CHAPTER VI

FINAL TRYOUT

Introduction.

"Test scores are meaningful and valuable to the extent that they can be interpreted in terms of capacities, abilities and accomplishments of educational significance".[1/] For this purpose there is a need of suitable procedures for describing, recording and comparing the performances of individuals in specific test situations. These procedures are mamed as standardization of a test.

"A standardized test is one in which the procedure, apparatus, and scoring have been fixed so that precisely the same test can be given at different times and places".[2/] In this sense the test has already been standardized on the basis of the results of the pilot tryout as described in the previous chapter.

Standardized tests provide norms which state what scores are earned by representative subjects. So there is a need of providing

---

1/Flanagan J.C., "Units, Scores and Norms". Chapter 17. Educational Measurement, Lindquist E.F. (Editor) American Council on Education, Washington D.C. 1966. pp. 695

2/Cronbach L.J., Essentials of Psychological Testing. Harper & Row, New York and John Weather hill, Ind., Tokyo. 1965. pp. 22

norms for this test. Broadly speaking, the steps involved in the process of establishing norms are as below:-

a. Administration of the test to a large representative sample of the population under uniform conditions

b. Scoring the test uniformly

c. Establishing the norms with suitable units.

Representative Sample

The best norms are those which have been established on the basis of the scores obtained by the whole population. But this procedure is not practicable. The other way is to establish them on the basis of the scores of the sample of the population. Sample is defined as, "A portion of a total so chosen that the characteristics of the whole may be judged from those of the part with a minimal degree of error".[1/] If the degree of error is to be minimum, the sample should have all the characteristics of population and to the extent to which they are possessed by the population. In short, it should be a representative sample. The reliability and validity of the results and of the interpretations based on these results will depend on the fact how far the sample selected is representative of the population. It also depends on the size of the sample. So the sample should be adequate and representative.

If it is to be representative it should be carefully chosen.

1/Goodenough F.L., Mental Testing. Holt, Rinehart and Winston, New York. 1961. pp 564.

Methods of Selecting the Sample:

Different techniques have been developed for this purpose. The commonly used methods are as follows:-

a. Random Sampling

b. Stratified Sampling

c. Incidental Sampling

d. Purposive Sampling.

a. Random Sampling.

"The criteria for randomness in a sample are met when (1) every individual (or animal or thing) in the population or supply has the same chance of being chosen for the sample; and (2) when the selection of one individual or thing in no way influences the choice of another".[1/] The selection of cases is done by some mechanical process like drawing lottery. Thus the selection of a case is merely a chance and is not affected by the whims or biases of the individual selecting it. This is the most reliable way of getting the representative sample.

b. Stratified Sampling.

This is a modified form of random sampling. When the population consists of groups of different sizes and if the proportions of the groups are to be maintained in the representative sample, this procedure is adopted. Though the size of each group in the sample is decided according to its size in the population, the

1/Garrett H.E., Statistics in Psychology and Education, Longmans, Green and Co., New York, London, Toronto. 1954. pp. 202.

selection of cases in each group is made by random sampling methods.

c. Incidental Sampling.

This procedure is applied to those groups which are readily available. Such groups rarely constitute random samples of any population. Generalization based on such samples may be misleading.

d. Purposive Sampling.

It is an arbitrarily selected sample because the available evidence shows that it is a representative sample of a population with reference to a given characteristic.

Size of the Sample:-

The reliability of mean, standard deviations and other statistics depend upon the standard errors (SE) of these statistics. SE varies inversely as the square root of the size of the sample. Thus if the above mentioned statistics are to be dependable, the size of the sample should be as large as possible.

The Present Sample.

This test is meant for the pupils of Standards VIII, IX and X of the Secondary Schools in Marathwada (Aurangabad Division). Thus the population of this study is the pupils of Standards VIII to X in the secondary schools of Marathwada.

The authentic information regarding this population may be obtained from the census report of 1971 or the statistical reports published by the Department of Education, Maharashtra State. The most recent statistical report "Sankyikeey Sankshep Granth" has been published in Marathi, by the Director of Education, Maharashtra in the year 1974, which contains statistics for the year 1970-71. This report gives information regarding the number of pupils in each age group, in each sex group and in each standard and district.

It was decided to standardize the test by taking a sample of about 10,000 pupils (about 2000 from each age group) from about 100 schools with an average of 120 pupils from each school.

Selection of Schools:

There are 46 talukas in the five districts of this region and it was decided to select at least one school from each taluka.

New schools are coming up in this region and the strength of these schools is very small. So while selecting the schools care was taken to see that at least one hundred pupils would be available for testing.

This area as mentioned earlier, is economically backward. There are many places with high-schools which can only be reached by walking a distance of 8 to 10 kilometers. Such schools were also not selected for testing.

The Maharashtra State Board of Secondary Education, Aurangabad Divisional Board, Aurangabad has prepared the list of Secondary Schools affiliated to the Board, district-wise. Five to six schools were selected by lottery from each taluka for testing.

Letters were written to the head-masters of these schools requesting them to permit testing in their schools. They were also requested to inform whether the place can be reached by S.T. bus and whether at least 100 pupils would be available for testing.

Only those schools which would be reached by S.T. bus or by train and where at least 100 pupils would be available were selected for testing.

Letters were again written to the headmasters of the selected schools, explaining them the nature of testing work and the nature of cooperation that was expected of them. The dates of testing were also decided according to the convenience of both the head master of the school and the administrator of the test.

So the sample is selected by applying neither purely random sampling nor stratified sampling procedures. However the principles of both the procedures have been broadly applied.

Table 20. Number of Schools from each district in which testing was done.

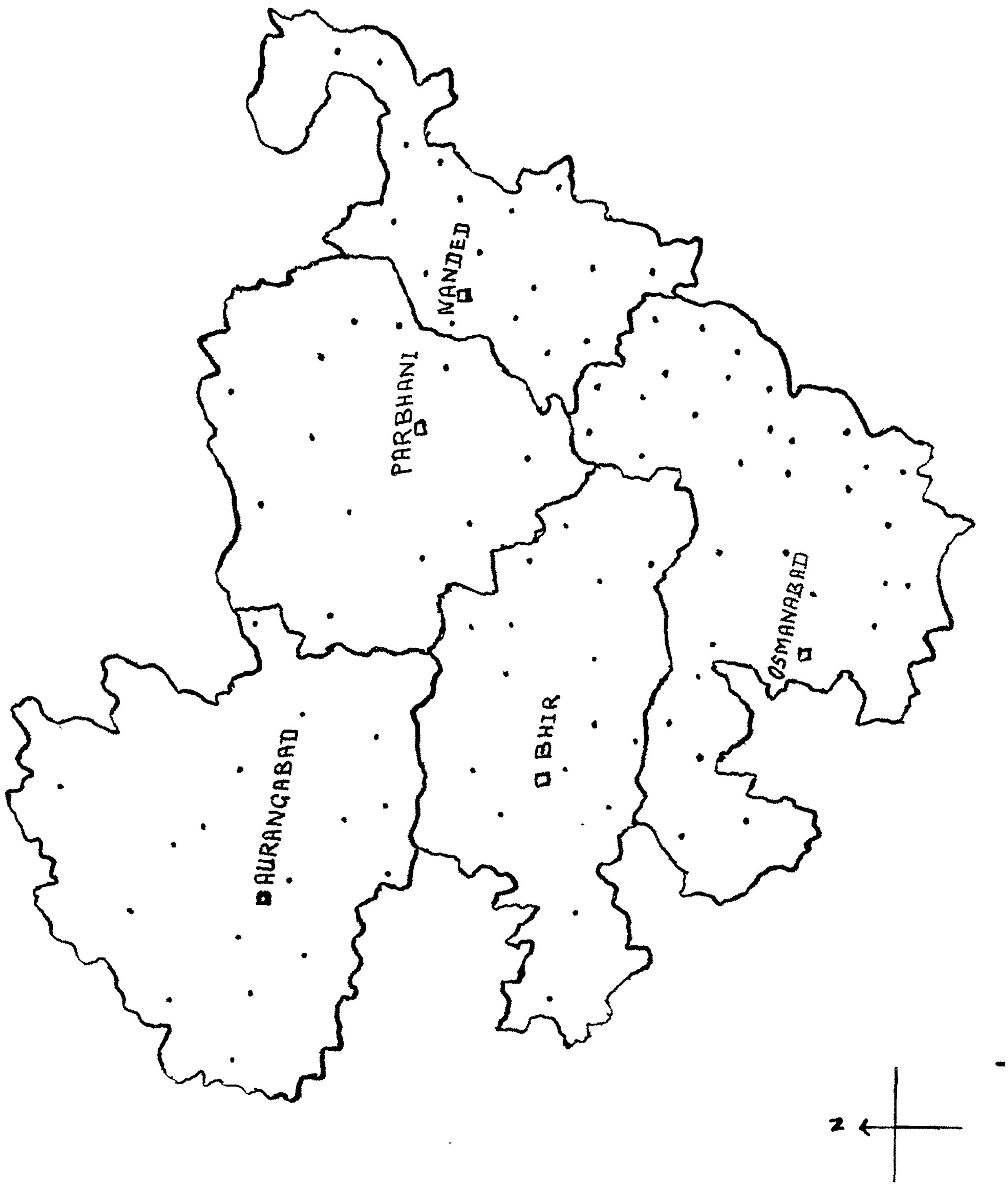
District		Number of Schools
Aurangabad	..	18
Parbhani	..	12
Bhir	..	18
Nanded	..	16
Osmanabad	..	23
Total	..	87

(For the names of the Schoolsplease refer to Apendix B)

Map of Marathwada

showing

The Places Where Testing Was Done



Administration of the Test.

As the author is serving in a College, it was not possible for him alone to administer the test in all the selected schools. So he had decided to take the cooperation of the lecturers in the colleges of education and also of the trained and experienced teachers in whose ability he had good faith. He had requested them to attend to at least two administrations of the test during the pilot try-out and to administer the test to atleast two classes under his supervision. Thus he had trained them to administer the test.

The administrator of the test was reaching the place one day earlier. He was meeting the head master and explaining him the procedure he was following in administering the test. He requested him to ask the respective class teachers to assist him in supervising the class while pupils were taking the tests. The test was administered in two sessions and strictly according to instructions finalised during the pilot try-out.

The class teachers had already told the pupils their datesof birth according to the school register. The pupils recorded them on the answer-sheets. The administrator compared these with the school register with the help of the class teachers and made corrections wherever necessary.

Experiences During Administration:

There were very scanty rains in this region during the three consecutive years 1970-71, 71-72 and 72-73 and thus this region was declared as the scarcity area. Many persons were leaving their villages and going to cities to find work. So the Government of Maharashtra started scarcity work in the viscinity of each village to check the migration of the people from the villages to the cities. The nature of the work was construction of water tanks, approach roads, widening of highways and preparing 'Khadi' for metalling of the roads. Even the children below 18 years of age were employed for this purpose and this had its own adverse impact on the attendance of the pupils in the schools especially in the rural areas.

Though no person died of starvation, they could not maintain the cattle they owned. They were either sold to the butchers or died of starvation. Thus during the year 1973-74, though there were adequate rains farmers had neither cattle nor seeds for cultivation. Though the Government sanctioned loans for these purposes, the amount was inadequate to meet their needs. The farmers faced the situation very bravely. They used children and woman to do the farm-work normally done by the bullocks. So the attendance in the schools was very low.

The testing of the pupils has been done during this period. In rural areas, though the strength of the school was above 200 the attendance in some schools on the day of testing was 50 or 60.

To add to this there was student agitation during the second term of the academic year 1973-74 (during which the testing was done). Most of the colleges were closed. The university had to postpone the examinations from March to July. The secondary school children in urban areas were holding demonstrations. So there was no guarantee of getting adequate pupils in schools on the day of testing.

Though the estimated availability of pupils from these 87 schools at an average of 120 per school from the three standards works out to be 10,440, the actual number available was less than 8,000.

The head masters, the teachers and the pupils in these schools gave good cooperation in the work. But at the same time about 200 head masters did not at all respond.

Description of the Tested Sample.

On the whole 8004 pupils were tested. Out of these 259 were either below 13 or above 17 and hence the performance of 7745 pupils was available for prescribing the norms.

Table 21. The number of pupils tested from each standard from the five districts.

District	Number of pupils tested from Standards			Total	Percentage in the Sample	Percentage Population
	VIII	IX	X			
1	2	3	4	5	6	7
Aurangabad	774	580	553	1907	24.62	25.00
Parbhani	486	322	299	1107	14.29	14.00
Bhir	554	399	257	1210	15.63	17.00

(concluded on next page)

Table 21. (concluded)

1	2	3	4	5	6	7
Nanded	650	398	275	1323	17.08	16.00
Osmanabad	881	633	684	2198	28.38	28.00
Total ..	3345	2332	2068	7745	100.00	100.00
Percentage in the sample	43.18	30.11	26.71	100		
Percentage in the population	45	30	25	100		

(Percentage in the population quoted in this table are adopted from Table No. 4, page 60).

From the above table it can be seen that the percentages of pupils in the sample tested from different standards is almost the same as these in the population. It also shows that the percentages of pupils tested from each district are almost the same as those in the population.

Table 22. Classification of pupils tested and their percentages.

Sex		Total Number tested.	Percentage	Percentage in the population
Boys	..	6,201	80.05	87.5
Girls	..	1,544	19.95	12.5

(Percentages in the population quoted in this table are adopted from Table No. 6, page 61)

The percentage of girls tested is slightly more in the sample than that in the population. The percentage of girls in the mixed schools is very low. There are separate girls' schools in many places and girls prefer to join these schools. To get adequate number of girls in the sample 2 to 3 girls' schools from each district were selected for testing. Moreover the percentage of girls attending the secondary schools is gradually increasing. For these two reasons the percentage of girls in the sample is slightly more than that in the population.

Table 23. Classification of pupils according to age.

Age	Number of pupils Tested	Percentage	Percentage in the population
13	1320	17.04	29.14
14	1825	23.57	23.97
15	2088	26.97	20.62
16	1484	19.15	15.82
17	1028	13.27	10.45
Total ..	7745	100.00	100.00

(Percentage in the population quoted in this table are adopted from Table No. 6, page 61)

The percentages show a variation. From table no. 4 on page no. 60 and table no. 6 on page no. 61 it can be seen that number of pupils studying in standards VIII to X is 1,37,980 and the number of pupils of the age groups from 13 to 17 years is 2,10,192. Thus out of 2,10,192 pupils of this age group, 72,212 pupils are studying in standards either below VIII or above X. Thus the percentages of pupils of age group in the population may not be the same as those in the pupils of standards VIII to X. Secondly this is likely to affect the percentages of extreme age groups namely 13 and 17 years.

New pattern of secondary education was introduced during the year 1971-72 and as said earlier standard XI was abolished from the secondary schools. The old X and XI were named as IX and X respectively. The pupils passing the old standards IX and X were placed in the new IX and X standards respectively. The pupils passing std. VII and VIII were retained in two different divisions of standard VIII (new and old). In one division old syllabus was followed and in the other the new one was introduced. Thus the average age of pupils in one division of standard VIII was one year more than that of those in the other. This batch had reached standard X in the academic year 1973-74 when the testing was done. This fact is probably responsible for increasing the percentages of pupils of age groups 15, 16 and 17 years.

If these two facts are taken into account, the variation observed in the sample is not much.

<u>Conclusion</u>.

As the districtwise, sexwise and agewise percentages of pupils in the sample tested and the populations are approximately the same, the sample tested may be treated as reasonably representative of the population.

Scoring the Answer Sheets.

The answer sheets were scored manually with the help of both the keys. The scores on each test were marked in the places provided on the front page of the answer sheet.

After scoring, the answer sheets were sorted according to age, standard, sex and area (rural or urban). Thus all the answer sheets were classified into 60 (5 x 3 x 2 x 2) groups. They were indexed and serial numbers were given to all the answer sheets starting with 1 in case of each group.

Weights to Scores.

The test is hetrogeneous. Each test has different 'g' saturation. They vary in difficulty levels. They also vary in validity in terms of their correlations with the objective criterion. Thus many feel that allotting equal scores to each item in each test is not justified. They prefer to assign weights to the raw scores.

Various methods of assigning weights to the obtained raw scores have been proposed. Number of formulae for deriving the weights have also been proposed.

Weighting the scores involve complicated computations. So research is done by Guilford, Lovell and Williams, by Phillips and by Harper and Dunlap etc. In many studies the weighted scores correlated with the raw scores to the extent of .95 to .99.

As Guilford points out, "Differential weighting of items.....

usually pays very little dividends when there are more than 10 to 20 items..... Thus weights of 1 for all items in long tests of ability are quite appropriate".[1/]

Goodenough also remarks, "In practice the advantages of such weighting have rarely been found to be great enough to justify the labour involved".[2/]

This test, in its final form, contains 154 items. Hence the weighting of scores has not been done. The raw score itself obtained by an individual is treated as the measure of his performance on this test.

Statistical Analysis of the Test Data

The scores on each of the eight sub-tests and scores on the whole test were tabulated for each of the 60 groups in which the answer sheets were classified.

The maximum and minimum scores that can be obtained on the whole test are 154 and 0 respectively. Actually maximum and minimum scores obtained are 144 and 16 respectively. Thus the range, i.e. the interval between the largest and the smallest scores, is 144-16 = 128. The data is meaningful only if it is grouped systematically into frequency distribution. Garrett is of opinion that the size of each group in classifying data should

1/Guilford J.P., Psychometric Methods, Mc Graw-Hill Book Company, Inc., New York, Kogakusha Company Ltd., Tokyo, 1954. pp. 447.

2/Goodenough E.L., Op.cit., pp. 137.

be such that it should not yield more than 20 and less than 12 groups. Thus the size of group is taken as 10 which has yielded 14 groups for drawing up frequency distribution. The frequency distribution of total scores on all the eight tests of all age groups and of the whole group are as given in the table below.

Table 24. Frequency Distribution of Scores of Individuals Belonging to Different Age Groups and of the whole Group.

Class Interval Scores	Frequency of Age Groups.					Whole group
	13 yrs.	14 yrs.	15 yrs.	16 yrs.	17 yrs.	
140-149	1	2	0	2	2	7
130-139	3	4	3	5	5	20
120-129	5	7	10	11	6	39
110-119	12	20	14	21	13	80
100-109	20	25	84	37	20	186
90-99	35	82	109	117	132	475
80-89	182	305	388	304	229	1408
70-79	339	488	588	417	290	2122
60-69	378	482	513	332	187	1892
50-59	236	282	245	183	110	1056
40-49	80	105	78	35	26	324
30-39	19	20	54	13	7	113
20-29	9	2	1	7	1	20
10-19	1	1	1	0	0	3
Total (N)	1320	1825	2088	1484	1028	7745

To attach meaning to these scores, measures of central tendancy are found out because, "....... it is a single measure which represents all of the scores made by the group and as such gives a concise description of the performance of the group as a whole; and..... it enables us to compare two or more groups in terms of typical performance".[1/]

The three types of measures of central tendancy are (1) mean (2) median and (3) mode. Mean is the average score obtained by an individual in the group. Median is the score obtained by the middle individual when the individuals are arranged in a serial order according to their scores. Mode is the score obtained by the greatest number of individuals.

