623	.2359
	15	109	.272	.728	.1980
	16	89	.222	.778	.1727
	17	89	.222	.778	.1727
	18	80	.2	.8	.16
				ΣPQ	17.5500

$$r_{11} = \frac{n}{n-1} \times \frac{\sigma^2 t - \Sigma pq}{\sigma^2 t}$$

where r_{11} = reliability coefficient of the whole test.

n = number of items in the test.

σt = the S.D. of the test scores.

p = the proportion of the group answering a test item correctly.

q = $(1-p)$ = the proportion of the group answering a test item incorrectly.

$$r_{11} = \frac{80}{79} - \left(1 - \frac{17.55}{(18.5)^2}\right)$$

$$= \frac{80}{79} - (1 - 0.051)$$

$$= .961 \pm 0.0098 \text{ (.99 confidence interval)}$$

It should be moted here that the higher coefficients of correlation in the split-half technique and rational equivalence method should be looked upon with caution. Speed tests are tests in which the time limit imposed is so short that usually not all examinees can attempt all of the items. Speed tests such as this are of low difficulty level. In fact, the odd-even split-half reliability procedure is vitiated by the element of speed.

As Garrett puts it:

> The split-half technique and the rational-equivalence methods should not be employed with speed tests.[1]

There are other measures of reliability which may be employed to corroborate and reinforce the indications of these two methods. As regards the present test the reliability coefficient as computed by the test-retest method is significantly high and it provides a reason why the indications of the other two methods should not be suspected.

III. Validity of the Test

The validity of a test depends upon the fidelity with which it measures what it purports to measure. The

1 Garrett, H.E., Statistics in Psychology and Education, Allied Pacific Private Ltd., p. 353.

present test is expected to measure intelligence and hence it should be made sure whether it actually does so. There are two main types of evidence bearing on the validity of a test: (i) rational and (ii) empirical or statistical.

Sometimes we judge the validity of a test by rational analysis. This analysis may be of the areas included in the test - its content. This type of analysis is done to ascertain the content validity of the test. We may also analyse the functions or processes it measures to see how well they correspond to the concept that we have set out to appraise. This analysis is done to ascertain the concept validity of the test. In the words of Thorndike and Hagen:

> In practice, establishing the content and the concept validity of a test are often closely interwoven.[1]

In the present investigation such rational analysis was done while selecting the subtests and constructing the items. The detailed account of the analysis is given in the chapter on 'Planning and Preparing the test'.

(ii) The second type of evidence of validity is

---

2 Thorndike and Hagen: Measurement and Evaluation in Psychology and Evaluation, New York, John Wiley and Sons, Inc., 1955, p. 114.

empirical or statistical. This type of evidence comes from the relationship of the instrument to some other measure. Sometimes we may gather statistical evidence of validity. The statistical evidence will usually be in the form of correlations with other measures. The correlation of a test with an existing similar measure of the same function provides evidence on congruent validity of the test. The correlation of a test with some other measure obtained at the same time provides evidence on the concurrent validity of the test.

In the present investigation empirical (or statistical) validity was determined by:

(a) correlating the test with the intelligence test of the Faculty of Education and Psychology, Baroda;

(b) correlating the test with the total of the examination scores of the four subjects (Science, Mathematics, English and Gujarati); and

(c) correlating the test with the teachers' estimates of intelligence.

(a) Correlation of the Present Test with the Verbal Test of the Faculty of Education and Psychology, Baroda

The Research Department of the Faculty of Education and Psychology, Baroda, has prepared a verbal group test of intelligence under the guidance of Lele. The present test scores were correlated with the scores of that test. The following table shows the correlation between the two sets of scores.

