AN ILLUSTRATIVE STUDY OF EXAMINATION MARKS BY THE METHODS OF FACTOR ANALYSIS AND THE ANALYSIS OF VARIANCE

Extract of Thesis to be presented for the Degree of Ph.D. in Psychology in the University of London, June 1942.

Introduction

The essence of an examination is stated to be measurement and the thesis analyses the measurements given by some examinations. The methods of scoring examination marks and the application of them are outlined.

(i) Factor Analysis of Marks with different Types of Scoring

The marks of two university examinations for eight consecutive years are analysed. In the case of one examination, the instructions given to the examiners demand that the marks should approximately fit a normal distribution curve. The other examination papers are marked without restrictions. In this section an endeavour has been made to determine whether the instructions to fit the marks to a normal distribution curve affect the factors obtained by analysis.

(ii) Analysis of Variance of Two Sets of Scores for Some Examination Candidates

The college assessment and the examiner's mark for one subject in a university examination are analysed to test the significance of the difference between the variances of the two sets of scores. The revising examiner has marked independently some of the border-line cases together with some of the poorest and best papers picked at random. The significance of the difference between the variances of these re-markings and the original marks is also tested.

(iii) Correlations between Persons

The marks of an examination for entrance to secondary schools are analysed. In this case the same candidates are examined in both 1937 and 1938. One purpose of this analysis is to determine the change, if any, in the factor saturations for the set of persons after the lapse of a year, using Burt's method. The second aim is to try out another method which has recently been evolved and to test the agreement of the results obtained by this method with those obtained by using Burt's method.

Results and Conclusions

The main conclusions are:-

(i) According to the data used in this research, fitting the score-scatter to a normal distribution curve does not influence the resulting factors. Some interesting facts concerning the individual examinations are revealed.

(ii) In many cases the difference between the college assessment and the examiner's marks is significant. There is evidence that there is no difference in the average mark assigned by the first and revising examiners, and that the marks do bear some relation to the ability of the individual concerned.

(iii) There is reasonable agreement in the first factor saturation of the marks of both years but there are differences in value and sign in the second and third factor saturations. The other method used gives results which agree to the degree of accuracy used.

Mary K.B.Harwood.
23rd April 1942.

AN ILLUSTRATIVE STUDY OF EXAMINATION MARKS BY THE METHODS OF FACTOR ANALYSIS AND THE ANALYSIS OF VARIANCE

MARY K.B. HARWOOD, M.A., M.Sc.,

UNIVERSITY OF LONDON UNIVERSITY COLLEGE

THESIS PRESENTED FOR THE DEGREE OF

Ph.D. IN PSYCHOLOGY

IN THE

UNIVERSITY OF LONDON, DECEMBER 1942.

CONTENTS

CHAPTER I

INTRODUCTION

CHAPTER II

SOME OF THE EFFECTS CAUSED BY MARKING IN ACCORDANCE WITH A NORMAL DISTRIBUTION

CHAPTER III

ANALYSIS OF VARIANCE BETWEEN TWO ESTIMATES OF MARKS FOR A UNIVERSITY EXAMINATION

CHAPTER IV

ANALYSIS OF MARKS AWARDED TO CHILDREN WHO WERE EXAMINED TWICE

CHAPTER V

CONCLUSION

APPENDIX

TWO ALTERNATIVE METHODS TO OBTAIN FACTOR MEASUREMENTS OF PERSONS

CHAPTER I

INTRODUCTION

CHAPTER I

INTRODUCTION

Aim of This Thesis

This thesis analyses the marks given by some examinations, and in particular it investigates:

(i) The influence of enjoining examiners to fit their marks to the normal distribution curve.

(ii) The variance between a personal assessment and an independent examination mark.

(iii) Correlations between persons when a second measurement was made a year after the first.