There are different formulae evolved for calculating these measures of central tendency from the frequency distribution.

The formula for calculating mean from the grouped data is

$$M = \frac{fX}{N}$$

where

M = mean score

f = frequency in the class interval

X = midpoint of the class interval

and N = total cases in the sample.

Another way of finding mean is by assumed mean or short method. The formula used is

$$M = AM + \frac{fx}{N} \times i$$

1/Garrett H.E., *Op.cit.*, pp. 28.

where

M = mean

AM = assumed mean

f = frequency in the interval

x = deviation (with sign) of the interval from the one whose midpoint has been taken as the assumed mean

N = size of the sample

i = size of the interval.

All the means in this study, have been calculated by the assumed mean method.

The formula used for calculating median from the grouped data is

$$Mdn = l + \left(\frac{\frac{N}{2} - F}{fm}\right) i$$

where

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 l

fm = frequency (number of scores) within the interval upon which the median lies

i = length of the class interval.

Mode has been calculated by using the formula

$$\text{Mode} = 3 \text{ median} - 2 \text{ mean}.$$

The nature of the group with reference to the performances of the group is decided with reference to the measures of variability. Four measures of variability that have been deviced are (a) the range, (b) the quartile deviation, (c) the mean deviation and (d) the standard deviation.

Standard deviation (SD) is the measure of variability that is usually employed in research. The formula used for computing SD by short method from grouped data is as given below

$$SD = i \times \sqrt{\frac{fx^2}{N} - c^2}$$

where

i = the size of the class interval

fx^2 =sum of the squared deviations in units of class intervals taken from the assumed mean

c^2 = squared correction in units of class interval ($c = \frac{fx}{N}$)

N = size of the sample.

The true mean, median, SD etc. are true hypothetical values obtained from the scores of all the members of the defined group known as population. But as it is almost impossible to find such values, one is content to get these values from the scores of a sample drawn from the population. Population values are called parameters and the values obtained from the samples are called statistics. Statistics are estimates of parameters.

However good the sample may be, the statistics are likely to deviate from the parameters. Parameters cannot be determined. They can however be estimated by computing the extent by which the statistics diverge from the parameters. The accuracy of the estimate is the measure of the reliability of the statistics. The reliability of a statistic is represented in terms of standard error (SE) of the statistics.

The formulae used for computing the SE of the mean, median and SD are as given below

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

$$SE\ median = \frac{1.253\ \sigma}{\sqrt{N}}$$

$$SE_{SD} = \frac{.71\ \sigma}{\sqrt{N}}$$

By using the above quoted formulae, the mean median, mode, standard deviation and also the standard errors of mean, median and standard deviation of each age group and the whole group have been calculated.

Table 25. Statistics and their standard errors of the obtained scores.

Particulars of statistics	Age Group					Whole group
	13 yrs.	14 yrs.	15 yrs.	16 yrs.	17 yrs.	
1	2	3	4	5	6	7
N	1320	1825	2088	1484	1028	7745
Mean	67.824	70.335	72.180	74.132	76.06	71.987
SE_{mean}	.4258	.3567	.3588	.4097	.4690	.1749

(concluded on next page)

Table 25. (concluded)

1	2	3	4	5	6	7
Median	67.833	69.9202	72.084	73.865	75.811	71.690
SE_{mdn}	.5344	.4467	.4503	.5141	.5887	.2196
Mode	67.851	69.0906	71.892	73.351	75.313	71.096
Standard Deviation	15.47	15.24	16.02	15.78	15.04	15.42
SE_{SD}	.3024	.2533	.2548	.2909	.3331	.2484

From the above table, it may be concluded that the deviations of the mean, median, and standard deviation, from the parameters are very narrow.

Nature of Distribution of Test Scores.

If the distribution of the score is normal then only we can say that the tool is satisfactory. If it is not normal then,

either the sample is biased,

or the size of the sample is not adequate,

or the items have not been properly selected,

or any other error might have been committed at a certain stage of test construction.

Following are the procedures adopted to study the nature of the distribution of the scores:-

a) Calculation of skewness and kurtosis

b) Graphical representation of test scores and their interpretation

c) Chi-square test.

The first two procedures have been adopted to study the distribution of the scores.

Calculation of Skewness.

If the distribution is normal, it is represented by the bell shaped curve known as the normal probability curve in which the mean, median and mode fall exactly on the crest of the curve and have the same numerical value. The curve is bilaterally symetrical. There is a perfect symetry between the two halves -right and left- in the figure. "A distribution is said to be "Skewed" when the mean, the median and the mode fall at different points in the distribution, and the balance (or centre of gravity) is shifted to one side or the other, to right or left".[1/] If skewness is observed in the obtained scores, it is necessary to know whether it is a real divergence from the normal or whether it is due to chance fluctuations.

The skewness of the distribution is computed by using the following two formulae:-

$$1)\quad Sk = \frac{3(\text{Mean} - \text{Median})}{\sigma}$$

where Sk = skewness

σ = standard deviation.

1/Ibid., pp. 97-98.

2) $$Sk = \frac{(P_{90} - P_{10})}{2} - P_{50}$$

where

Sk = Skewness

P_{90} = 9 0th percentile

P_{10} = 10th percentile

P_{50} = 50th percentile in the frequency distribution.

One cannot say whether the skewness of a curve is significant unless the standard error of skewness is known. It is calculated by the use of the formula:-

$$SE_{Sk} = \frac{.5185 \times D}{\sqrt{N}}$$

where

$D = (P_{90} - P_{10})$

N = Size of the sample.

Critical ratio is computed by using the formula:-

$$CR = \frac{Sk}{SE_{Sk}}$$

If the CR is less than 1.96 the obtained skewness is insignificant at both .05 and .01 levels of significance. If it is more than 1.96 and less than 2.58 it is significant at .05 level but not at .01 level and if it is more than 2.58 it is significant.

The skewness, standard error of skewness and CR of distribution of scores of each age group and of the whole group are computed by applying the above mentioned formulae.

Table 26. Skewness, Standard Error and Critical Ratio of the distribution of scores.

Particulars of Statistics	Age group					Whole group
	13 yrs.	14 yrs.	15 yrs.	16 yrs.	17 yrs.	
N	1320	1825	2088	1484	1028	7745
P_{90}	86.423	88.106	90.528	93.312	95.197	90.126
P_{10}	50.4745	51.433	52.416	54.602	55.755	52.491
P_{50}	67.833	69.9202	72.084	73.865	75.811	71.690
Sk	0.193	-0.133	-0.612	0.092	-0.335	-0.382
SE_{Sk}	0.5137	0.4446	0.4323	0.5210	0.6377	0.2218
CR	0.5304	0.2985	1.416	0.1766	0.5253	1.723

From the table given above it may be observed that the CR in all the cases is not significant at .05 level of significance as it is less than 1.96. Hence it can be concluded that the observed skewness is not a real one but is only due to chance fluctuations.

Calculations of Kurtosis.

Kurtosis refers to the peakedness or flatness of a curve, as compared to that of a normal one. If the distribution is more peaked, it is said to be leptokurtic and if it is more flat, it is said to be platykurkic. The normal curve is called mesokurtic.

The kurtosis is computed by using the formula:-

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

where

Ku = Kurtosis

$$Q = \frac{P_{75} - P_{25}}{2}$$

P_{90} = 90th percentile

P_{10} = 10th percentile

P_{75} = 75th percentile

P_{25} = 25th percentile in the frequency distribution.

If the distribution is normal the kurtosis obtained by the use of this formula is .263. If the obtained kurtosis is more than .263 the distribution is platykurtic and if it is less than .263 it is leptokurtic.

The question howmuch deviation of kurtosis from .263 is to be treated as insignificant or is due to chance fluctuation can only be answered by computing the standard error of kurtosis and the critical ratio. The formulae used for computing these are as given below:-

$$SE_{Ku} = \frac{.28}{\sqrt{N}}$$

where

N = size of the sample

$$CR = \frac{.263 - Ku}{SE_{Ku}}$$

where

Ku = obtained kurtosis

SE_{Ku} = standard error of kurtosis

By using the above quoted formulae the kurtosis, standard error of kurtosis and CR of the distribution were computed.

Table 27. Kurtosis, Standard Error of Kurtosis and the Critical ratio of the distribution of scores.

Particulars of statistics	Age Group					Whole group
	13 yrs.	14 yrs.	15 yrs.	16 yrs.	17 yrs.	
N	1320	1825	2088	1484	1028	7745
P_{75}	77.375	79.270	81.717	84.137	86.051	80.480
P_{25}	58.865	60.4596	62.287	63.658	65.543	61.726
P_{90}	86.423	88.106	90.528	93.312	95.197	90.126
P_{10}	50.4745	51.433	52.416	54.602	55.755	52.491
Ku	0.2574	0.2701	0.2549	0.2647	0.2599	0.2621
SE_{Ku}	0.007707	0.006555	0.006128	0.007269	0.008734	0.003182
D	-0.0056	0.0071	-0.0081	0.0017	-0.0031	-0.0009
CR	0.7266	1.803	1.321	0.2338	0.3550	0.2828

From the above table it may be observed that in all cases the obtained CR is less than 1.96 and hence the deviations obtained in all the cases are insignificant at both the levels of significance. Thus kurtosis obtained in case of all the distributions is due to chance fluctuations.

The observed deviations of skewness and kurtosis of the distributions are insignificant and hence distribution of scores of each age group and of the whole group are normal.

Graphical Representation of the Scores.

The obtained distribution can be compared merely by seeing if it is represented graphically, though the measures of variability from normal are not computed. "In fact, the direction and extent of asymetry often strike us more convincingly when seen in a graph than when expressed by measures of skewness and kurtosis".1/

Four methods of representing a frequency distribution that are commonly used are (a) frequency polygon (b) histogram (c) the cummulative frequency graph (d) cummulative percentage curve or ogive.

Following graphs have been drawn to see the symmetry of the graphs in case of each age group and whole group:-

a) Frequency polygon

b) Smoothed frequency polygon

c) Histograms

Frequency polygons.

The midpoints of the class intervals are plotted along 'X' axis and the frequency along the 'Y' axis. Garrett suggests that

1/Ibid., pp. 101

"The ratio of height to width may vary from 60-80 % and figure will have good proportions"[1/].

The frequency polygon gives information regarding the distribution of scores of the group and also about the nature of the test. It tells whether the scores are piled up at low or high end or whether they are distributed symetrically. If they are piled up at the high end the test is too easy and if they are piled up at low end, it is too hard. If the test is suitable for the group, the scores will tend to be distributed symetrically around the mean. When this happens it approximates the normal frequency curve.

By following the method suggested by Garrett, the frequency polygons have been plotted in case of each age group and for the whole group. The data for each group have been given in the following tables:-

Table 28. Original and Smoothed Frequencies, of Age Group 13 years.

Class interval Scores	Mid point	Original Frequency	Smoothed Frequency
1	2	3	4
140-149	144.5	1	1.33
130-139	134.5	3	3.00
120-129	124.5	5	6.67
110-119	114.5	12	12.33

(concluded on next page)

1/*Ibid*, pp.12.

Table 28. (concluded)

1	2	3	4
100-109	104.5	20	22.33
90-99	94.5	35	79.00
80-89	84.5	182	185.33
70-79	74.5	339	299.67
60-69	64.5	378	317.67
50-59	54.5	236	231.33
40-49	44.5	80	111.67
30-39	34.5	19	36.00
20-29	24.5	9	9.67
10-19	14.5	1	3.33

Table 29. Original and Smoothed Frequencies
Age Group 14 years.

Class interval scores	Mid point	Original frequency	Smoothed frequency
140-149	144.5	2	2.00
130-139	134.5	4	4.33
120-129	124.5	7	10.33
110-119	114.5	20	17.33
100-109	104.5	25	42.33
90-99	94.5	82	137.33

(concluded on next page)

Table 29. (concluded)

1	2	3	4
80-89	84.5	305	291.67
70-79	74.5	488	425.00
60-69	64.5	482	417.33
50-59	54.5	282	289.67
40-49	44.5	105	135.67
30-39	34.5	20	42.33
20-29	24.5	2	7.67
10-19	14.5	1	1.00

Table 30. Original and Smoothed Frequencies
Age Group 15 years

Class interval scores	Mid point	Obtained frequency	Smoothed frequency
140-149	144.5	0	1.00
130-139	135.5	3	4.33
120-129	124.5	10	9.00
110-119	114.5	14	36.00
100-109	104.5	84	69.00
90-99	94.5	109	193.67
80-89	84.5	388	361.67
70-79	74.5	588	496.33

(concluded on next page)

Table 30. (concluded)

1	2	3	4
60-69	64.5	513	448.67
50-59	54.5	245	278.67
40-49	44.5	78	125.67
30-39	34.5	54	44.33
20-29	24.5	1	18.67
10-19	14.5	1	0.67

Table 31. Original and Smoothed Frequencies
Age Group 16 years.

Class interval Scores	Mid point	Obtained frequency	Smoothed frequency
140-149	144.5	2	2.33
130-139	134.5	5	6.00
120-129	124.5	11	12.33
110-119	114.5	21	23.00
100-109	104.5	37	58.33
90-99	94.5	117	152.67
80-89	84.5	304	279.33
70-79	74.5	417	351.00
60-69	64.5	332	310.67
50-59	54.5	183	183.33

(concluded on next page)

Table 31. (concluded)

1	2	3	4
40-49	44.5	35	77.00
30-39	34.5	13	18.33
20-29	24.5	7	6.67
10-14	14.5	0	2.33

Table 32. Original and Smoothed Frequencies
Age Group 17 years

Class interval Scores	Mid point	Obtained frequency	Smoothed frequency
140-149	144.5	2	2.33
130-139	134.5	5	4.33
120-129	124.5	6	8.00
110-119	114.5	13	13.00
100-109	104.5	20	55.00
90.99	94.5	132	127.00
80-89	84.5	229	217.00
70-79	74.5	290	235.33
60-69	64.5	187	195.67
50-59	54.5	110	107.67
40-49	44.5	26	47.67
30-39	34.5	7	11.33

(concluded on next page)

Table 32. (concluded)

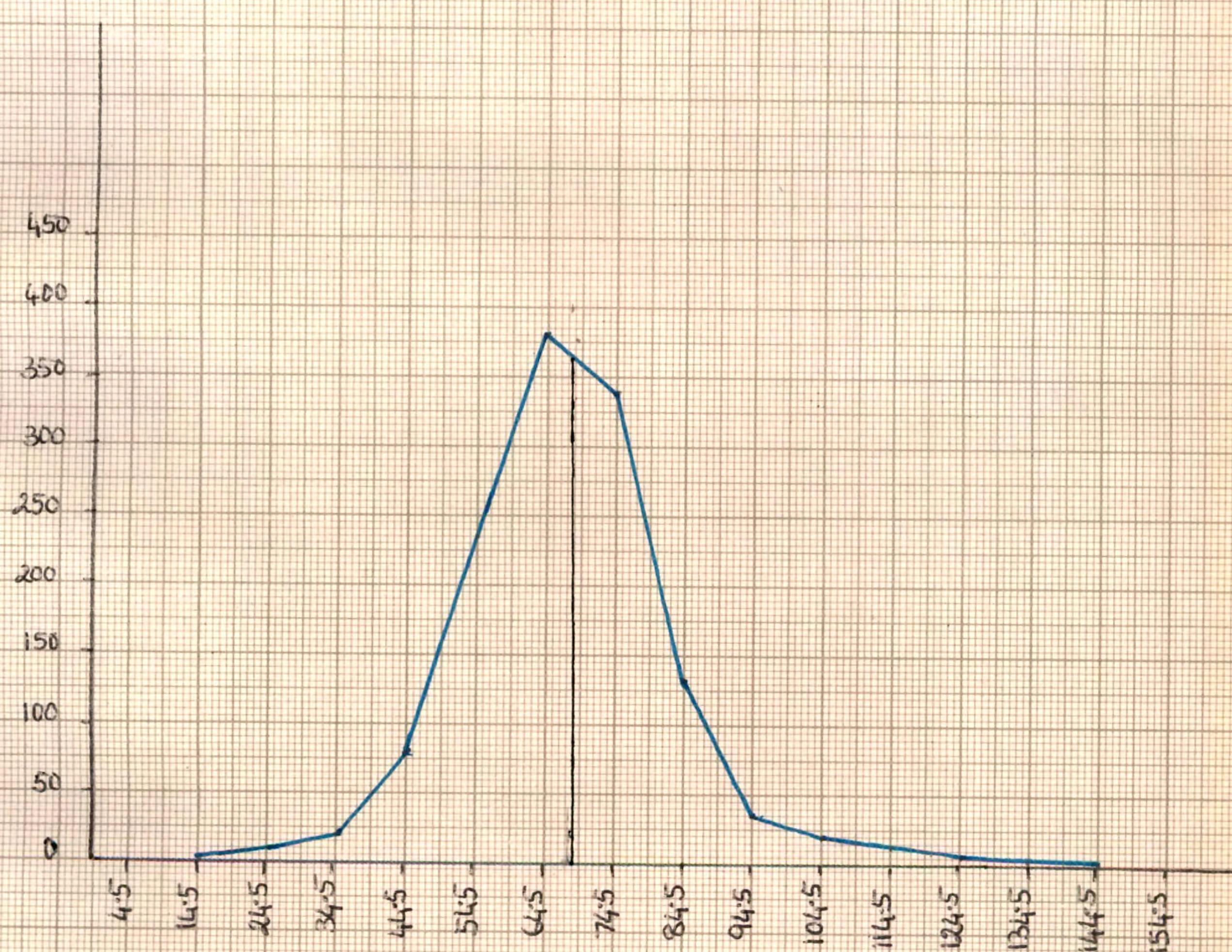
1	2	3	4
20-29	24.5	1	2.67
10-19	14.5	0	0.33

Table 33. Original and Smoothed Frequencies
Whole Group.