TABLE 40

Correlation of the Scores of the Present Test with the Verbal Test

Scores on the Verbal Test	Scores on the Non-verbal Test: 0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	Total
90-99	-	-	-	-	-	-	2	-	-	2
80-89	-	-	-	-	-	3	3	-	-	6
70-79	-	-	-	3	6	6	3	-	-	18
60-69	-	2	8	6	15	9	5	3	-	48
50-59	-	-	17	23	29	23	2	1	2	97
40-49	-	3	18	15	14	3	1	1	-	55
30-39	-	4	17	8	4	3	-	-	-	36
20-29	4	16	14	-	-	-	-	-	-	34
10-19	6	15	8	-	-	2	-	-	-	31
0- 9	-	-	3	-	-	-	-	-	-	3
Total	10	40	85	55	68	49	16	5	2	330

$r = 0.7 \pm .072$ (.99 confidence interval)

(b) Correlation of IQs with Examination Marks

Scores of four subjects viz: Science, Mathematics, English and Gujarati were taken instead of taking the aggregate score at the annual examination which includes scores of physical training and other subject as well. Moreover, the scores in these subjects were converted into standard scores with 100 as the mean; and 20 as the standard deviation so that the marks of the schools might be comparable. The following table shows the correlation of the IQs with the standard scores.

TABLE 41

Correlation of IQs with the Total of Examination Scores of Four Subjects (Science, Mathematics, English and Gujarati)

Total scores in Z scores	I.Q.s 70-79	80-89	90-99	100-109	110-119	120-129	130-139	Total
155-174				1	5	3		9
135-154			30	12	4	3		49
115-134		17	80	63	10			170
95-114	11	21	15	8				55
75- 94	5	14	26					45
55- 74	2							2
Total	18	52	151	84	19	6		330

$\therefore$ $r = 0.55 \pm 0.09$ (.99 confidence interval)

(c) Correlation of IQs with Teachers' Estimates of Intelligence

The validity of the test was tested by correlating IQs of pupils with teachers' estimates of their intelligence. The teachers were instructed to classify the pupils into the following categories:

A. Very superior
B. Superior
C. Average
D. Inferior
E. Very inferior.

The following table shows the correlation of IQs with teachers' estimates of intelligence:

TABLE 42

Correlation of IQs with Teachers' Estimates of Intelligence

Teachers' estimates of intelligence	IQs 70-79	80-89	90-99	100-109	110-119	120-129	Total
A	-	-	-	1	4	2	7
B	-	-	32	12	5	4	53
C	-	17	75	64	10	-	166
D	12	23	17	7	-	-	59
E	6	12	27	-	-	-	45
Total	18	52	151	84	19	6	330

$r = 0.53 \pm 0.01$ (.99 confidence interval)

From what has been discussed above it can be concluded that the present test is a fairly reliable and valid instrument of measuring intelligence of the pupils of the prescribed age limit. The above correlational evidence are indicative to the extent to which the present test is dependable.

Factor Analysis

In the statistical method called factor analysis, the intercorrelations of subtests are examined and if possible accounted for in terms of a much smaller number of more general "factors" or trait categories. The factor presumably run through the often complex abilities measured by the individual tests. This is a specialized mathematical technique widely used and highly important in modern test construction.

According to Cronbach, "It is hard to gain even a partial understanding of factor analysis."[3] Although, the treatment is mathematical the technique involves considerable judgement. In the present case, the centroid method developed by Thurstone was used. For this purpose, the internal correlations of the 6 subtests calculated for the final test have been used. The sample selected for this was 1000. The results of the factor analysis

3 Cronbach, Essentials of Psychological Testing, New York, Harper & Brothers, Publishers, 1960, p. 247.

were verified by Spearman's formula for 'g' saturation.

TABLE 43

Original Correlation Matrix (N = 1000)

	1	2	3	4	5	6
1	-	.51	.60	.29	.41	.37
2	.51	-	.59	.48	.56	.51
3	.60	.59	-	.45	.56	.50
4	.29	.48	.45	-	.35	.32
5	.41	.56	.56	.35	-	.48
6	.37	.51	.50	.32	.48	-

In the correlation matrix rewritten below, the diagonal cells are filled in with the highest correlation of respective columns.