For many years correlations have been applied to educational tests which are virtually school examinations; naturally enough, psychologists have applied the same method to the examinations which, in their academic capacity, they are constantly carrying out themselves. Though much factorial work has been applied to school tests, little factor analysis has been made of university examinations. The first, probably, was the unpublished analysis made by Burt with results of the same examination as that with which Examination I in Chapter II is concerned. However, these analyses never really answered the frequently asked question as to what particular aspects of teaching capacity the examination measures. The analyses of this research show that the examinations considered measure certain special aptitudes other than general intelligence. They also measure the relative weights of those aptitudes year after year.

Examination Scores

One of the main issues of an examination is the placing of a set of candidates in a relative order of merit. The system of scoring should be suitable to the purpose of the examination. The examiners in many universities mark their degree and post-graduate examination papers with a system of symbols such as $\alpha+$, $\beta-$, γ, etc., and this system has been found most satisfactory where the number of candidates is small. It is thought that by this method there is a greater likelihood of maintaining a constant pass-level for different groups of candidates, a task which is very onerous and calls for experience and sound judgment coupled with a somewhat rare ability in setting suitable examination papers. When some thousands of candidates are entered for an examination, the more objective method of fixing the pass-level at a point such that a definite percentage always pass is used. Some boards of examiners are instructed that the marks that they allot to the candidates should fit the normal frequency curve. In this way about 40% of the candidates score between 40% and 60%. In some examinations, the teacher submits a mark on the relative merit of the candidates he is entering. Though this estimate is not altogether reliable as it is influenced by personal bias, it is a means of helping to secure fair treatment for a candidate who is nervous or not well at the time of the examination.

CHAPTER II

SOME OF THE EFFECTS CAUSED BY MARKING IN ACCORDANCE WITH A GIVEN DISTRIBUTION

CHAPTER II

SOME OF THE EFFECTS CAUSED BY MARKING IN ACCORDANCE WITH A GIVEN DISTRIBUTION

Material

By the kindly aid of two examiners and with the permission of the university authorities in question, the sets of marks for two different examinations for eight consecutive years were obtained. The university required that neither the names of the examinations nor the actual marks be disclosed. Only the following few facts about these marks can therefore be mentioned.

(a) Examination I These are the marks for a post-graduate examination set in the years 1931 to 1938, inclusive, the subjects of which are:

(i) Principles of Education
(ii) Psychology and Hygiene
(iii) Special Method
(iv) Educational Psychology or Comparative Education or History of Education
(v) Educational System
(vi) Essay
(vii) Practical Teaching

All the candidates have the same question papers in the first two and the fifth subjects. The fourth subject comprises three different courses only one of which may be chosen. In actual practice, Educational Psychology is very much more popular than the other two, so that the marks for the fourth of the papers which the candidates have to take may be treated as the marks for one main subject only. The questions in Special Method, also Essay and Practical Teaching, are related to the subject in which the candidate took his degree. The candidates, both men and women, are graduates of the faculties of arts and science: there are about 400 entries each year, only few of whom are graduates of other universities. Owing to the high standards of their selection, these men and women form an extremely homogeneous group.

Examination II The marks, again for the years 1931 to 1938 inclusive, are from an undergraduate examination whose entrants are all women. The names of the subjects may not be disclosed so they will be described as follows:

(i) P, an organic science
(ii) Q, an inorganic science
(iii) R, a mathematical science
(iv) S, a social science

The four subjects are compulsory for all candidates and the number examined is small, rising from 38 in 1931 to 63 in 1938. All the candidates belong to the same faculty and they also may be considered to form an extremely homogeneous group.

Examiners and Their Methods of Scoring

Different methods were used in the scoring of the two examinations and the purpose of the analysis of these marks is to attempt to estimate the influence of the method used on the marks obtained. In Examination I, the examiners were instructed to mark their papers so that the resulting scores form a normal distribution curve. Also they were to use literal marks, A+, A, A-, B+, B etc., to E-, i.e. 15 grades. The examiners of Examination II received no instructions about the scatter of their marks; they formed their own maxima and then calculated the marks as a percentage. In this set there were no changes in the examiners throughout the eight years, while in Examination I there were a few changes. The influence of these changes will be discussed with the results. The distributions of marks for the eight consecutive years for both examinations are given on pages 10, 11, 12, 13 and 14. (figs. I, II, III).