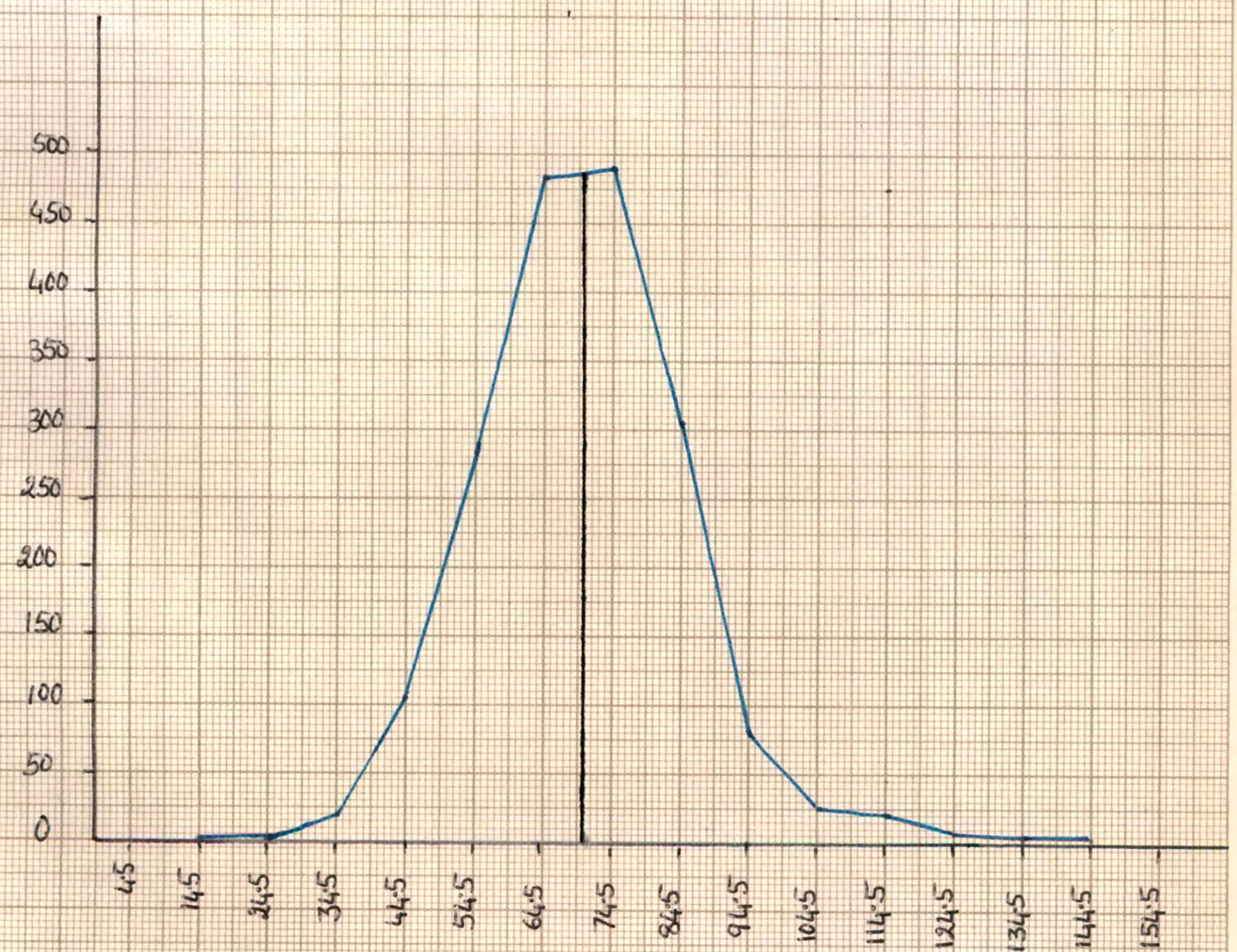
Class interval Scores	Mid point	Obtained freqnency	Smoothed frequency
140-149	144.5	7	9.00
130-139	134.5	20	22.00
120-129	124.5	39	46.33
110-119	114.5	80	101.67
100-109	104.5	186	247.00
90-99	94.5	475	689.67
80-89	84.5	1408	1335.00
70-79	74.5	2122	1807.33
60-69	64.5	1892	1690.00
50-59	54.5	1056	1090.67
40-49	44.5	324	497.67
30-39	34.5	113	152.33
20-29	24.5	20	45.33
10-19	14.5	3	7.67

Graph Numbers 1 to 6 show the nature of polygons from the obtained frequencies in each case.

Graph No. 1
Frequency Polygon
Age Group 13 years.



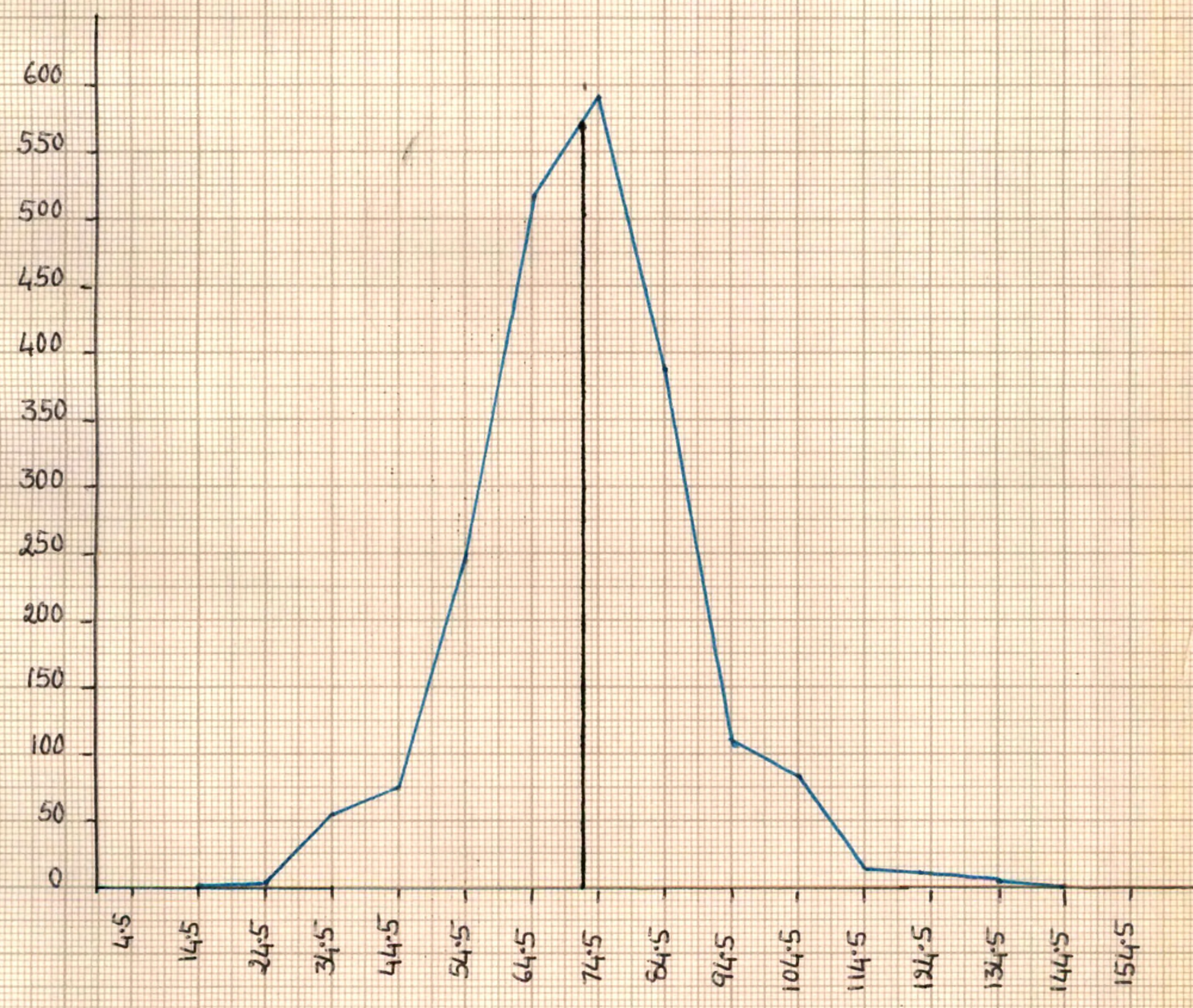
Graph No. 2
Frequency Polygon
Age Groups 14 years.



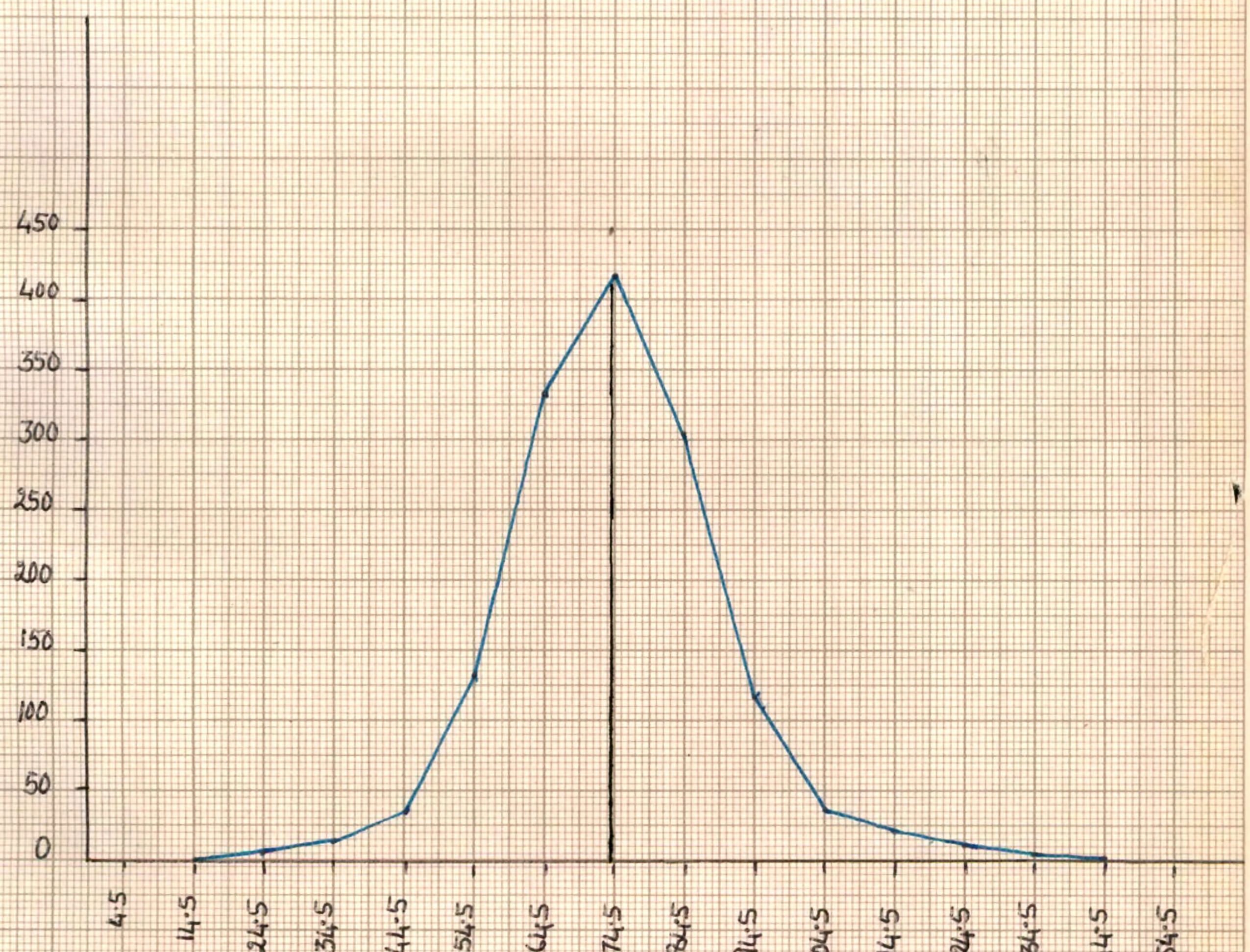
Graph No. 3

Frequency Polygon

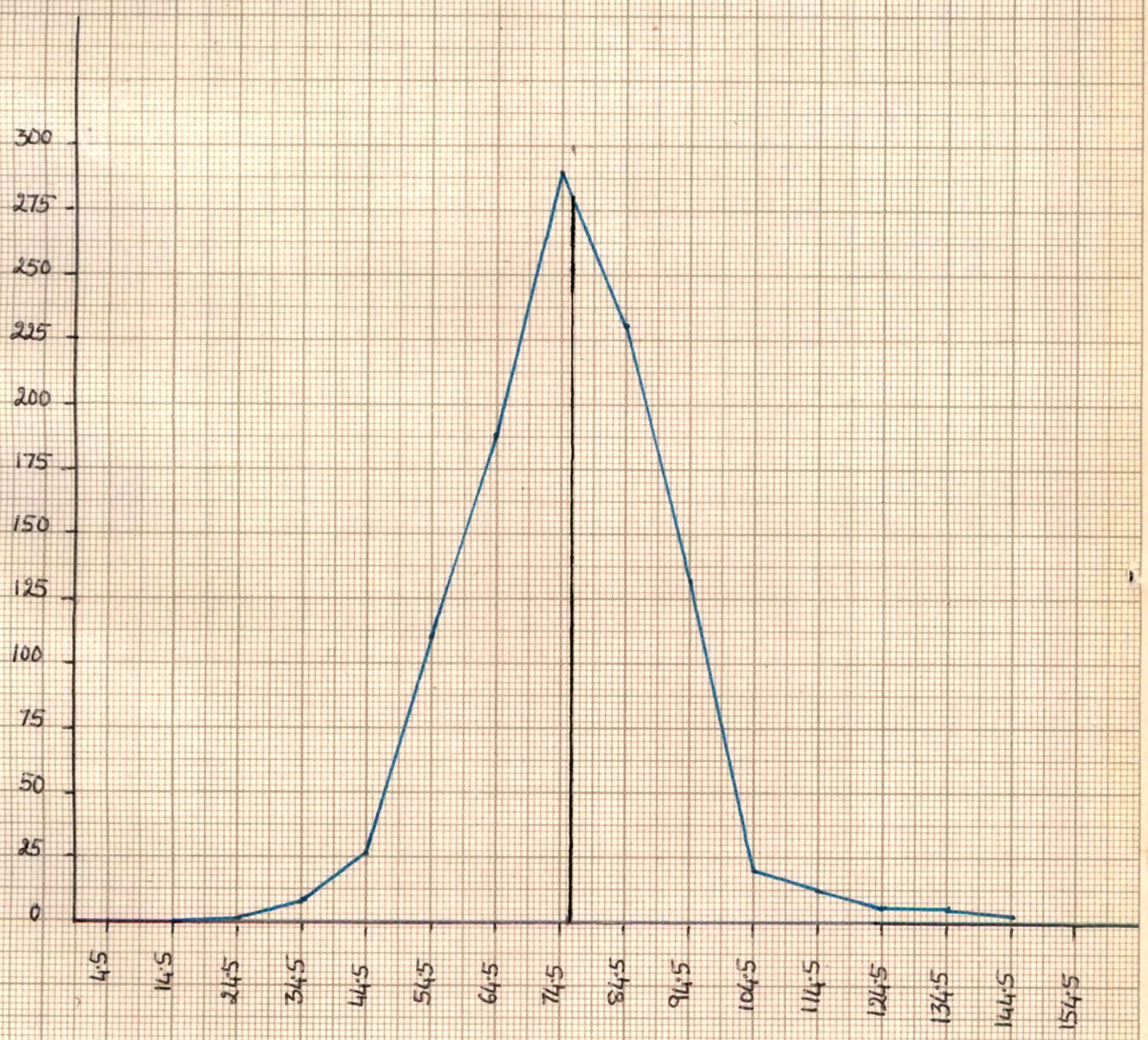
Age Group 15 years.



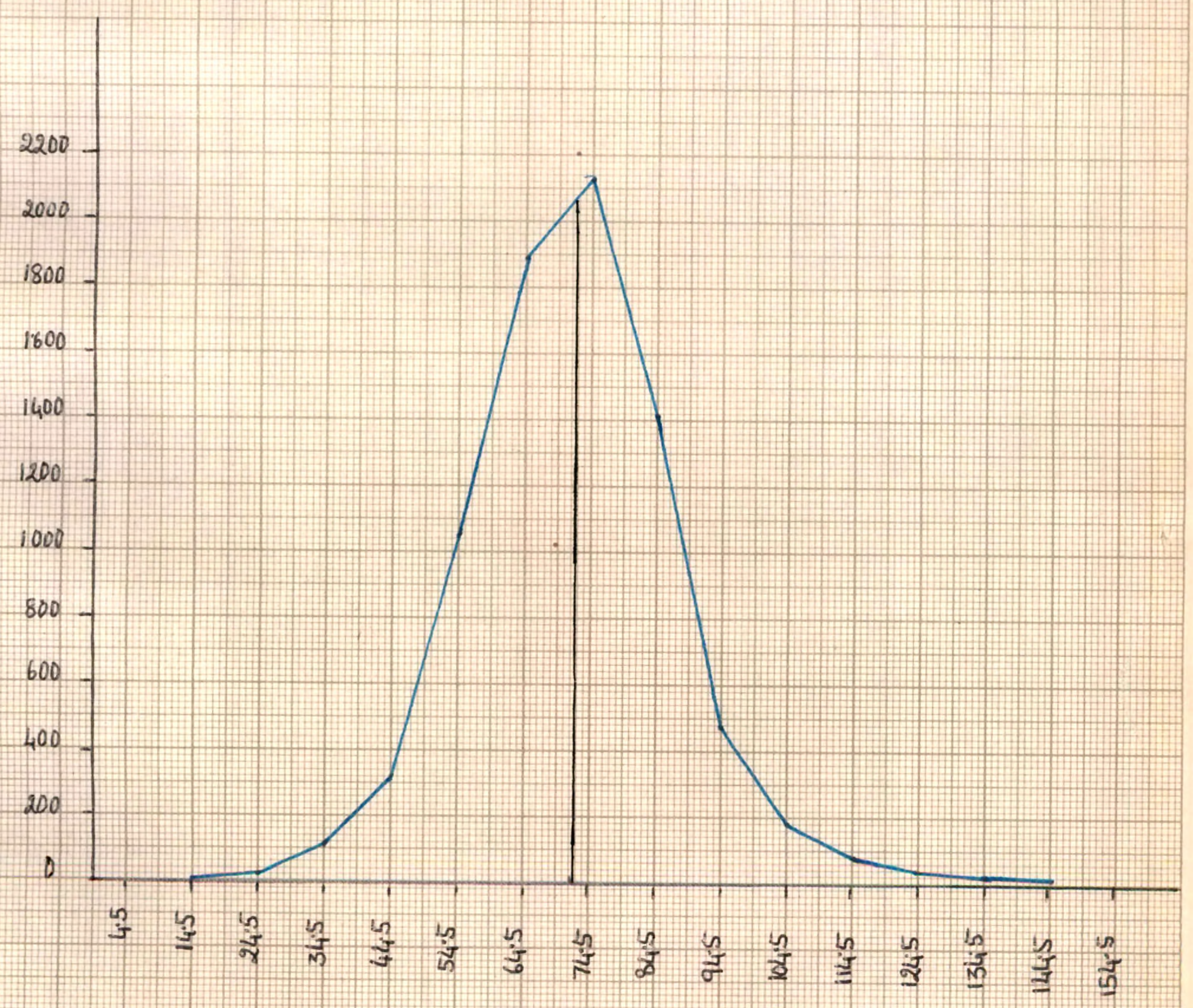
Graph No. 4
Frequency Polygon
Age Group 16 years.



Graph No. 5
Frequency Polygon
Age Group 17 years.



Graph No. 6
Frequency Polygon
Whole Group.



Smoothed Frequency Polygons

Due to chance irregularities the polygons may tend to be jagged in outline. "To iron out chance irregularities, and also to get a better notion of how the figure might look, if the data were more numerous, the frequency polygon may be "smoothed"....".[1/]

Thus to smoothen the chance irregularities, the frequencies in each of the class intervals, are smoothed by following the method suggested by Garrett.[2/] Smoothed frequencies have also been given in table Nos. 28 to 33.

By smoothing the frequencies, the surface of the polygon becomes more continuous. Smoothing gives a picture of what an investigator would have got if his data were more numerous.

With the help of the data in tables 28 to 33 smoothed frequencies were plotted against midpoints. The graph thus obtained was superimposed on the frequency polygon with obtained frequencies.