TABLE 44

Rearranged Correlation Matrix (N = 1000)

	1	2	3	4	5	6	
1	(0.60)	.51	.60	.29	.41	.37	
2	.51	(0.59)	.59	.48	.56	.51	
3	.60	.59	(0.60)	.45	.56	.60	
4	.29	.48	.45	(0.48)	.35	.32	
5	.41	.56	.56	.35	(0.56)	.48	
6	.37	.51	.50	.32	.48	(0.51)	
j1	2.18	2.65	2.70	1.89	2.36	2.18	= 13.96
tj1	2.78	3.24	3.30	2.37	2.92	2.69	= 17.30
aj1	0.6684	0.7790	0.7934	0.5698	0.7020	0.6467	= 4.1593

$$T_1 = 17.30$$

$$\sqrt{T_1} = 4.1593$$

$$\frac{1}{\sqrt{T_1}} = 0.2404251$$

$$\Sigma aj1 = 4.1593$$

TABLE 45

First Residual Correlation Matrix

Test No.	1	2	3	4	5	6	
	.6684	* .7790	.7934	* .5698	* .7020	* .6467	
1	.0909 (+.1532)	± .0107	.0697	± .0909	± .0592	- .0623	
* 2	±.0107	.0361 (- .0168)	±.0281	+ .0361	+ .0131	+ .0062	
3	+.0697	± .0281	.0697 (-.0295)	- .0020	∓ .0030	± .0131	
*4	±.0909	+ .0361	±.0020	.0909 (+ .1553)	- .0500	- .0485	
*5	±.0592	+ .0131	∓.0030	- .0500 (+ .0672)	.0592	+ .0260	
*6	±.0623	+ .0062	±.0131	- .0485	+ .0260	.0623 (+ .0918)	
Col. 4	+.0284	- .0556	.0335	.1553	+ .0329	+ .0053 =	+ 0.1998
2	.0498	.0556	.0897	.2275	.0067	- .0071 =	.4222
6	.1744	.0680	.1159	.1305	- .0453	.0071 =	.4506
5	.2928	.0942	.1099	.0305	.0453	.0591 =	.6318
tj2	.3837	.1303	.1796	.1214	.1045	.1214	1.0409
aj2	.3761	.1277	.1760	.1190	.1024	.1190	

T2 = 1.0409 $\sqrt{T2}$ = 1.0203 $\frac{1}{\sqrt{T2}}$ = 0.9801039 $\Sigma aj2$ = 1.0202

TABLE 46

Second Residual Correlation Matrix

	1	* 2	* 3	* 4	5	6	
Test No.	.3761	.1277	.1760	+.1190	.1024	.1190	
1	.0461 (-.0506)	±.0373	-.0035	-.0461	.0207	.0175	
* 2	±.0373	.0373 (.0198)	.0056	.0209	-.0000	±.0090	
* 3	-.0035	.0056	.0210 (.0387)	-.0189	±.0210	±.0078	
* 4	-.0461	.0209	-.0189	+.0627 (.0767)	±.0622	±.0627	
5	.0207	-.0000	±.0210	±.0622	.0622 (.0487)	.0138	
6	.0175	±.0090	±.0078	±.0627	.0138	.0627 (.0481)	
	-.0001	.0000	+.0001	-.0001	.0000	-ø.0001	
j3	.0505	-.0198	-.0386	-.0768	-.0487	-.0482	- .1816
Col. 4	-.0417	-.0616	-.0008	.0768	.0757	.0772	.1256
Col. 2	+.0329	+.0616	-.0120	.1186	.0757	.0952	+ .3720
Col. 3	.0259	+.0728	.0120	.0808	.1177	.1108	.4200
•	.0720	.1101	.0330	.1435	.1799	.1735	.7120
aj3	.0853	.1305	.0391	.1701	.2132	.2056	.8438

$T3 = .7120 \quad \sqrt{T3} = .8438 \quad \frac{1}{\sqrt{T3}} = 1.1851149 \quad \Sigma aj3 = 0.8438$