Correlations Obtained

As early as 1917 Burt stated that "school achievements are due to mental qualities of two kinds: first, a general ability entering into all school work; secondly, special aptitudes for particular subjects."[1] After collecting test

---

[1] Burt, C. The Distribution and Relations of Educational Abilities, P.S.King & Son Ltd., London, 1917. p. 64.

results from many hundreds of children of all ages and capacities he wrote "with younger children, and particularly, it would appear, with younger girls, one can often demonstrate little but the existence of the general factor; with older children, and particularly with college students, little but specific talents or specialised interests."[2] The generally-accepted view seems to be that human abilities differentiate or become specialised more and more as the children examined get older and pass through the adolescent to the adult stage. Burt suggests that one important factor at work is the increasing selective character of our educational system. The children in an infant school form an almost perfect sample of the population of the age they represent and hence correlations between test results should be high. Heterogeneity is lessened in the junior schools with the removal of the mental defectives. The senior, central and secondary schools give three definite levels of ability in the child population of their age group. Finally, college students form groups with varying degrees of homogeneity according to the type of college, the age of the student and the ability level of the course which the student is attending. It has been found that as the degree of homogeneity of groups increases, so the correlations

---

[2] Burt, C. Mental and Scholastic Tests, P.S.King & Son Ltd., London, 1927. p. 266.

obtained between group test results become smaller. In Examination I, where the candidates are all graduates, the correlation coefficients are all smaller than those of Examination II where the candidates are studnets in their first year of college life. Recently Vernon[3] has found evidence in his own researches which enables him to state "in the educational field at least, our data show that there is positive overlapping of widely diverse abilities among training college students." The results of the present research seem to be in agreement with those of Burt and as far as correlations are concerned with those of Vernon: all his correlations are positive and statistically significant except one. In Examination I of this research, out of a total of 168 correlations, three were not significant, while in Examination II, only one out of 48 was not significant. Again, in Examination I, the correlations are small, the average ranging from .248 in 1935 to .365 in 1932. Vernon found small correlations too, his median being .279. In Examination II, the correlations are larger, ranging from .235 to .881, with an average of .574. (See Tables III & IV, pages 21 and 22).

---

[3] Vernon, P.E. Educational Abilities of Training College Students, B.J.Ed.P., Vol. IX, Nov. 1939. pp. 233-50.

Statistical Methods

A. Correlations

The method used throughout for calculating the correlation coefficients was the formula[4], $R = \frac{X + Y - Z}{2\sqrt{XY}}$, where X, Y and Z are the total square deviations along the two axes ($x = 0$, $y = 0$) and along the ($y = x$) diagonal of the scatter diagram. The correlation was assumed to be significant if greater than $3PE.\sigma_r$, where $\sigma_r = \frac{1 - r^2}{\sqrt{n - 1}}$ and n is the number of pairs of observations.[5]

B. Analysis of Factors

Several methods were tried. As there were many analyses to be made, a method was sought which gave reliable factors and was, at the same time, quick to compute. For rhis purpose, Burt's method by Weighted Summation[6] was found to be most satisfactory. Each of the tables of correlations for the eight consecutive years for Examinations I and II were analysed for two factors and the saturation coefficients will be given later.

---

[4]Thomson, G.H. *How to Calculate Correlations*, Harrap, London, 1924. p. 16.