Graph numbers 7 to 12 show the nature of smoothed polygons obtained from the smoothed frequencies in each case.

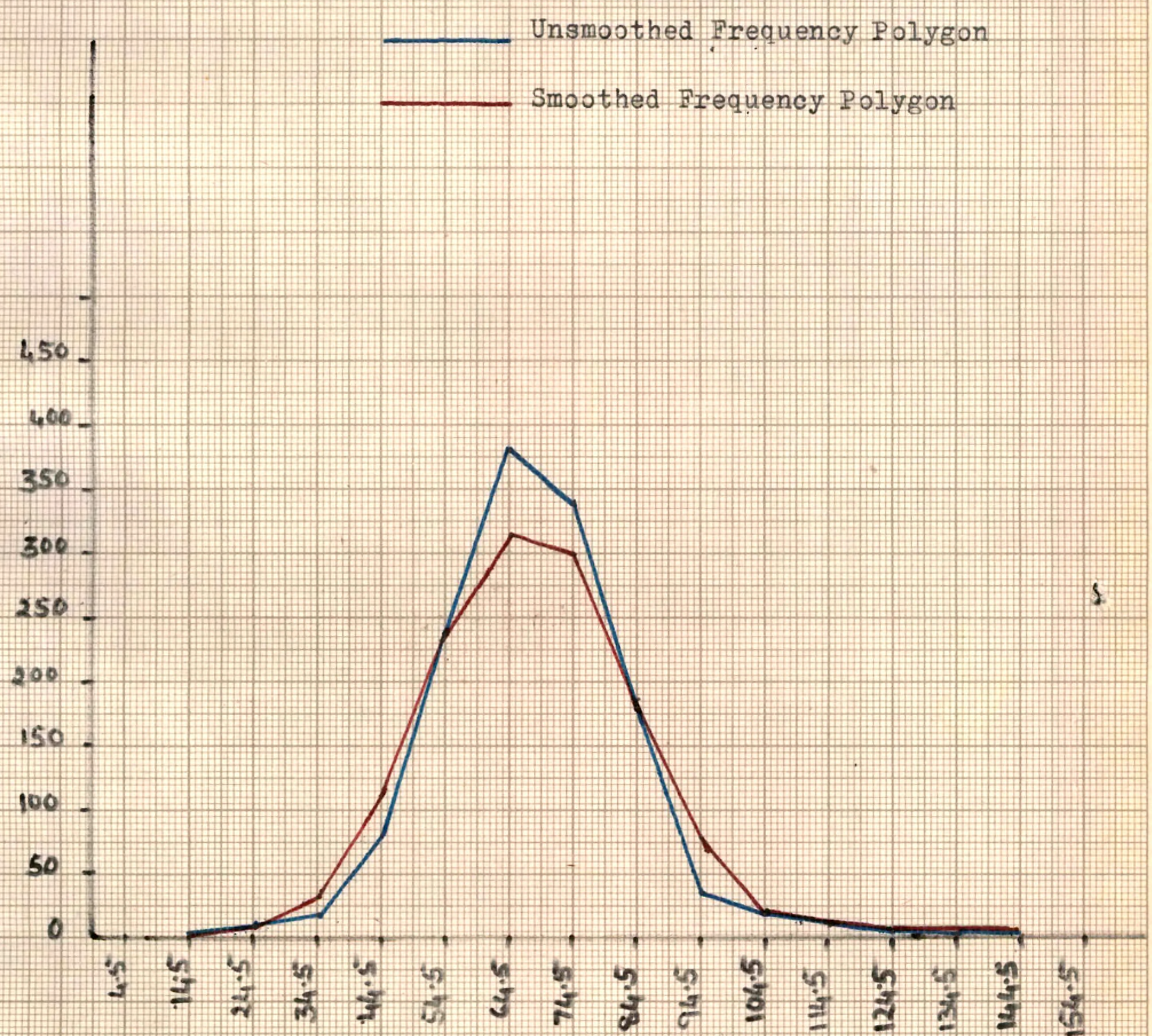
1/*Ibid*. pp. 14.

2/*Ibid*. pp. 14-15.

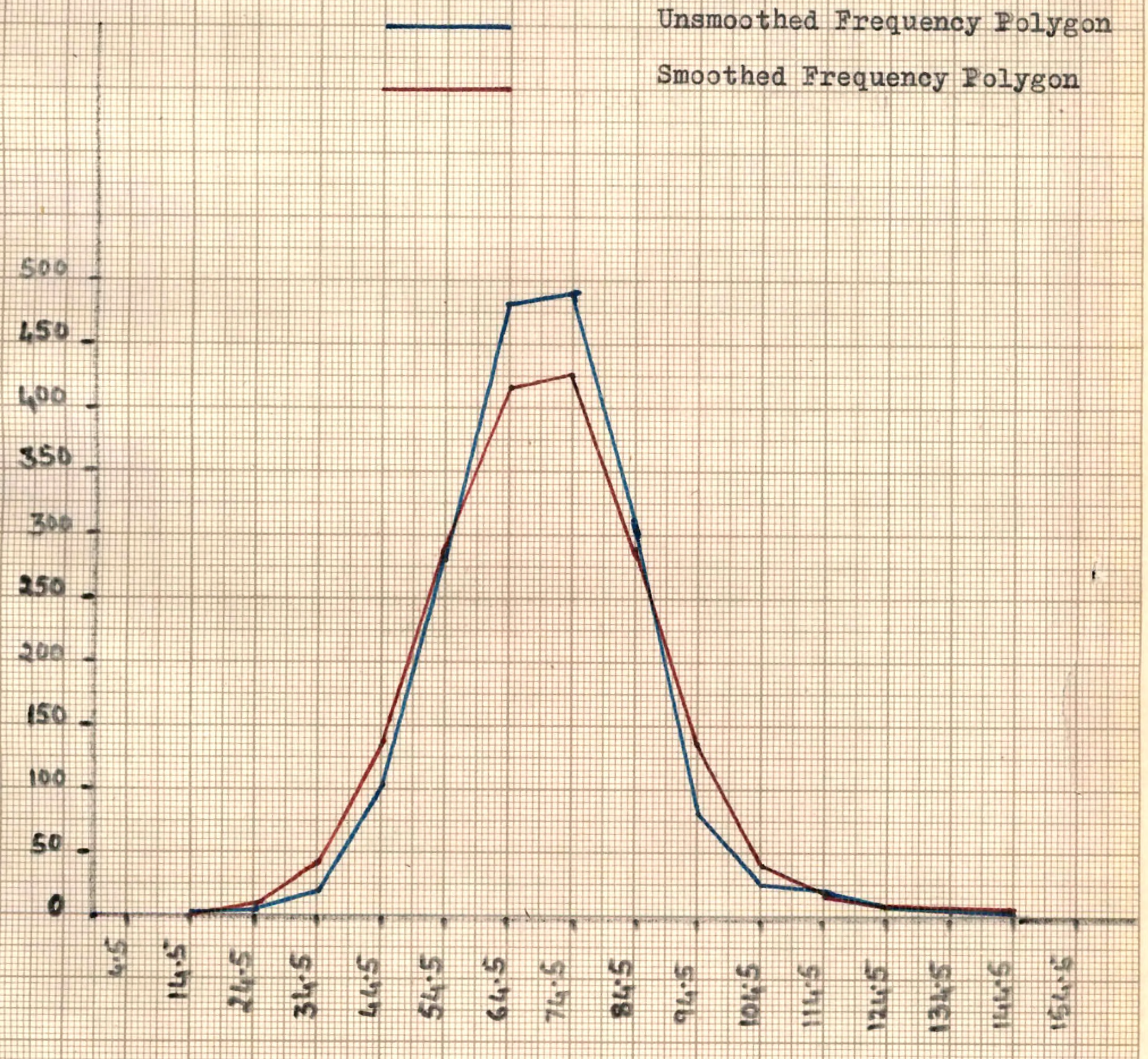
Graph No. 7

Smoothed and Unsmoothed Frequency Polygons

Age Group 13 years.



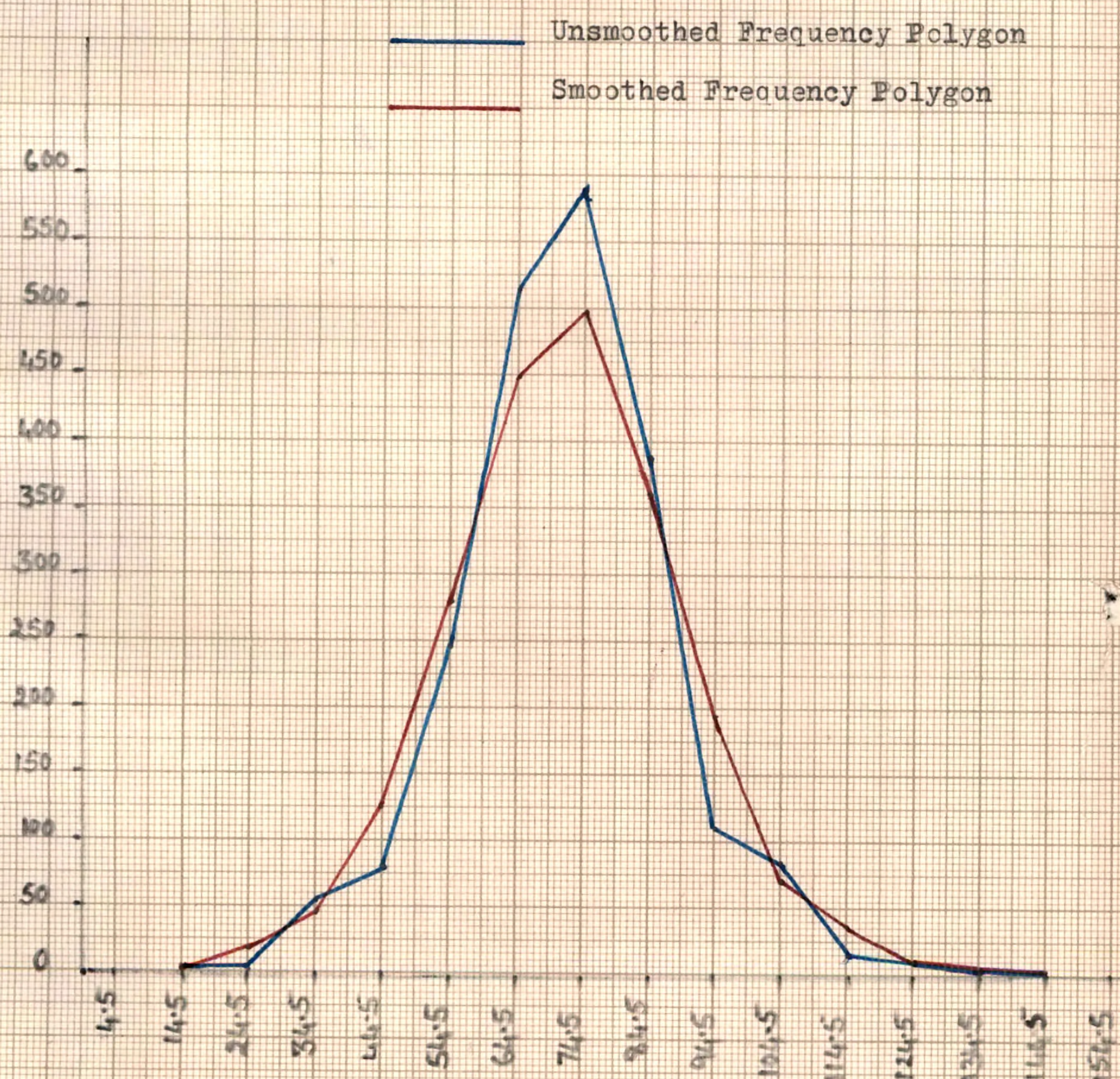
Graph No. 8
Smoothed and Unsmoothed Frequency Polygons
Age Group 14 years.



Graph No. 9

Smoothed and Unsmoothed Frequency Polygons

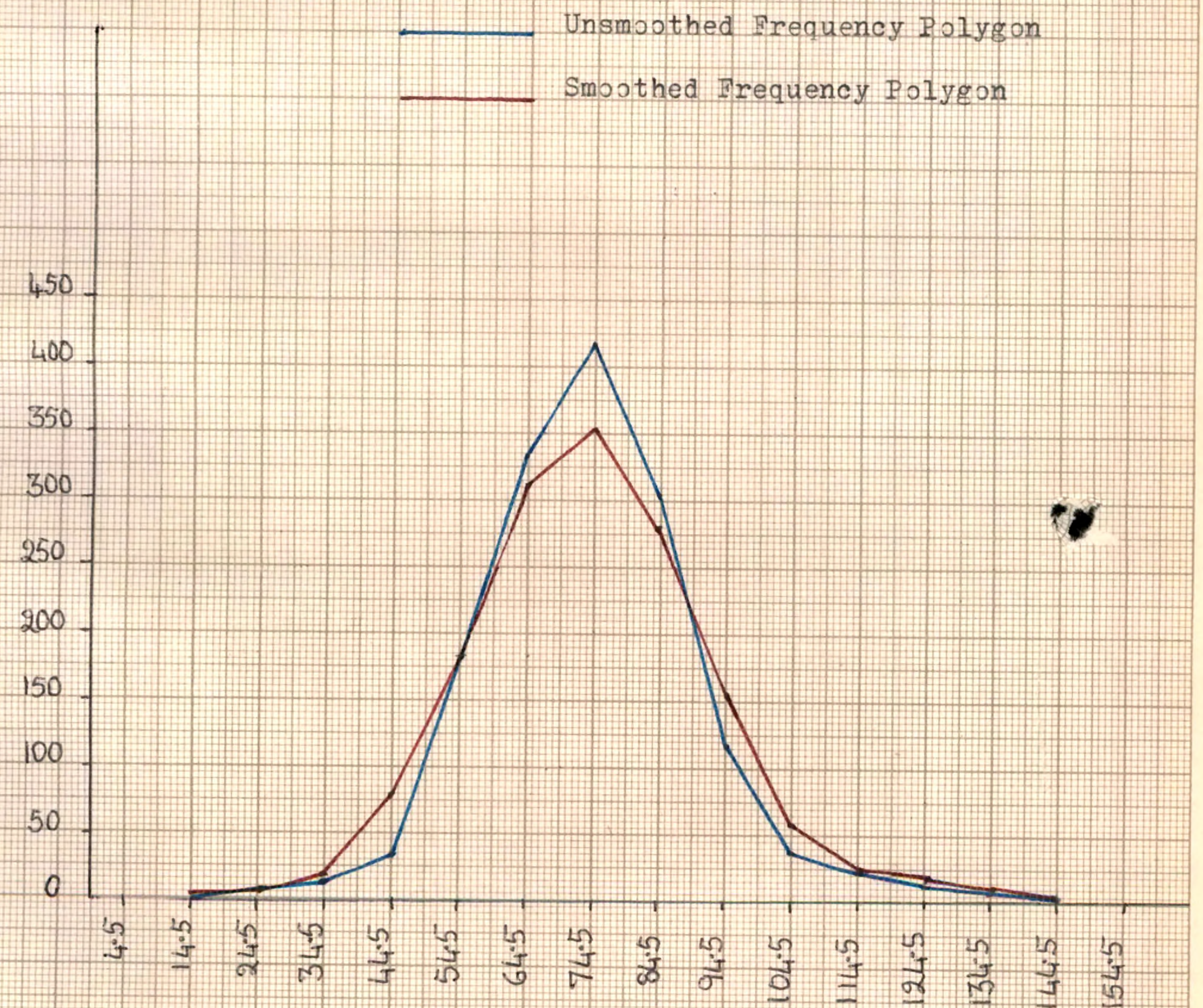
Age Group 15 years.



Graph No. 10

Smoothed and Unsmoothed Frequency Polygons

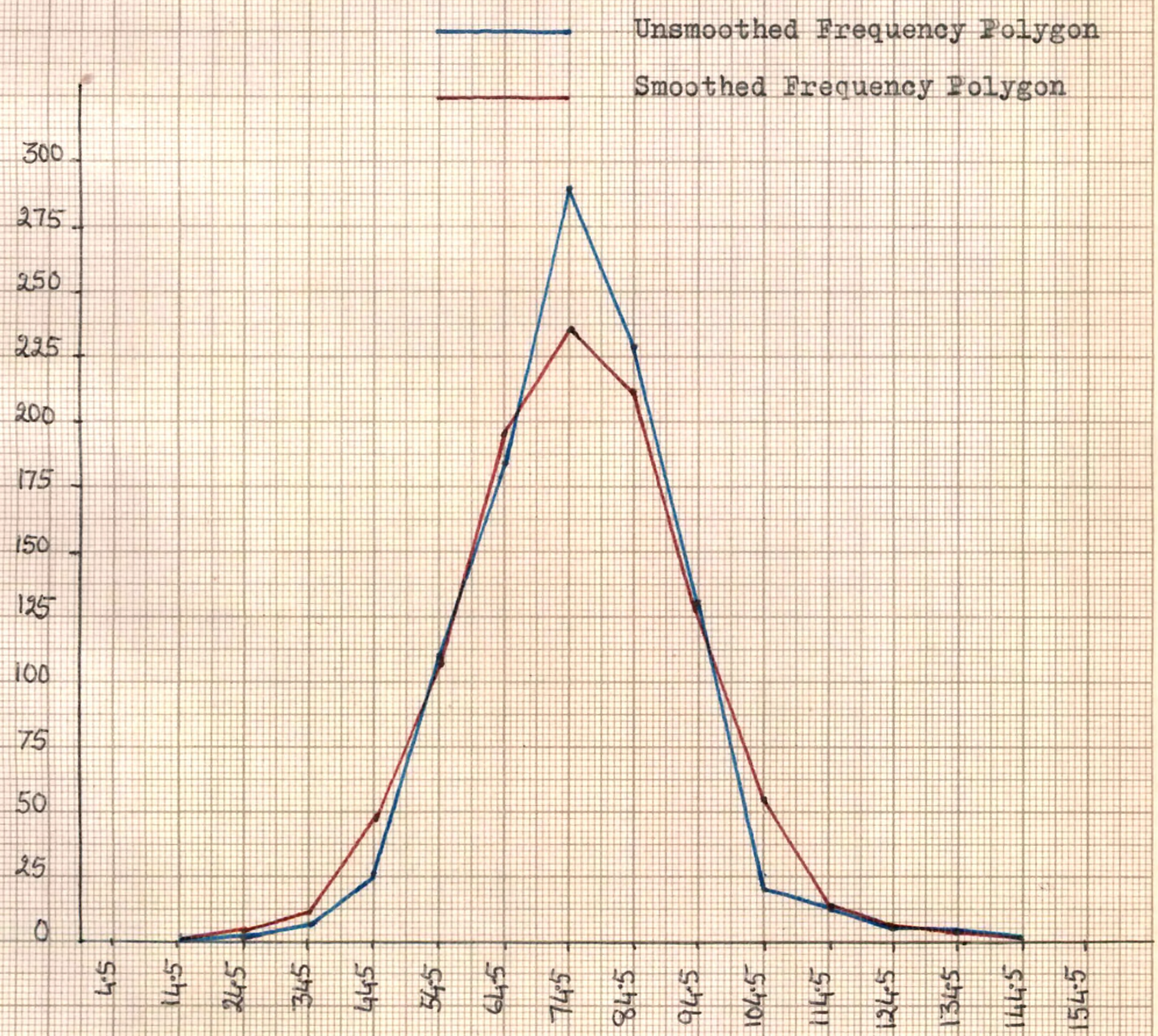
Age Group 16 years.