TABLE 47

Third Residual Correlation Matrix

Test No.	.0853 1	.1305 2	.0391 3	.1701 4	.2132 5	.2056 6	
1	.0606 (+.0338)	+.0262	-.0068	-.0606	+.0025	.0000	
2	+.0262	.0278 (+.0203)	+.0005	-.0013	-.0278	-.0178	
3	-.0068	+.0005	.0255 (+.0195)	-.0255	+.0127	-.0002	
4	-.0606	-.0013	-.0255	.0606 (+.0338)	+.0259	+.0277	
5	+.0025	-.0278	+.0127	+.0259	.0300 (+.0167)	-.0300	
6	.0000	-.0178	-.0002	+.0277	-.0300	.0300 (+.0205)	
j4	+.0001	+.0001	+.0002	.0000	.0000	+.0002	
Col.1	-.0387	-.0202	-.0193	-.0338	-.0167	-.0203	-.1490
Col.1	.0387	-.0726	-.0057	+.0874	-.0217	-.0203	+.0058
Col.2	+.0911	+.0726	-.0067	+.0900	+.0339	+.0153	+.2962
Col.3	.0775	.0736	+.0067	+.1410	.0085	+.0157	+.3230
tj4	.1381	.1014	.0322	.2016	.0385	.0457	
aj4	.1850	.1358	.0431	.2700	.0516	.0612	

$T4 = .5575 \quad \sqrt{T4} = .7466 \quad \frac{1}{\sqrt{T4}} = 1.3392977 \quad \Sigma aj4 = .7467$

TABLE 48

Centroid Factor Matrix

No.	Subtest	Factor I	Factor II	Factor III
1	Similarity	0.54	.48	.26
2	Classification	0.72	.00	.36
3	Analogy	0.73	.31	.19
4	Absurdity	0.51	-.02	.33
5	Progressive series	0.74	.02	.01
6	Substitution	0.69	-.01	.00

It can be seen from the table that there is one common factor (say 'g'). The second factor common to test first and third is not exactly identified and hence no attempt is made to specify it. The third factor common to first four tests can be identified as eduction of correlates. The next step will be to see how the first general factor common to all the subtests corresponds to 'g' saturation found out by Spearman's method.

<u>Spearman's Method of Determining 'g' Saturations of the Tests</u>

'g' saturations of the tests may be directly determined by the formula:[4]

4 Fruchter Benjamin, <u>Introduction to Factor Analysis</u>, New York, D.Van Nostrand Company, Inc., 1954, p.9.

$$g = \sqrt{\frac{Aj^2 - A'j}{T - 2Aj}}$$

where T = Sum of all the correlations (where each occurs twice and the diagonal cells are empty);

Aj = Sum of all the correlations in raw j;

$A'j$ = Sum of the squared correlations in row j;

Aj^2 = the square of Aj.

Step 1

Calculation of Aj^2

Original Matrix

	1	2	3	4	5	6	Aj^2
1	-	.51	.60	.29	.41	.37	4.7524
2	.51	-	.59	.48	.56	.51	7.0225
3	.60	.59	-	.45	.56	.50	7.29
4	.29	.48	.45	-	.35	.32	3.5721
5	.41	.56	.56	.35	-	.48	5.5696
6	.37	.51	.50	.32	.48	-	4.7524
	2.18	2.65	2.70	1.89	2.36	2.18	T=13.96

Step 2

Calculation of A'j

	1	2	3	4	5	6	A'j
1	-	.260	.360	.084	.168	.137	1.009
2	.260	-	.348	.230	.314	.260	1.412
3	.360	.348	-	.202	.314	.250	1.474
4	.084	.230	.202	-	.122	.102	.740
5	.168	.314	.314	.122	-	.230	1.148
6	.137	.260	.250	.102	.230	-	.979

TABLE 49

Calculation of 'g' Saturation

Test No.	Aj^2	A'j	T	2Aj	$\frac{Aj^2 - A'j}{T - 2Aj}$	'g' saturation
1	4.7524	1.009	13.96	4.36	.3899	.6244
2	7.0225	1.412	13.96	5.3	.6478	.8048
3	7.29	1.474	13.96	5.4	.6792	.8241
4	3.5721	.740	13.96	3.78	.2782	.5274
5	5.5696	1.148	13.96	4.72	.4785	.6916
6	4.7524	.979	13.96	4.36	.3930	.6368

TABLE 50

Comparison of the 'g' SaturationsObtained by Thurstone's Centroid Method and Spearman's Method

No.	Test	'g' Saturation	
		By Thurstone's Centroid Method	Spearman's Method
1	Similarity	.54	.6244
2	Classification	.72	.8048
3	Analogy	.73	.8241
4	Absurdity	.51	.5271
5	Progressive series	.74	.6916
6	Substitution	.69	.6368

It will be seen from the table that the 'g' saturations of the tests calculated by both the methods are almost the same. Furthermore, they are quite high showing that the tests are good measures of intelligence.

IV. Fixing the Norms

In order to compare any two pupils on the test, we need a standard. Instead of putting an arbitrary pass-mark, we evolve a relative standard of performance from the results of this large group of subjects. Such standards are called norms. By norms are meant specimens of work which represent the commonest type of achievement for the whole group in question. They constitute the means by which the degrees of abnormality, shown by testees above or below the normal, can be measured. It is generally recognised that valid norms are essential for dependable interpretations of individual and group measures.

Generally two types of norms are established for intelligence tests:

(i) Age norms;
(ii) Grade norms.

Sex norms, occupational norms and norms for rural and urban areas are also sometimes determined to show the effects of varying environments. But age-norms are the most useful in intelligence testing. In the present case, age norms have been prepared by the method of indirect standardization. The steps followed were as follows:

(1) Average scores were computed half-yearly for all the ages. The mean scores from quarter of an year to quarter of an year were not significantly different.

(2) The standard deviations were also computed half-yearly for all the ages.

(3) The arbitrary range fixed for IQs is 60-140 i.e. 100 ± 40. If the mean average of an

age score distribution is equated at 100 IQ, then the maximum score that can be reached by pupils is 40 unit scores on either side i.e. 100 ± 40 unit scores. IQs are assumed to be normally distributed over a range of 3σ on either side. So, the unit score will be $40/3\sigma$

(4) This standard scatter is superimposed on the scatter of various age groups. Hence the formula for deriving IQs is:

$$IQ * = 100 - (M - S) \; 40/3\sigma$$

where M is the mean of age group,
S is the score obtained on the test.

The credit of first introducing this method goes to Prof. T.P.Lele of the Faculty of Education and Psychology, Baroda and it has been found to give good results. He has prepared the verbal group test of intelligence and norms for the test have been established by using the same method. By following the steps listed above, the ready reckoner showing IQs corresponding to various scores of different age-groups has been prepared as shown below:

TABLE 51

Ready Reckoner for IQs

Age-Group (8-14+) (Half Yearly)

Age	8		9		10		11		12		13		14	
In years and months	7- 6 to 7-11	7-12 to 8- 5	8- 6 to 8-11	8-12 to 9- 5	9- 6 to 9-11	9-12 to 10- 5	10- 6 to 10-11	10-12 to 11- 5	11-6 to 11-11	11-12 to 12- 5	12- 6 to 12-11	12-12 to 13- 5	13- 6 to 13-11	13-12 to 14- 5
Score														
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	75	74	69	68	65	63	64	62	55	55	55	53	47	44
2	76	74	70	69	66	64	65	63	56	56	56	54	48	45
3	77	75	71	70	67	65	66	64	57	57	57	55	49	46
4	78	76	72	71	68	66	66	65	58	58	57	56	50	46
5	79	77	73	72	69	67	67	66	59	59	58	56	51	47
6	80	78	74	73	70	68	68	66	59	59	59	57	52	48
7	81	79	75	74	71	69	69	67	60	60	60	58	53	49
8	82	81	76	75	72	69	70	68	61	61	61	59	54	50
9	83	82	77	76	73	70	71	69	62	62	62	60	54	51
10	84	83	78	76	74	71	72	70	63	63	62	60	55	52
11	85	84	79	77	74	72	73	71	64	63	63	61	56	53
12	85	85	79	78	75	73	73	72	64	64	64	62	57	53