Graph No. 11

Smoothed and Unsmoothed Frequency Polygons

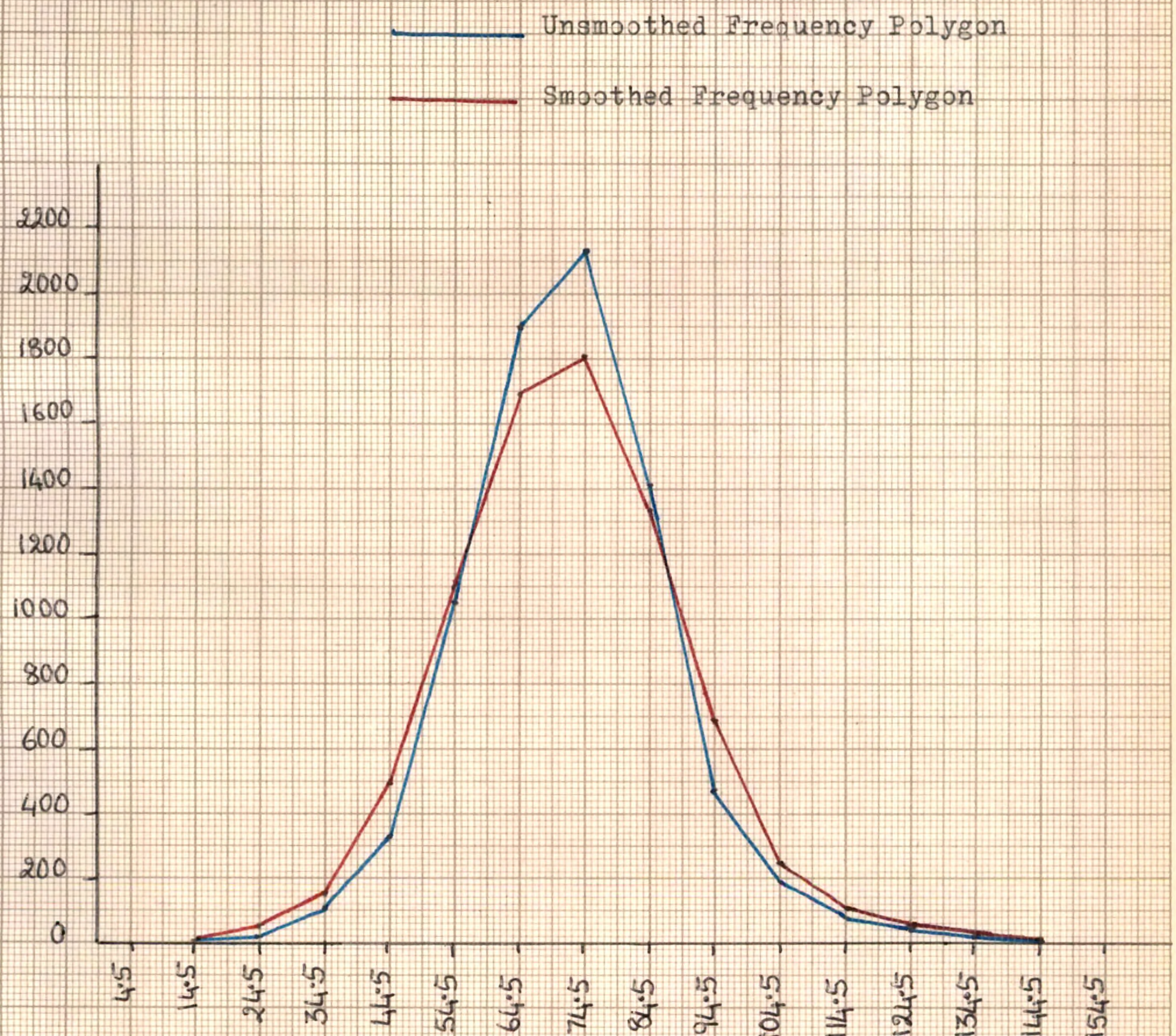
Age Group 17 years.



Graph No. 12

Smoothed and Unsmoothed Frequency Polygons.

Whole Group.



These graphs show that the scores are neither piled at the left end nor at the right but they have been distributed almost symmetrically around the mean.

Histogram.

Another way of representing the scores graphically is by means of drawing histograms or column graphs. In case of histograms the assumption is made that all the scores have uniformaly spread in their interval. Thus they are represented by rectangles. The base of each rectangle is equal to the interval and the height is equal to the frequency in the interval. As all the intervals are equal, the bases of all rectangles are ~~the~~ equal. They differ in their heights according to the frequencies. The rise or fall in the height of the rectangle indicates the increase or decrease respectively in the number of scores. The areas of the rectangles represent the frequencies in the intervals. "In contrast to frequency polygon, the area of each rectangle in a histogram is directly proportional to the number of measures within the interval"[1/]. So the histograms represent more accurate picture of the distribution of the frequency from interval to interval.

The data in tables 28 to 33 have been used to draw the histograms of each age group and of the whole group.

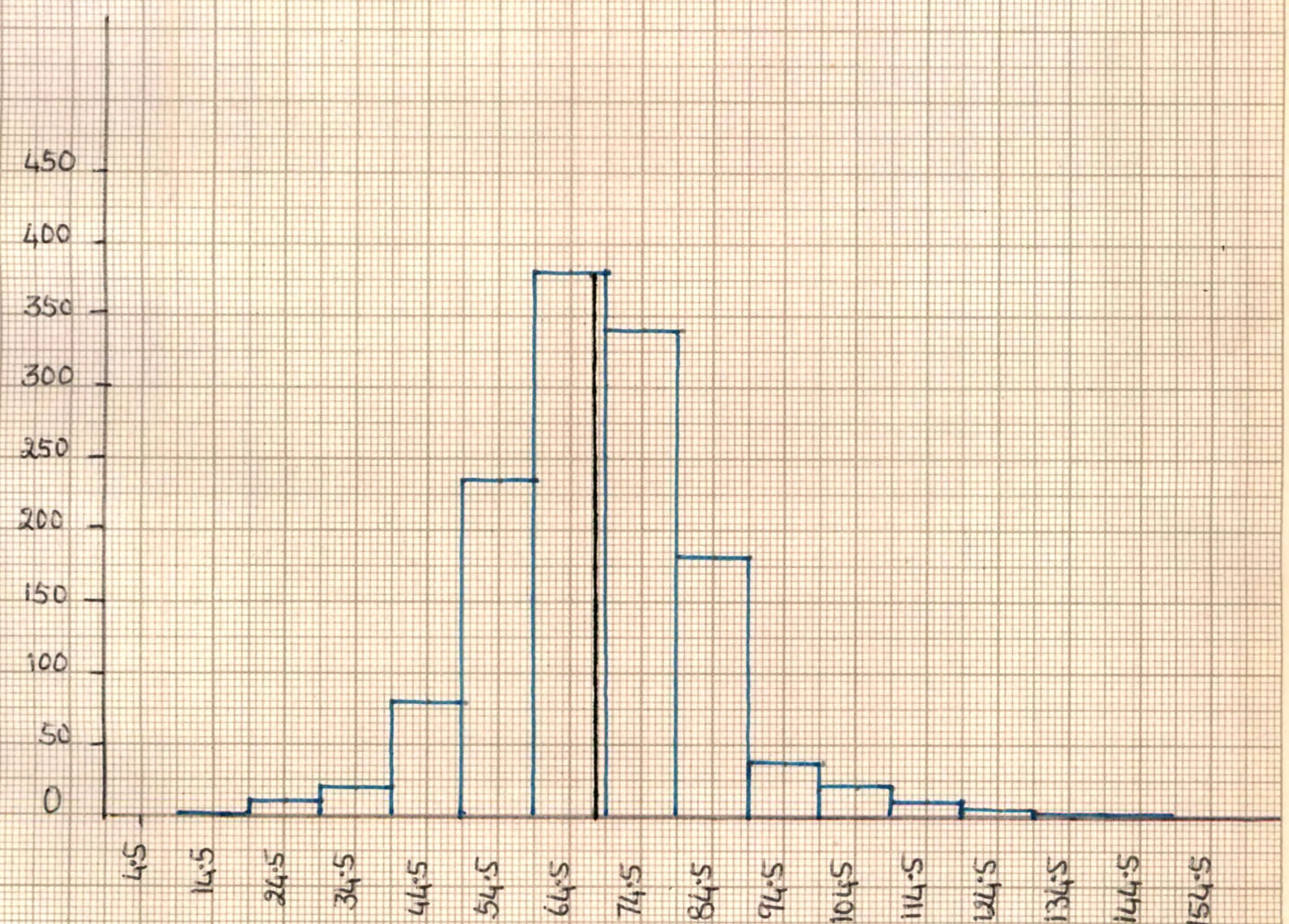
Graph numbers 13 to 18 represent the nature of the histograms of the different groups.

1/Ibid. pp. 17.

Graph No. 13

HISTOGRAM

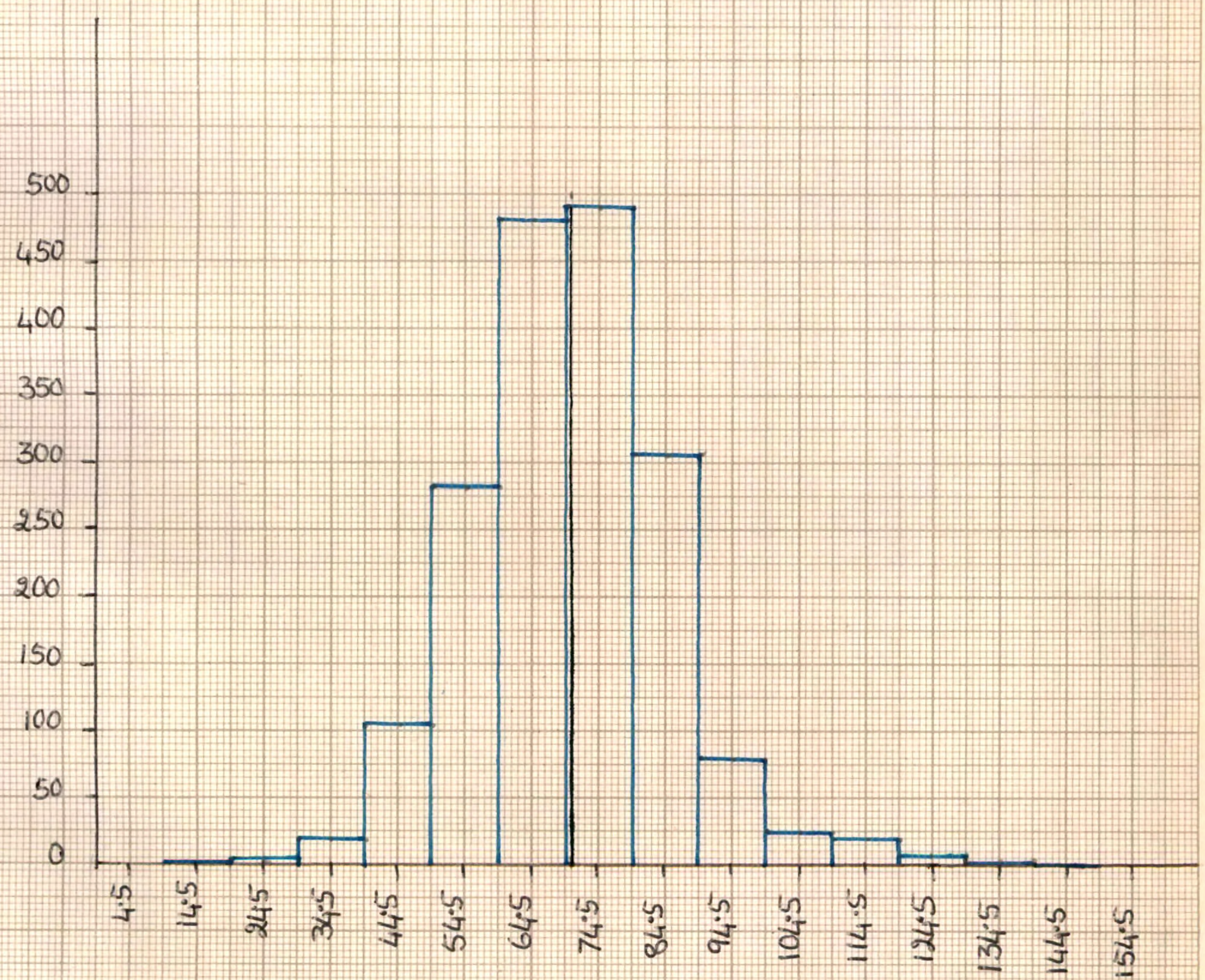
Age Group 13 years.



Graph No. 14

HISTOGRAM

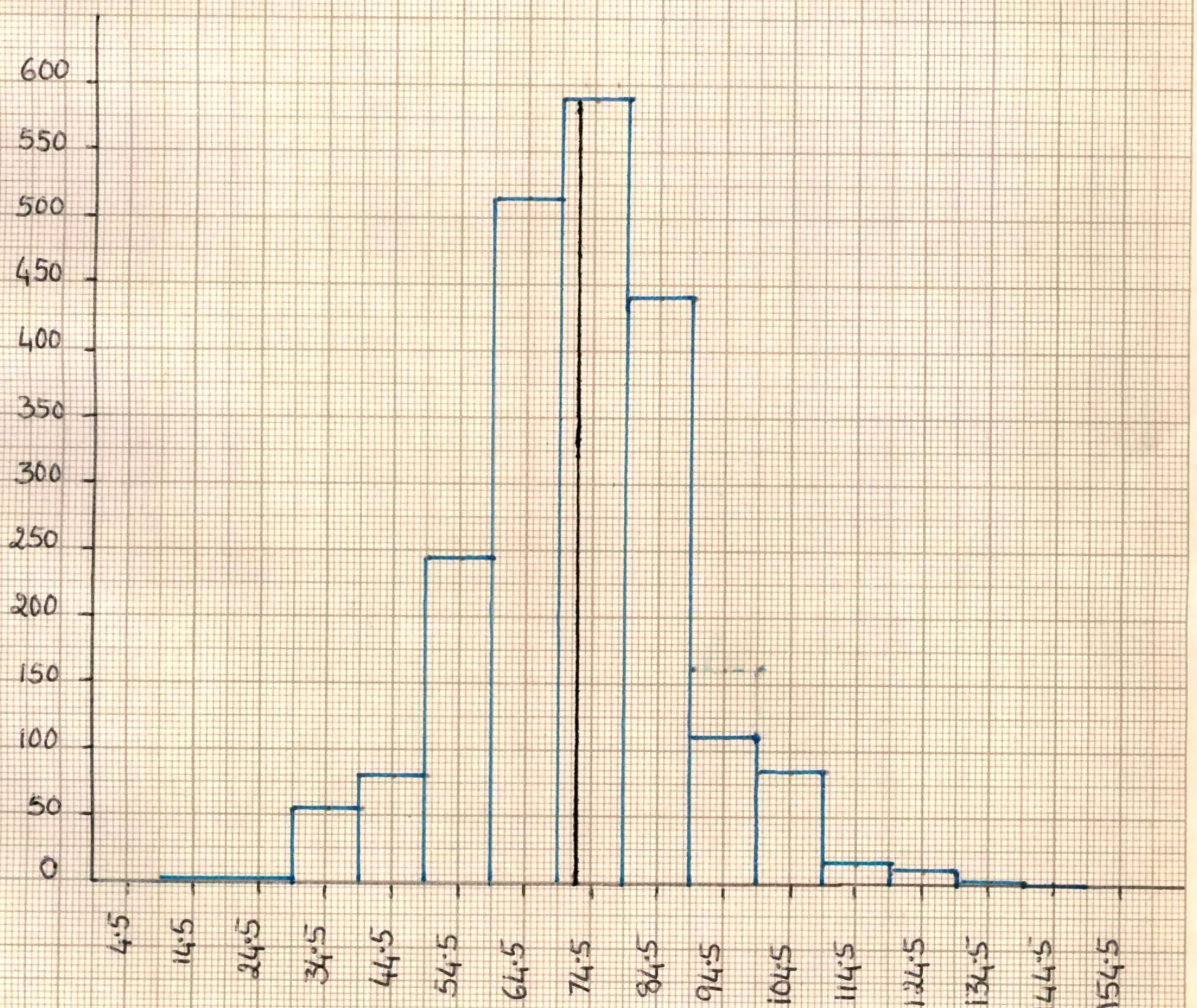
Age Group 14 years.



Graph No. 15

HISTOGRAM

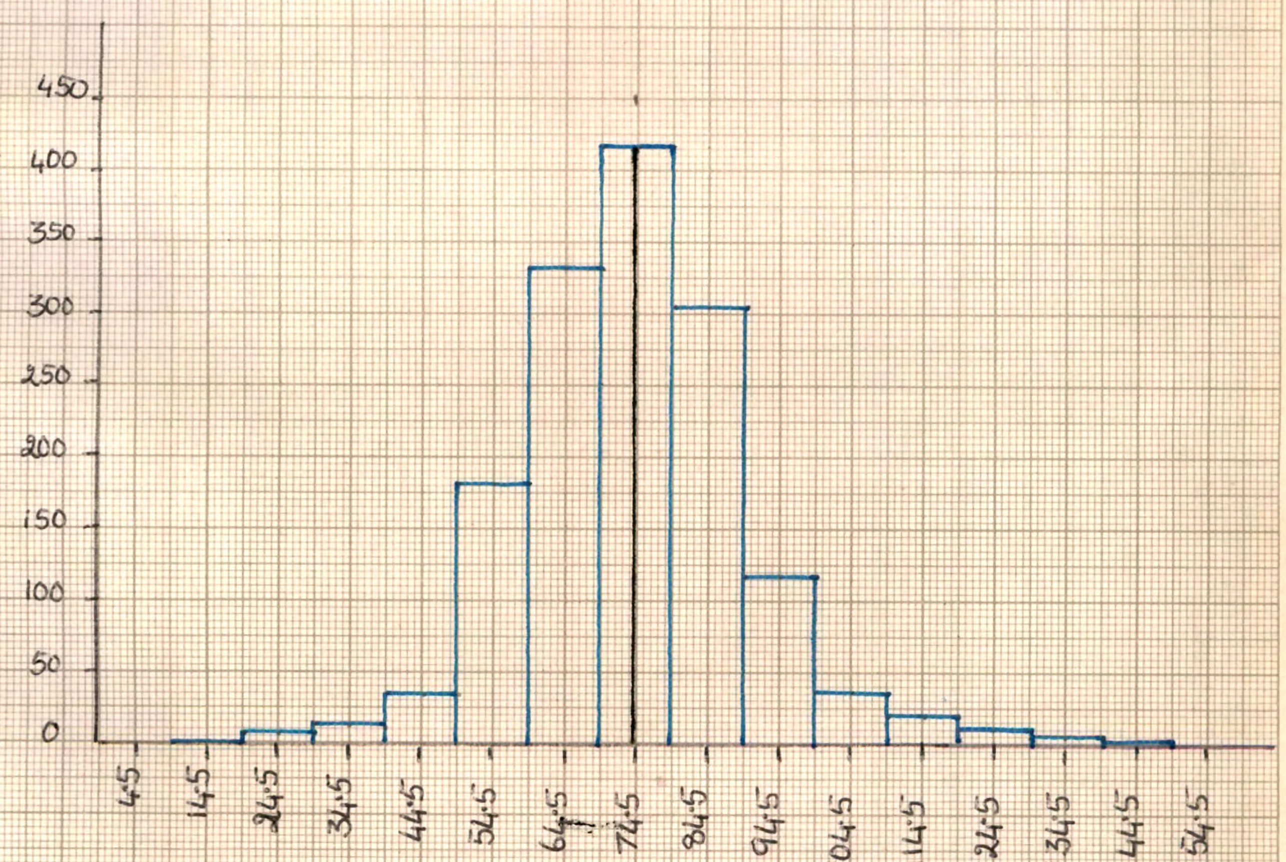
Age Group 15 years.