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
13	86	85	80	79	76	74	74	73	65	65	65	63	58	54
14	87	86	81	80	77	75	75	73	66	66	66	64	59	55
15	88	87	82	81	78	76	76	74	67	67	66	65	60	56
16	89	88	83	82	79	77	77	75	68	68	67	65	61	57
17	90	89	84	83	80	77	78	76	69	68	68	66	61	58
18	91	90	85	84	81	78	79	77	70	69	69	67	62	59
19	92	91	86	85	82	79	80	78	70	70	70	68	63	60
20	93	92	87	86	82	80	80	79	71	71	71	69	64	61
21	94	93	88	86	83	81	81	80	72	72	71	69	65	61
22	95	94	89	87	84	82	82	80	73	73	72	70	66	62
23	96	95	90	88	85	83	83	81	74	73	73	71	67	63
24	97	95	91	89	86	84	84	82	75	74	74	72	68	64
25	98	96	91	90	87	85	85	83	75	75	75	73	68	65
26	98	97	92	81	88	85	86	84	76	76	75	74	69	66
27	99	98	93	92	89	86	86	85	77	77	76	74	70	67
28	100	99	94	93	90	87	87	86	78	78	77	75	71	68
29	101	100	95	94	90	88	88	86	79	78	78	76	72	68
30	102	101	96	95	91	89	89	87	80	79	79	77	73	69
31	103	102	97	96	92	90	90	88	80	80	80	78	74	70
32	104	103	98	96	93	91	91	89	81	81	80	79	75	71
33	105	104	99	97	94	92	92	90	82	82	81	79	76	72
34	106	105	100	98	95	93	93	91	83	83	82	80	76	73
35	107	105	101	99	96	94	93	92	84	83	83	81	77	74
36	108	106	102	100	97	94	94	93	85	84	84	82	78	75
37	109	107	102	101	98	95	95	93	85	85	84	83	79	75
38	110	108	103	102	98	96	96	94	86	86	85	83	80	76
39	111	109	104	103	99	97	97	95	87	87	86	84	81	77
40	112	110	105	104	100	98	98	96	88	88	87	85	82	78
41	112	111	106	105	101	99	99	97	89	88	88	86	83	79
42	113	112	107	106	102	100	100	98	90	89	89	87	83	80

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
43	114	113	108	106	103	101	100	99	91	90	89	88	84	81
44	115	114	109	107	104	102	101	100	91	91	90	88	85	82
45	116	115	110	108	105	102	102	100	92	92	91	89	86	83
46	117	116	111	109	105	103	103	101	93	93	92	90	87	83
47	118	116	112	110	106	104	104	102	94	93	93	91	88	84
48	119	117	113	111	107	105	105	103	95	94	94	92	89	85
49	120	118	114	112	108	106	106	104	96	95	94	92	90	86
50	121	119	114	113	109	107	107	105	96	96	95	93	90	87
51	122	120	115	114	110	108	107	106	97	97	96	94	91	88
52	123	121	116	115	111	109	108	106	98	98	97	95	92	89
53	124	122	117	116	112	110	109	107	99	98	98	96	93	90
54	125	123	118	117	113	110	110	108	100	99	98	97	94	91
55	125	124	119	117	113	111	111	109	101	100	99	97	95	91
56	126	125	120	118	114	112	112	110	101	101	100	98	96	91
57	127	126	121	119	115	113	113	111	102	102	101	99	97	92
58	128	126	122	120	116	114	113	112	103	103	102	100	98	93
59	129	127	123	121	117	115	114	113	104	103	103	101	98	94
60	130	128	124	122	118	116	115	113	105	104	103	101	99	95
61	131	129	125	123	119	117	116	114	106	105	104	102	100	96
62	132	130	125	124	120	118	117	115	106	106	105	103	101	96
63	133	131	126	125	120	118	118	116	107	107	106	104	102	97
64	134	132	127	126	121	119	119	117	108	108	107	105	103	98
65	135	133	128	126	122	120	120	118	109	108	107	106	104	99
66	136	134	129	127	123	121	120	119	110	109	108	106	105	100
67	137	135	130	128	124	122	121	120	111	110	109	107	105	101
68	138	136	131	129	125	123	122	120	112	111	110	108	106	102
69	138	136	132	130	126	124	123	121	112	112	111	109	107	103
70	139	137	133	131	127	125	124	122	113	113	112	110	108	104
71	140	138	134	132	128	126	125	123	114	113	113	110	109	104
72	141	139	135	133	129	126	126	124	115	114	113	111	110	105
73	142	140	136	134	129	127	127	125	116	115	114	112	111	106

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
74	143	141	137	135	130	128	127	126	117	116	115	113	112	107
75	144	142	137	136	131	129	128	126	117	117	116	114	113	108
76	145	143	138	136	132	130	129	127	118	118	117	115	113	109
77	146	144	139	137	133	131	130	128	119	118	117	115	114	110
78	147	145	140	138	134	132	131	129	120	119	118	116	115	111
79	148	146	141	139	135	133	132	130	121	120	119	117	116	111
80	149	146	142	140	136	134	133	131	122	121	120	118	117	112
81	150	147	143	141	137	134	133	132	122	122	121	119	118	113
82	151	148	144	142	138	135	134	133	123	123	122	120	119	114
83	152	149	145	143	138	136	135	133	124	123	123	121	119	115
84	152	150	146	144	139	137	136	134	125	124	123	121	120	116
85	153	151	147	145	140	138	137	135	126	125	124	122	121	117
86	154	152	148	146	141	139	138	136	127	126	125	123	122	118
87	155	153	148	146	142	140	139	137	127	127	126	124	123	118
88	156	154	149	147	143	141	140	138	128	128	127	124	124	119
89	157	155	150	148	144	142	140	139	129	128	127	125	125	120
90	158	156	151	149	145	142	141	140	130	129	128	126	126	121
91	159	156	152	150	146	143	142	140	131	130	129	127	126	122
92	160	157	153	151	146	144	143	141	132	131	130	128	127	123
93	161	158	154	152	147	145	144	142	133	132	131	129	128	124
94	162	159	155	153	148	146	145	143	133	133	132	129	129	125
95	163	160	156	154	149	147	146	144	134	133	132	130	130	126
96	164	161	157	155	150	148	147	145	135	134	133	131	131	127
97	165	162	158	156	151	149	147	146	136	135	134	132	132	128
98	165	163	159	156	152	150	148	146	137	136	135	133	133	129
99	166	164	160	157	153	150	149	147	138	137	136	133	133	129
100	167	165	160	158	154	151	150	148	139	138	137	134	134	130
101	168	166	161	159	154	152	151	149	140	139	137	135	135	131
102	169	166	162	160	155	153	152	150	141	139	138	136	136	132

Decile Points of Raw Scores

This is the graphic way of fixing the point of reference. Here the whole range of distribution is divided into as many percentile groups as deemed feasible. These percentiles give fairly reliable picture of the distribution of scores. For example, we may calculate certain percentiles as 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th and 100th percentiles.

Suppose the 30th percentile comes out to be 21.8, it indicates that the persons securing 21.8 are better than 30 per cent of the group or in other words are inferior to 70 per cent of the group. If a pupil of 8 years secures 40, then to decide his place on the percentile scale, we will compare his score with the near about percentile scores. This will give the idea of relative standing of a pupil in a particular age-group. In the table below are given the age-wise decile points. It can be seen from the chart that there is no overlapping of scores and the distinctions are clear from one decile point to the other for each age-group.