Graph No. 16

HISTOGRAM

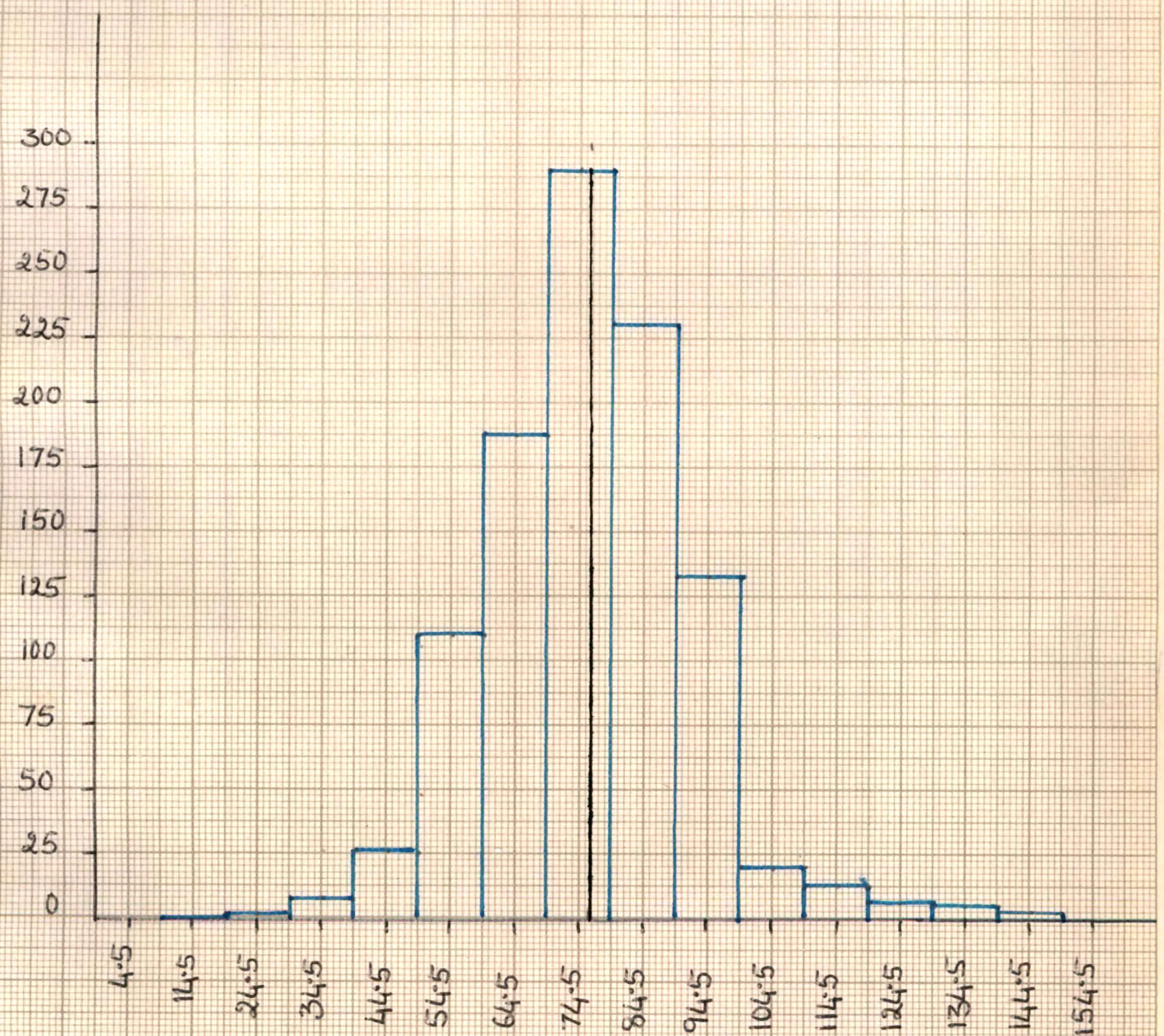
Age Group 16 years.



Graph No. 17

HISTOGRAM

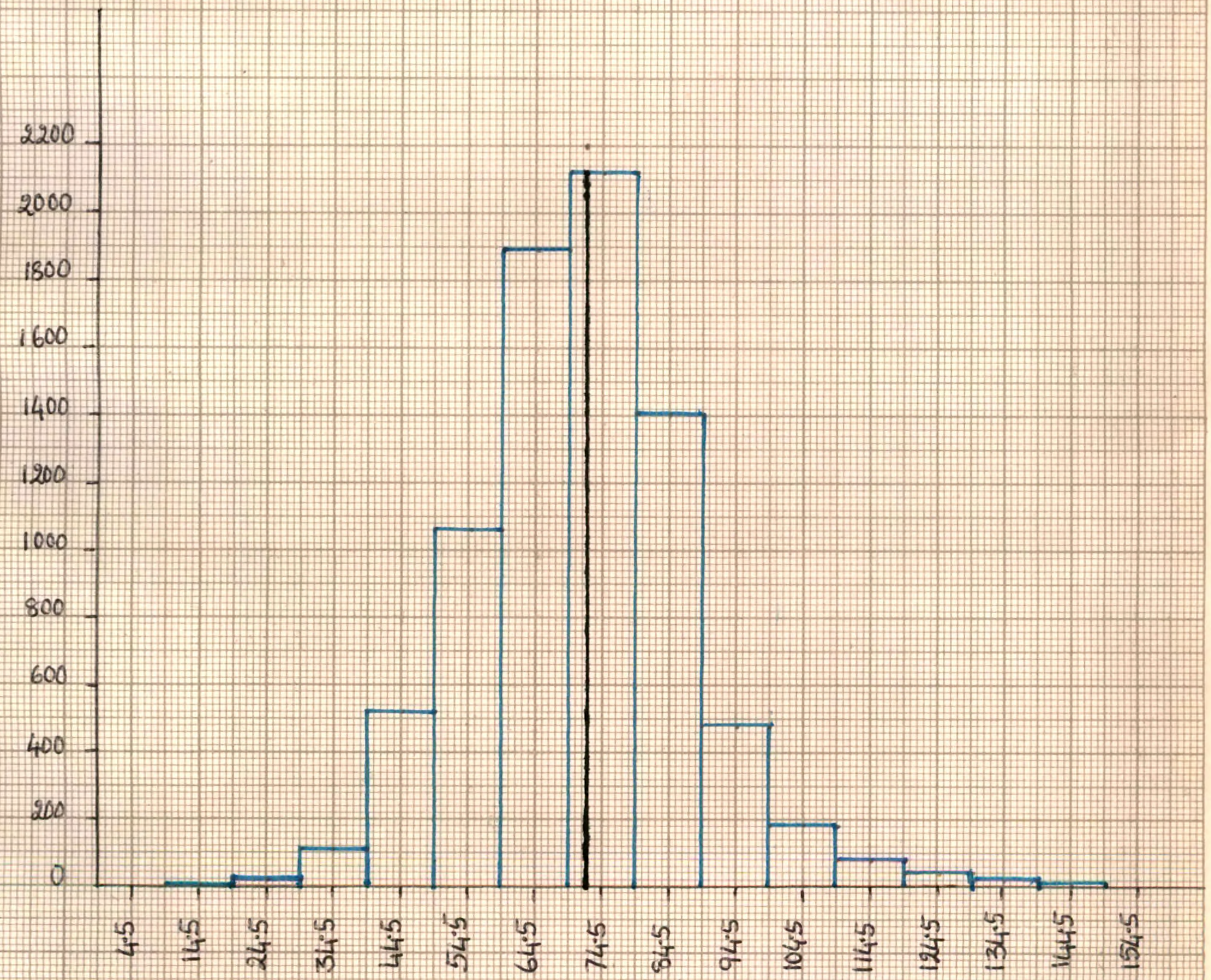
Age Group 17 years.



Graph No. 18

HISTOGRAM

Whole Group.



The mean and median are also shown in the histograms. The casual inspection of the histograms indicate that the scores have almost been symmetrically distributed around their means.

Comparision With a Normal Probability Curve.

The histogram or freqnency polygon can be compared with a normal probability curve of the same area, mean and standard deviation. A normal probability curve can be constructed if N and SD of a given distribution are known.

Below is given the equation of the normal probability curve:-

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

where

x = scores (expressed as deviations from the mean) laid off along the baseline or 'X' axis

y = the height of the curve above the 'X' axis, i.e. the frequency of a given x value or the number achieving a certain score

N = number of cases

σ = standard deviation of the distribution

π = 3.1416

e = 2.7183 (base of the Napierian system of logarithms)

Thus if N and σ are known it is possible to find the height of the ordinate (frequency) for any value of x i.e. the number of persons making a particular score.

To draw a curve to be superimposed on the obtained histogram, the maximum height of the ordinate y_0 is to be calculated.

The value of x at the mean of the normal curve is 0. So, when $x = 0$, $e^{-\frac{x^2}{2\sigma^2}} = 1$

Therefore, $y_0 = \frac{N}{\sigma\sqrt{2\pi}}$

Value of σ in interval units is used in the above formula as the values on 'X' axis are in terms of class intervals. 'y_0' is the frequency at the point where the mean lies. So the values of y_0 for all age groups have been calculated by using the above formula.

The values of y's, the heights of the ordinates of the normal probability curve expressed as fractional parts of the mean ordinate y_0 at distances 1 σ, 2 σ, and 3 σ, were found from the table B [1/] and from these were computed the values of y's corresponding to y_0 computed for each age group and the whole group.

Table 34. Normal Curve Ordinates at Mean, $\pm$ 1 σ, $\pm$ 2 σ and $\pm$ 3 σ corresponding to obtained y_0

Particulars	Age Group					Whole Group
	13 yrs.	14 yrs.	15 yrs.	16 yrs.	17 yrs.	
1	2	3	4	5	6	7
N	1320	1825	2088	1484	1028	7745
SD	15.47	15.24	16.02	15.78	15.04	15.42

(concluded on next page)

1/Ibid, pp. 425.

Table 34. (concluded)

1	2	3	4	5	6	7
σ (in class internal) units	1.547	1.524	1.602	1.578	1.504	1.542
$y_0 = \frac{N}{\sigma\sqrt{2\pi}}$	339.9	477.2	519.30	374.7	272.4	2003
$\pm 1\ \sigma = y_0 \times .60653$	206.2	289.5	315.0	227.3	165.2	1213
$\pm 2\ \sigma = y_0 \times .13534$	45.99	64.57	70.26	50.70	36.84	270.4
$\pm 3\ \sigma = y_0 \times .01111$	3.773	5.302	5.763	4.163	3.023	22.18

From this data, seven points on the normal probability curve are obtained in each case. The curve is drawn with the help of these seven points to superimpose it on the obtained histograms, The obtained curves are shown in graphs number 19 to 24.

The normal curves plotted on the histograms are fitting the obtained distributions well enough to treat the data as adequately normal.

Conclusion.

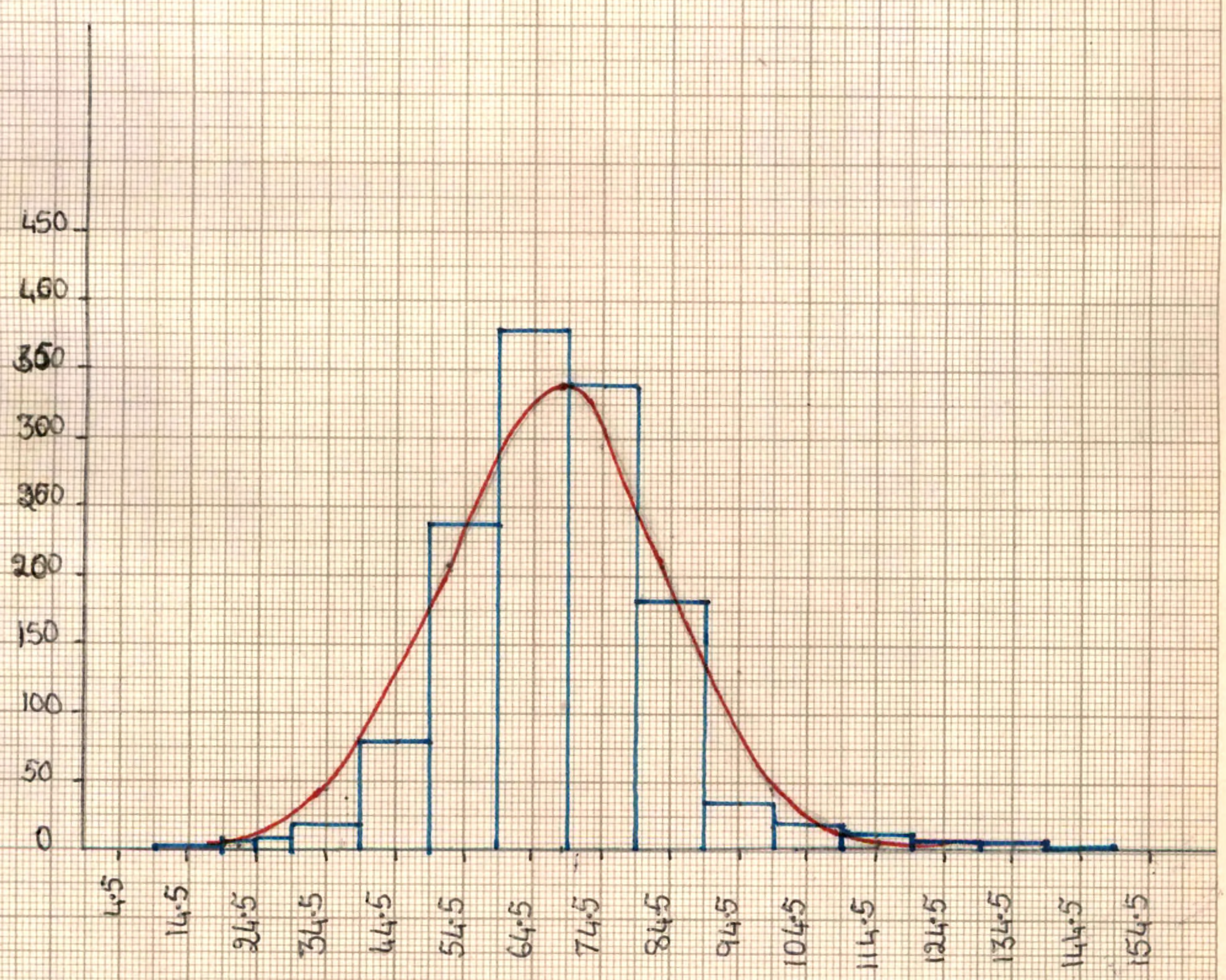
Thus both the tests applied to the data indicate that the distributions of scores of each age group and the whole group can safely be treated as normal.

Norms.

As the sample is representative of the defined population

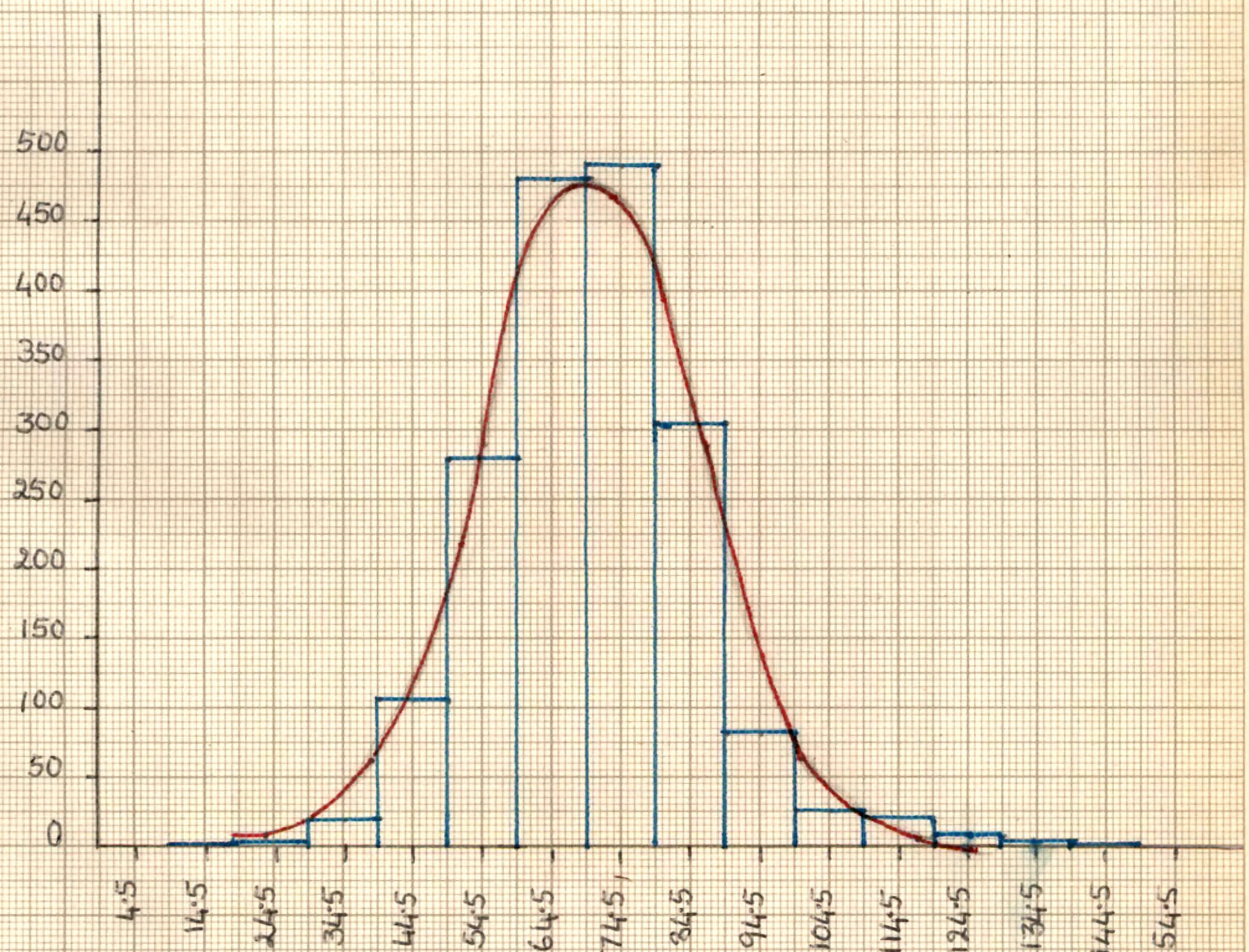
Graph No. 19

Best-Fitting Normal Curve Superimposed On Histogram
Age Group 13 years.