TABLE 52

Decile Points of Raw Scores

Decile point	8 years	9 years	10 years	11 years	12 years	13 years	14 years
90	51.2	55.1	61.5	62.9	73.2	78.5	82.2
80	39.9	44.5	53.7	56.8	67.8	72.6	77.4
70	36.0	43.4	48.3	52.7	64.5	67.1	74.1
60	32.3	38.9	46.1	48.7	61.2	61.7	70.9
50	29.1	36.9	41.9	44.7	57.3	57.9	67.2
40	25.3	32.2	38.5	40.6	52.8	53.3	63.3
30	21.8	28.3	34.8	35.1	47.9	50.9	59.3
20	17.5	22.7	30.3	31.9	41.8	43.5	51.5
10	11.3	15.6	22.1	22.5	31.0	31.2	41.8

Classification of Pupils

The widespread application of intelligence tests to a large number of children and adults has established quite clearly the range of intelligence to be expected in the population at large. In statistical language intelligence is said to be distributed normally i.e. in accordance with the normal probability curve. Terman has classified the children in his investigation into the following categories:

Range of I.Q.	Category	Percentage of general population
140 and above	Near genius or Genius	1.5
120 - 139	Superior or Very bright	11.0
110 - 119	Bright	18.0
90 - 109	Average or Normal	48.0
80 - 89	Dull Normal or Backward	14.0
70 - 79	Border line or Very dull	5.0
0 - 69	Feeble-Minded	2.5

The Feeble-Minded are further divided into three categories:

Class	Range of I.Q.	Mental age	Remarks
Morons	50 to 70	8 to 11	To be shifted to special schools for mentally deficient
Imbeciles	20 to 50	3 to 8	To be engaged in occupational centres
Idiots	Below 20	Below 3	Absolutely ineducable. Need only to be protected.

Desai has classified the secondary school population of Gujarat in the following way:

Near Genius or Genius	140 and above
Extra ordinary	130-139
Very Superior	120-129
Superior	110-119
Normal	90-109
Backward	80- 89
Very Backward	70- 79
Borderline Deficiency	Below 70

In the present case the pupils were classified into the following five categories:

Category	Range of I.Q.
Very superior	140 and above
Superior	116 - 139
Normal	85 - 115
Dull	70 - 84
Borderline Deficiency	Below 70

Classifying the 6037 cases according to this classification we get the distribution of cases as under:

TABLE 53

Agewise Classification of Pupils
(N = 6037)

Age	Borderline Deficiency	Dull	Normal	Superior	Very superior
8	45	130	592	150	33
9	30	169	612	89	15
10	40	98	673	111	40
11	57	120	528	75	29
12	28	81	480	102	27
13	39	123	648	55	34
14	25	159	485	101	14
N=6037	264	880	4018	683	192

The percentage of cases would be as follows:

Very superior	3.18 per cent
Superior	11.31 per cent
Normal	66.55 per cent
Dull	14.57 per cent
Borderline Deficiency	4.37 per cent

Thus the present test can be profitably used for broad classification of pupils into the above categories.

To conclude, the present chapter comprises the complete account of the procedure followed for standardization of the test. The statistical analysis of the results given in this chapter reveals that the test is a dependable instrument of measuring intelligence of the pupils within the prescribed age group.

REFERENCES

Cronbach, L.J., Essentials of Psychological Testing, Harper & Brothers, Publishers, New York, 1960.

Desai, K. G., The Construction and Standardization of a Battery of Group Tests of Intelligence in Gujarati, Bharat Prakashan, Ahmedabad, 1954.

Downie, N.M., Fundamentals of Measurement, Oxford University Press, New York, 1958.

Fruchter Benjamin, Introduction to Factor Analysis, D.Van Nostrand Company, Inc., New York.

Garett, H.E., Statistics in Psychology and Education, McGraw Hill Book Co., New York, 1956.

Kamat, V.V., Measuring Intelligence of Indian Children, Oxford University Press, Indian Branch, Bombay, 1940.

Mehrotra, L.P. and Mehrotra, K., Mental Testing & Standardization of Tests, Ram Narain Lal, Allahabad.

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