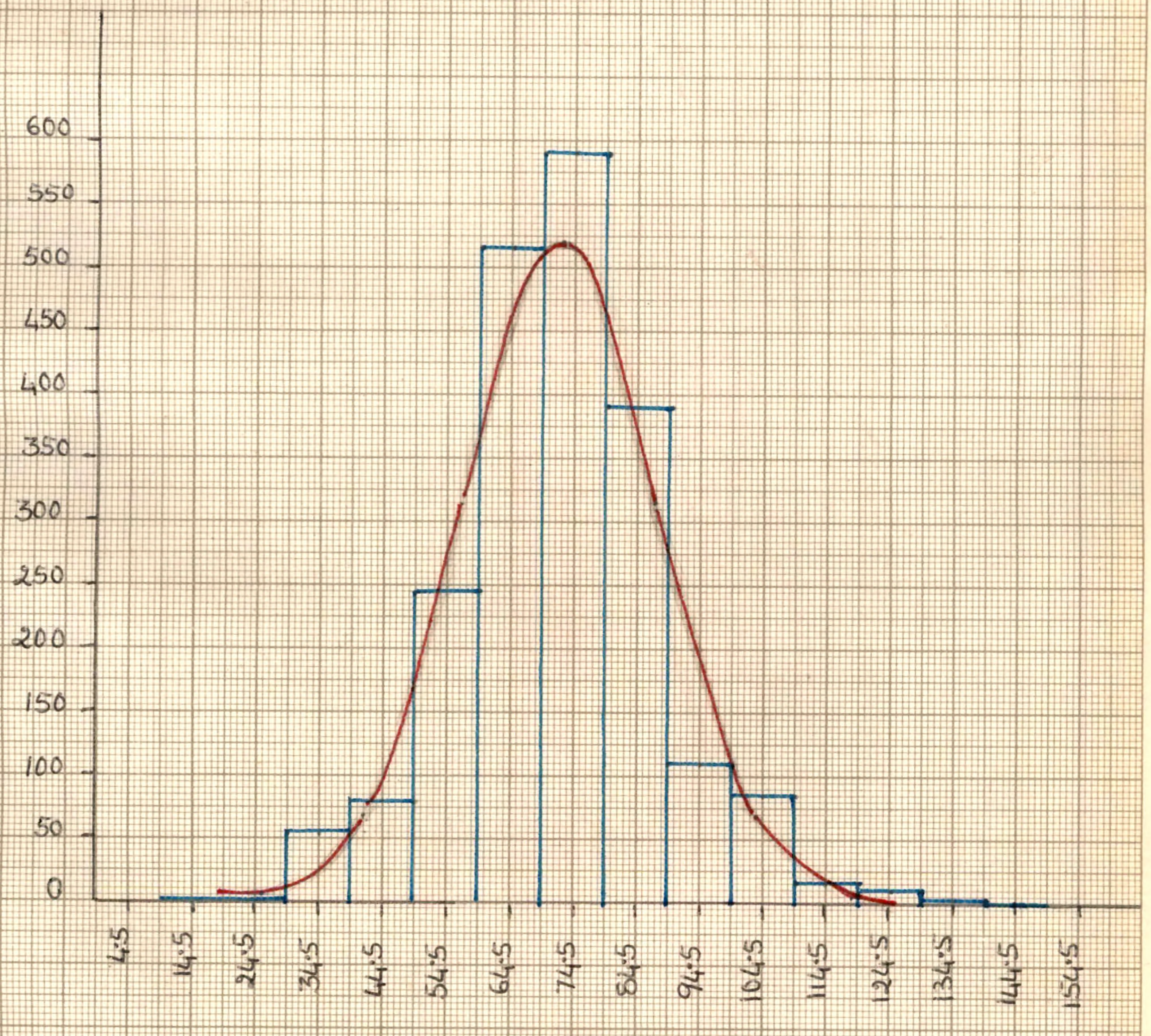
Graph No. 20

Best-Fitting Normal Curve Superimposed On Histogram
Age Group 14 years.



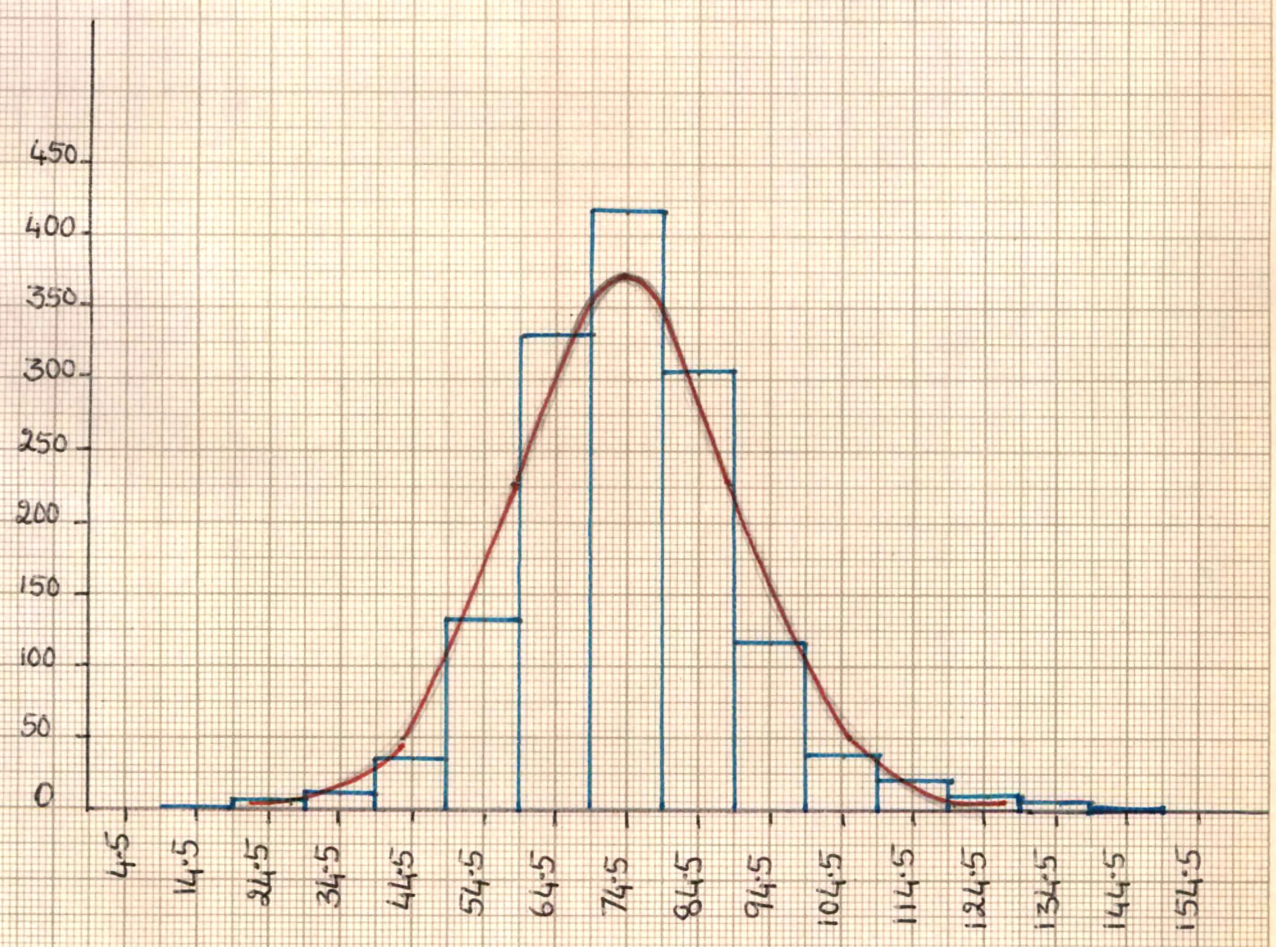
Graph No. 21

Best-Fitting Normal Curve Superimposed On Histogram
Age Group 15 years.



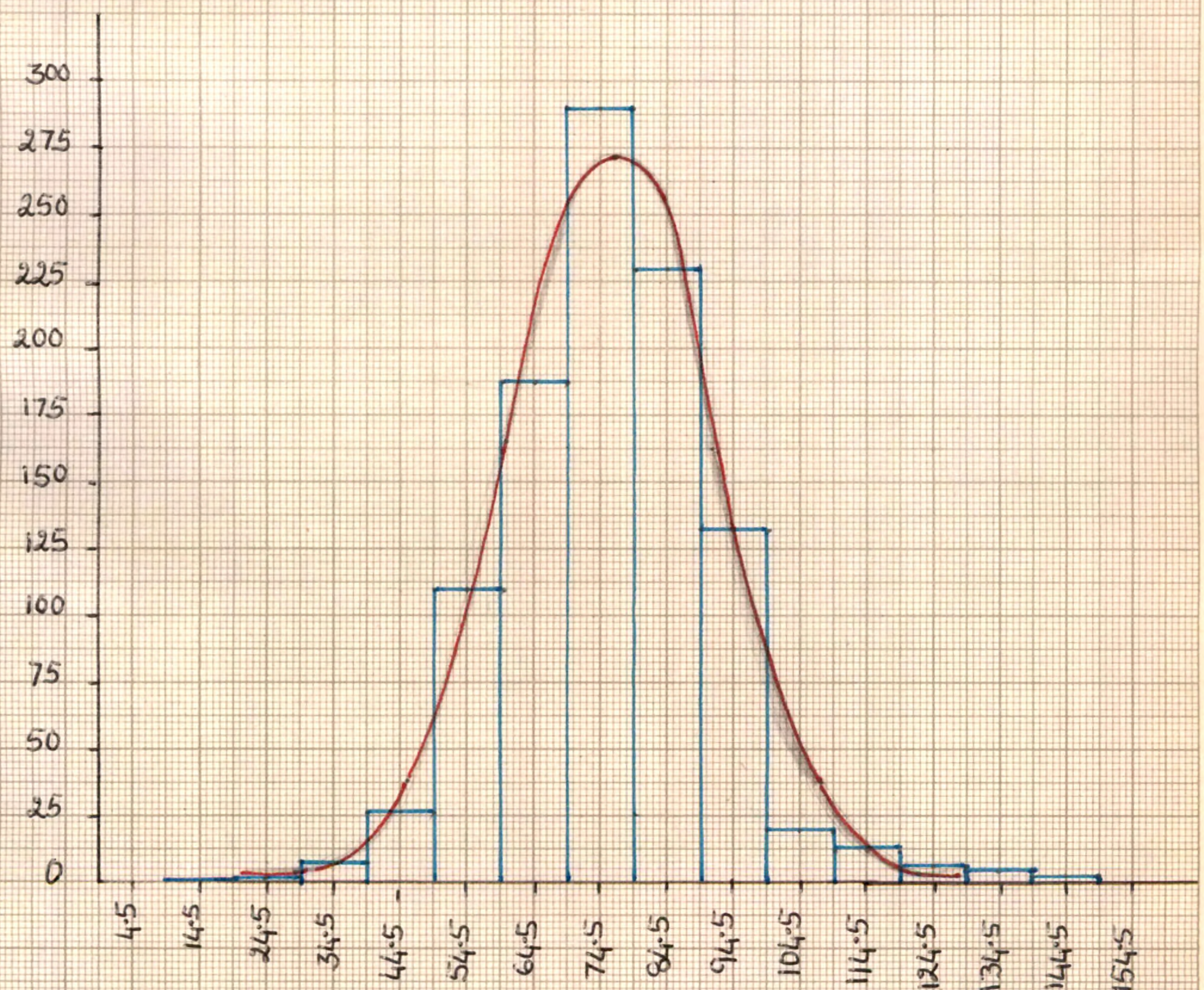
Graph No. 22

Best-Fitting Normal Curve Superimposed On Histogram
Age Group 1 6 years.



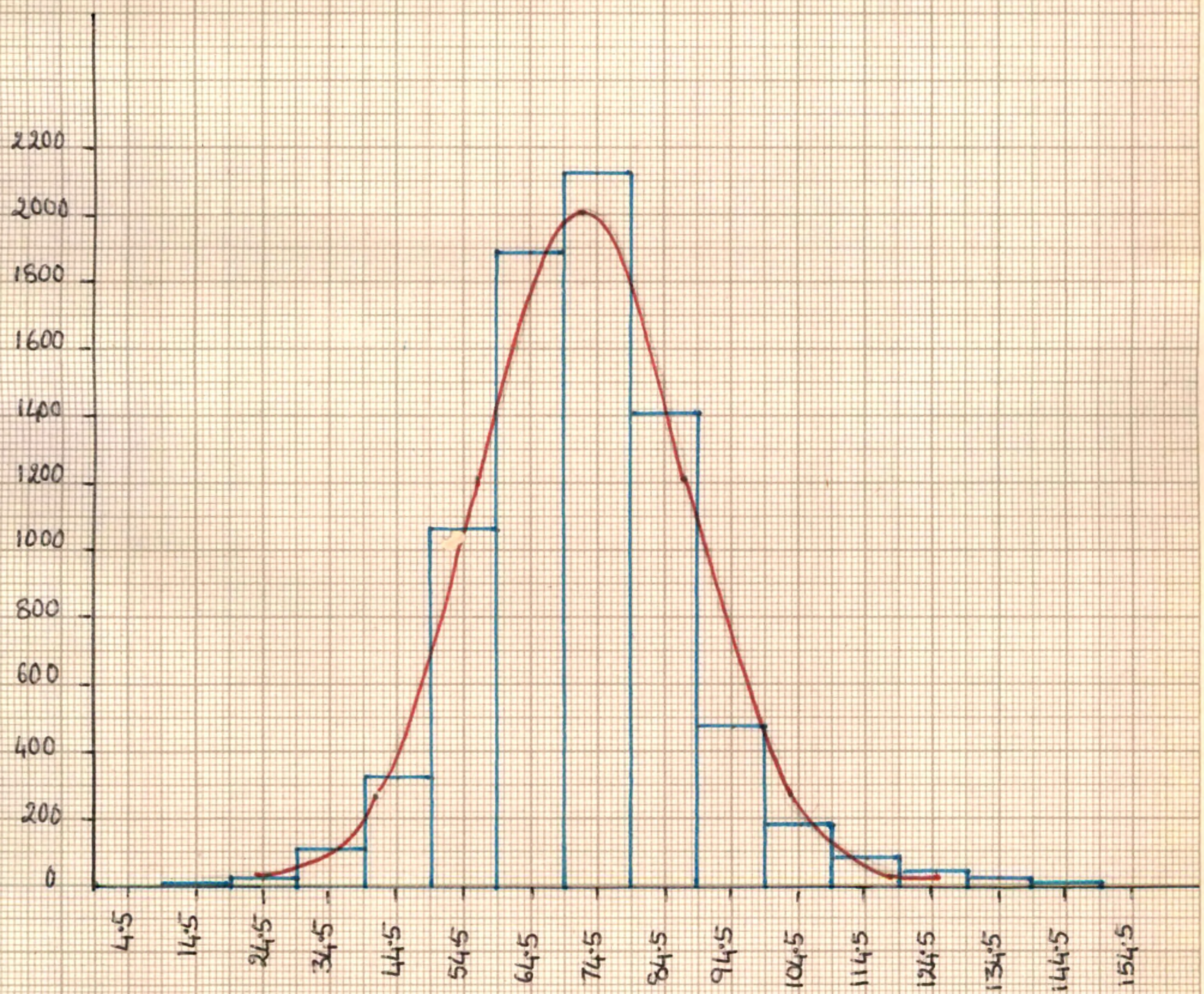
Graph No. 23

Best-Fitting Normal Curve Superimposed On Histogram
Age Group 17 years.



Graph No. 24

Best-Fitting Normal Curve Superimposed On Histogram
Whole Group.



and as the scores are fairly distributed along the normal probability curve, and as the obtained statistics are reliable, the norms established on the basis of these scores will be reliable and stable.

Norm is defined as "A qualitative or quantitative standard of reference assumed to be typical of a given population and therefore used as a base with which individuals or groups may be compared".[1] Raw scores can be given meaning only by refering it to some type of group or groups.

"Test norms may be defined as estimates of some characteristic of a distribution of test scores for a specified population".[2]

Norms are sometimes confused with the word standards. However they are quite different. As Flanagan points out, "Norms describe the actual performance of specified groups of individucals. Standards on the other hand, are desirable, or desired levels of attainment, preferably expressed in terms of outcomes of instruction".[3]

There are two ways of expressing the norms. "One approach describes a score in terms of the group for which it is average. Age norms (or age equivalents) and grade norms (or grade equivalents) exemplify this approach. The other approach describes a score in terms of whatever relative standing corresponds to it in a

1/Goodinough E.L., Op.cit., pp. 558.

2/Flanagan J.C., Op.cit., pp. 698.

3/Ibid, pp. 698.

particular defined group. Percentile ranks, standard scores, and normalized standard scores are by far the most common methods used in this approach".[1]

Types of Norms.

The four types of norms used for educational and psychological tests are as mentioned below:-

a. Age norms

b. Grade norms

c. Percentile norms

d. Standard score norms.

a. Age Norms.

"The norm for any age, in this sense, is the average value of the trait for persons of that particular age".[2] In this sense it is nothing more than the average value. The score of the individual is matched to the average score of a particular age group. For this, average of the scores are established for successive age groups. We say that the score of the individual is equal to that of a particular age group.

The major advantage of this type of norm lies in its familiarity and convenience. The assumption (on which these are based) that one year's growth can be treated as a standard and uniform unit is doubtful. "A trait showing no continuous improvement over an age range (......) cannot possibly be expressed in terms of a

1/Schrader W.B., "Norms". Encyclopedia of Educational Research Harris C.W. (Editor, The Macmillan Company, New York, 1960. pp.925

2/Thorndike R.L. & E. Hagen, Measurement and Evaluation in Psychology and Education, John Wiley & Sons, Inc. New York, London, 1961. p. 127.

scale of age units".[1/]

b. Grade Norms.

They are similar to age norms in many respects. Reference groups are successive grade groups instead of age groups. The average scores for each of the successive grades like school standards are established. The individual's score is matched with the average score of a group and is said that individuals performance is equal to that of a particular grade.

There is also the problem of equality of units. These are useful in interpreting the academic achievement of an individual in the schools. They are of very little use for other types of measures.

c. Percentile Norms.

These norms interprete score of an individual in terms of his standing in some particular group. For each score we can find the percentage of cases falling below it. So percentile rank can be assigned to any score in the distribution. However we will have to use different norm groups for different ages and different grades. There is again the problem of equality of units, because the difference between the 50th and 55th percentiles is not the same as that between 20th and 25th percentiles. The units are typically and systematically unequal.

d. Standard Score Norms.

These norms also try to interprete the scores of an individual

---

1/Ibid, pp. 130

with those of a reference group. They are expressed in terms of the standard deviations of the reference group. The scores are expressed in terms of SD of the group from the mean and these units are equal.

This introduces the use of the + and - signs and also of the decimals. These can be avoided by converting the scores on some assumed mean and standard deviation.

"These standard scores have the distinctive feature that they are guaranteed to have a normal distribution, at least for the population comparable to that on which the original norms were obtained"[1].

Moreover the unit of score, in terms of standard deviation has essentially the same meaning from one test to another.

Thus in psychological measurements the standard score norms are generally used.

<u>Units</u>.

The next consideration in establishing norms is to decide about the units of expressing the ability. As pointed out by Thorndike and Hegen,[2] units should have below mentioned properties:

1) They should possess uniform meaning from test to test.

2) They should have uniform size.

3) They should have a zero point.

The characteristic no. 3 of the units cannot probably be

1/*Ibid*, pp. 140.

2/*Ibid*, pp. 126.

obtained in the psychological measurements. There has been no person, so far, with zero intelligence. Thus we cannot add or multiply the intelligence of different persons. "If you put together two morons, you will not get a genius...".[1/]

"The fact that a given individual is unable to pass any of the items of a given test is, of course, no indication that he is completely lacking in the ability which the test is designed to measure".[2/]

It will be in the fitness of things to trace the evolution of the units of measurement of intelligence.

Mental Age.

Binet was the first person to assign a numerical value to intelligence. In 1908 he introduced the concept of mental age as an interpretive measure of intelligence. He established, so to say, the age norms for measuring intelligence.

The items in the test are grouped according to age levels. The child is assigned the mental age of the highest year level in which he passes all items or all but one items. To this basic year is added the age in proportion to the number of items done from the higher age groups. These are called age scales. Some times, scores are assigned according to the number of item done correctly and the score earned is converted into mental age from the specially prepared tables.

1/*Ibid*, pp. 127.

2/Goodinough F.L., *Op.cit*. pp 145.

In both cases the performance on standardized items is compared with that of the average age group of a representative sample of each age level. "Hence we define mental age as the level of development in mental ability expressed as equivalent to chronological age at which the average group of individuals reach that level".[1/]

The concept of mental age has a very important place in the history of development of units for measuring intelligence. But while using it, there are following disadvantages.

1. The test provides a very limited sample of universe from which the items are selected
2. The reference standards are derived from different samples
3. Spacing of the items in this scale is a function of time and not of growth. Though the mental growth takes place with reference to time there is no evidence to show that it is proportional to time.

This concept is able to indicate that the child is as bright as the child of a certain chronological age. But it is not able to tell howmuch better the child is than the average child of that chronological age.

The Intelligence Quotient.

Attempt was made to represent intelligence in terms of growth of mental age with chronological age. As Goodinough points out,

1/Freeman F.S., Theory and Practice of Psychological Testing. Holt, Rinehart and Winston, New York, 1963. pp. 131.

"the use of the ratio of mental age to chronological age, ..., was first proposed by Stern and also by Kuhlmann as early as 1912, but it was not until Terman made it an integral part of his 1916 revision that it attained the widespread popularity it holds today".[1/] It was Terman who named the ratio as intelligence quotient. If the ratio MA/CA is used as it is, it would involve fractions and the use of it would become more complicated. Terman multiplied the ratio by 100 to remove the decimals. Thus IQ is calculated by using the formula,

$$IQ = \frac{MA}{CA} \times 100$$

where

IQ = intelligence quotient,

MA = Mental age,

and CA = Chronological age.

If the rate of growth of mental age and chronological age is the same, which should happen in case of normal child, mental age will be equal to the chronological age and hence the ratio MA/CA will be equal to one and IQ of the normal individual shall be 100.

It was found that IQ's of the same individual at different ages, as measured by this 1916 scale was relatively constant. This fact was responsible in making the concept of IQ as a measure of intellectual ability very popular. It was assumed that this was a very stable unit. So the studies were made to investigate the

1/Goodinough F.L., Op.cit., pp. 161.

problem of stability of IQ which is usually represented in terms of correlation between scores at different ages or in terms of the average number of IQ changes. Some of the studies mentioned by Pinneau [1/] are given below:-

By Bayley on the Berkeley Growth Study; by Honzik, and by Honzik, Macfarlane and Allen on the Guidance Study; by Ebert and Simons on the Brush foundation study; by Sontag, Baker and Nelson on the Fels study

"Evidence .... shows that on several available tests a given conventional IQ value indicates a different relative position at different ages. As a consequence, a child's IQ on such tests must change as he grows older if he maintains his same standing relative to those of his age". [2/]

"On the average there is greater change in conventional IQ for bright children from test to retest on a Stanford-Binet scale than for those of less ability of the same age". [3/] The IQ changes at ages 10, 11, 17 as compared with their IQs at age 6, in Berkely Growth Studies [4/] were 12.4, 14.8 and 11.4. This difference has been shown to be significant. These shortcomings create problems in interpreting the relative meaning of IQ.

1/Pinneau S.R., Changes in Intelligence Quotient Infancy to Maturity. Houghton Mifflin Company, Boston 1961. pp. 10.

2/*Ibid*. pp. 29

3/*Ibid*. pp. 46.

4/*Ibid*. pp. 16

Deviation IQ (DIQ)

In order to avoid the problem in using the IQ, the concept of 'deviation IQ' is recently used as a unit for representing the intelligence. This is an adaptation of the standard score (Z) technique.

The mean weighted score of the group is given a deviation IQ value 100. The standard deviation of IQ in the population is found to be 15 to 16. Thus the raw scores are converted into Standard scores with a mean 100 and SD 15 or 16, as found with Stanford-Binet which has been regarded as criterion. Thus the choice of values of mean as 100 and SD 15 or 16 are not arbitrary but have been taken to represent the general distribution of IQ in the population. "The principle is that an individual's intelligence quotient should be determined by the relative extent to which his score on the test deviates from the mean of his age group, and that an intelligence quotient of a given value should have the same relative significance throughout the age range".[1]

This unit has many advantages.

Firstly the mean 100 and SD 16 for deriving standard scores are not taken arbitrarily.

Secondly "a standard deviation of 15 points is similar to the one to which psychologists and educators have become accustomed and in which values at each of the several levels have acquired

1/ Freeman. F.S., Op.cit., pp. 133.

qualitative significance".[1/]

Thirdly deviation IQ is useful at age levels 16 or 18 years. It was found by Terman that the average mental age of the adult was between 15 and 17, with the assumed midpoint at 16 years. This means that in the case of an average adult, his maximum level of measured intelligence is reached at the age of 16 and that there are no increments thereafter. So conventional formula for calculating IQ (MA/CA) are considered to be inappropriate for children older than 16 years. However Deviation IQs can be found out in case of children older than 16 years.

Lastly, "DIQs are more stable with age than are conventional IQs".[2/]

In the 1960 revision, Terman and Merrill have used the mean of 100 and standard deviation of 16. Dr. V.V.Kamat and Dr. N.N. Shukla have pointed out that this standard deviation for Indian children is 16.4.

So it may be concluded that "percentiles, standard deviations, standard scores,and intelligence quotients are intimately related. Whatever index is used its principal significance is found in the relative rank it represents in its psychological, educational and vocational connotations ".[3/]

1/Ibid, pp. 133-134

2/Pinneau S.R., Op. cit., pp. 47.

3/Freeman F.S., Op.cit., pp. 134.

IQs in this case are in the form of standard scores with a mean of 100 and standard deviation of 16.4 (as per Dr. Kamat and Dr. Shukla).

The scores had already been grouped into five consecutive age groups from 13 to 17 years. As the mental ability grows with age, it is necessary to prescribe different norms for different age groups.

It is worth establishing different norms for different age groups, only if the performance of each age group on this test is significantly different from the consecutive age group. As the performance of the group is evaluated in terms of the mean, the difference between the means of consecutive age groups should be significantly different.

Table 35. Means, Difference of Means and Critical Ratios of consecutive age groups.

Age	13		14		15		16		17
N	1320		1825		2088		1484		1028
SD	,15.47		15.24		16.02		15.49		15.04
σ_M	.4258		.3567		.3588		.4022		.4690
σ^2_M	.1772		.1273		.1287		.1617		.2195
σ_D		.6191		.5059		.5388		.6173	
Mean	67.824		70.335		72.180		74.132		76.06
D		2.511		1.845		1.952		1.928	
CR		4.056		3.646		3.707		3.123	

As all the CRs are more than 2.58, the difference between the means of any two consecutive age groups is significant at 0.01 level of significance. Thus the performance of each age group is significantly different from that of the consecutive age group. So, different norms for different age groups may be established on the bases of the scores obtained on this test.

Sex Differences.

Various studies have been made to see whether there is any relation between sex and intelligence. Some of them are, by Freeman and Flory based on the VACO tests, by Miles and Miles based on Otis self administering test, Jones and Conard based on Army Alpha tests etc. Conclusions drawn from all such studies may be stated in the following words of Tyler:

"It seems that at the present stage of mental measurement it is not possible to give an unequivocal answer to the question, which sex is superior in intelligence ? When sex differences are found on specific tests, the results may reflect opportunity for learning cultural expectations and attitudes towards the content of the tests".[1/]

The results obtained in case of boys and girls, in this study, are as below:-

	Boys	Girls
N	6201	1544
Mean	72.036	71.788

1/Tayler F.T., "Individual and Sex Differences". Encyclopedia of Educational Research, Harris C.W. (Editor),The Macmillan Company, New York, 1960. pp. 685.

	Boys	Girls
SD	15.76	15.50
σM	.2002	.4426
$(\sigma M)^2$	.04006	.1954

$\sigma D = \sqrt{(\sigma M_1)^2 + (\sigma M_2)^2}$ .23546

$D = (M_1 - M_2)$ .248

$$CR = \frac{D}{\sigma D} = \frac{.248}{.23546} = 1.053$$

As the obtained critical ratio is less than 1.96, this difference in the means of the scores of boys and girls is not significant. This proves the null hypothesis that there is no significant difference in the means of scores of boys and girls.

No separate norms for boys and girls have been prescribed because,

a) there is no unequivocal evidence to support the need of separate norms from the studies done so far,

b) the present study also does not ask for it,

and c) the size of the sample of girls is not adequate enough to find them.

Thus only one set of norms in terms of standard scores with a mean of 100 and standard deviation of 16.4 have been prescribed for each age group from 13 to 17 years.

Standard Scores

Standard score is "A score expressed in terms of the number of standard deviations by which it exceeds or falls below the mean of the group to which the study belongs".[1/] Following procedures are followed for converting the raw scores into standard scores with M = 100 and SD = 16.4

1. The obtained raw scores were first rounded up to the nearest integers.

Table 36. Obtained means and rounded means of different age groups.

Age Years	13	14	15	16	17
Obtained Mean	67.824	70.335	72.180	74.132	76.06
Rounded up Mean	68.000	70.000	72.000	74.000	76.00
Difference	0.176	0.335	0.180	0.132	0.06
SE_{Mean}	0.4258	0.3567	0.3588	0.4097	0.4690

(figures indicating SE Mean are adopted from Table No.25 on page 191)

From the table above it can be seen that the differences in the obtained and rounded up means are less than the SE of the means and thus rounding up of the means is not going to have any significant effect on the results based on them. However this change simplifies the computations to a great extent.

2. The SD of each age group is slightly different from the other. So σ unit for each age group is different. Consequently σ unit for unit deviation of the raw score would be different for each age group. They will have to be computed separately for each

1/Goodinough F.L., Op.cit., pp. 566.

group. Dr. Desai has followed a procedure which simplifies the calculations to a great extent. As the difference in SD of each group is very small he has assumed that the SD of each group is the same and has selected a value arbitrarily which is almost the average of SD of all groups, and has treated it as SD of all the groups. Thus σ value of each unit deviation of the raw score from the mean is the same. Thus no such computations for each age group will be necessary. This procedure simplifies camputations without affecting the results significantly.

Table 37. SD of each age group and the arbitrarily assumed SD of the groups.

Age group yrs.	13	14	15	16	17
Obtained SD	15.47	15.24	16.02	15.78	15.04
Assumed SD	15.50	15.50	15.50	15.50	15.50
Difference	0.03	0.34	0.52	.28	0.46
SE_{SD}	0.3024	0.253	0.2548	.2909	0.331

The actual deviation of obtained SD from the assumed SD seems to be more than SE in case of age groups 14, 15 and 17. The maximum and minimum errors for a deviations of 50 scores from the mean (which is more than 3 σ units where SD is 15.5) introduced are .4 and 1.8 IQ units. Thus 15.5 can safely be treated as the SD of all age groups.

3. The formula used for converting the raw scores into

standard scores is as below:-

$$\frac{X' - M'}{\sigma'} = \frac{X - M}{\sigma}$$

or $$X' = \frac{\sigma'}{\sigma}(X - M) + M'$$

where

X = a score on original distribution

X'= a standard score on new distribution

M & M' = means of the raw score and standard score distributions respectively

σ & σ'= SD's of raw and standard scores respectively.

In this particular case,

$X' = IQ$

$M' = 100$

$\sigma = 15.5$

$\sigma' = 16.4$

Thus in this particular case the above formula changes to

$$IQ = \frac{16.4}{15.5}(X - M) + 100$$

$$= 1.056\,(X - M) + 100$$

$X - M$ is the deviation of the raw score X from the mean of the group to which it belongs. $\frac{\sigma'}{\sigma}(X - M)$ indicates the deviation of IQ from the mean IQ(100). From this deviation of IQ corresponding to each multiple of unit deviation of raw score is calculated.

Table 38. Deviation of IQ from the Mean IQ 100 for every unit deviation of raw score from the mean.

Deviation of raw score from Mean	Deviation of IQ from Mean IQ 100	Rounded up deviation IQ
0	0.000	0
1	1.056	1
2	2.112	2
3	3.168	3
4	4.224	4
5	5.280	5
6	6.336	6
7	7.392	7
8	8.443	8
9	9.504	10
10	10.560	11
11	11.616	12
12	12.672	13
13	13.728	14
14	14.784	15
15	15.840	16
16	16.896	17
17	17.952	18
18	19.008	19
19	20.064	20

(continued on next page)

Table 38. (continued)

1	2	3
20	21.120	21
21	22.176	22
22	23.232	23
23	24.288	24
24	25.344	25
25	26.400	26
26	27.456	27
27	28.512	29
28	29.568	30
29	30.624	31
30	31.680	32
31	32.736	33
32	33.792	34
33	34.848	35
34	35.904	36
35	36.960	37
36	38.016	38
37	39.072	39
38	40.128	40
39	41.184	41
40	42.240	42

(continued on next page)

Table 38. (continued)

1	2	3
41	43.296	43
42	44.352	44
43	45.408	45
44	46.464	46
45	47.520	48
46	48.576	49
47	49.632	50
48	50.688	51
49	51.744	52
50	52.800	53
51	53.856	54
52	54.912	55
53	55.968	56
54	57.024	57
55	58.080	58
56	59.136	59
57	60.202	60
58	61.258	61
59	62.314	62
60	63.360	63
61	64.416	64

(continued on next page)

Table 38. (Continued)

1	2	3
62	65.472	65
63	66.528	67
64	67.584	68
65	68.640	69
66	69.696	70
67	70.752	71
68	71.808	72
69	72.864	73
70	73.920	74
71	74.976	75
72	76.032	76
73	77.088	77
74	78.144	78
75	79.200	79
76	80.256	80
77	81.312	81
78	82.368	82
79	83.424	83
80	84.480	84
81	85.536	86
82	86.592	87

(concluded on next page)

Table 38. (concluded)

1	2	3
83	87.648	88
84	88.704	89
85	89.760	90

The actually obtained IQ deviations have been rounded up to the nearest integers to simplify computations.

The above table can very easily be used for computing the IQ corresponding to a particular score of an individual belonging to any age group from 13 to 17 years. The steps are as described below:-

1) Substract from the obtained score X, the mean of the age group to which it belongs. This is the deviation of the score from the mean. It may have + or - value.
2) From the above table find out the rounded up deviation of IQ corresponding to that deviation of raw scores. Put + or - sign to this deviation according to the sign possessed by the deviation of the raw score.
3) Add this deviation to 100 taking into account the sign possessed by the IQ deviation.

This sum represents IQ corresponding to the raw score X. Any raw score, the corresponding deviation IQ of which is more than

30, will be treated as 30.

Anticipating that the user is likely to commit errors in computing the IQs, table is provided in the mannual from which IQ corresponding to any score belonging to any of the age groups can be read directly.

Now the test is ready for use because,

1. its contents have been standardized by finding the validity, discrimination power and difficulty index of each item and by arranging them according to the decending order of difficulty indices,
2. the procedures for the administration and scoring have been standardized,
3. the time limits for the test have been fixed,

and 4. the norms have been prescribed for evaluating the obtained scores, by administering the test on an adequately large representative sample of a defined population for whom the test has been constructed.

Preparation of the Manual

For getting reliable and valid results, the test is to be administered, performance is to be scored and the obtained scores are to be interpreted according to the standard procedures laid down for the purpose. Thus manual for this purpose is needed.

To guide the administrators of the test, a manual has been prepared. It contains information regarding the following:-

1. A brief description of the test.

2. Sitting arrangement
3. Instructions to the administrator
4. Instructions to pupils
5. Manner of scoring
6. Evaluating the obtained scores
7. Norms for each age group.

References.

1. Ballard F.B., Group Tests of Intelligence. University of London Press Ltd., London E.C.4. 1957.

2. Brown F.G., Principles of Educational and Psychological Testing. The Dryden Press Ind., Hinsdate Illinois, 1970.

3. Cronbach L.J., Essentials of Psychological Testing. Harper & Row, New York and John Weatherhill, Inc., Tokyo. 1965.

4. Dyal J.A., Readings in Psychology: Understanding Human Behaviour. Mc Graw-Hill Book Company, New York. 1967.

5. Freeman F.S., Theory and Practice of Psychological Testing. Holt, Rinehart and Winston, New York. 1963.

6. Garrett H.E., Statistics in Psychology and Education. Longmans, Green and Co., New York, London, Toronto. 1954.

7. Goodenough F.L., Mental Testing., Holt, Rinehart and Winston, New York. 1961.

8. Gronlund N.E., Readings in Measurement and Evaluation. The Macmillan Company, New York and London. 1968.

9. Guilford J.P., Psychometric Methods. McGraw-Hill Book Company, Inc., New York. Kogakusha Company Ltd., Tokyo. 1954.

10. Harris C.W., Encyclopedia of Educational Research. The Macmillan Company, New York. 1960.

11. Hunt J.MCV., Intelligence and Experience. The Ronold Press Company, New York. 1961.

12. Jenkins J.J., Paterson D.G., Studies in Individual Differences. Methuen & Co., Ltd., London W.C. 2.1961.

13. Lindquist J.C., Educational Measurement. American Council on Education, Washington D.C. 1966.

14. Mc Call J.R., Sex Differences in Intelligence. The Catholic University of America Press, Washington D.C. 1955

15. Pinneau S.R., Changes in Intelligence Quotient Infancy to Maturity. Houghton Mifflin Company, Boston. 1961

16. Thorndike R.L., E.Hagen., Measurement and Evaluation in Psychology and Education. John Wiley & Sons, Inc, New York London. 1961.

17. Tyler Leona E., The Psychology of Human Differences. Appleton Century-Crofts, New York. 1956.