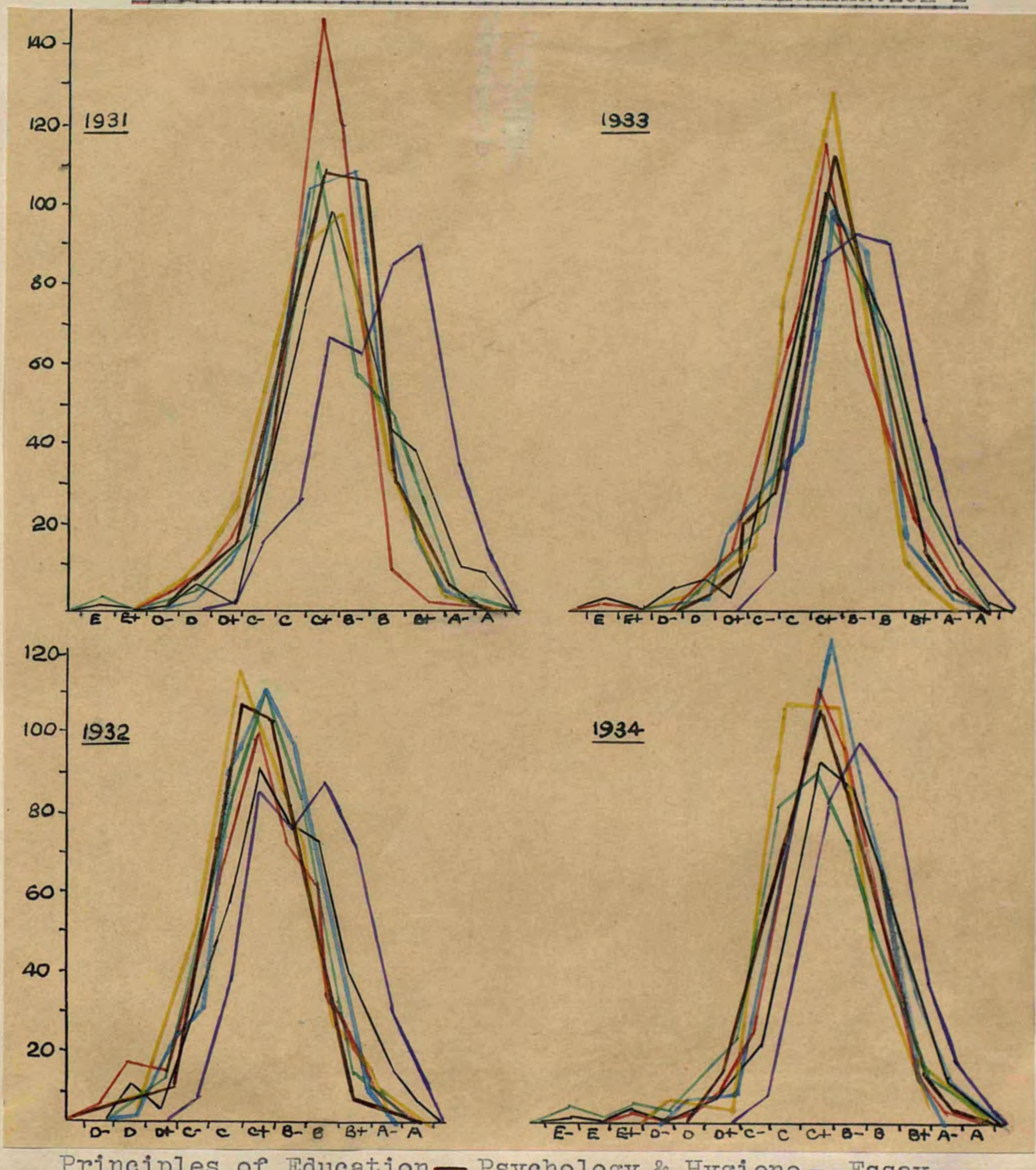
also Kelley, T.L. *Statistical Method*, Macmillan Co., New York, 1924. p. 180. $r = \frac{\sigma^2 x + \sigma^2 y - \sigma^2 (x-y)}{2\sigma_x \sigma_y}$

[5]Fisher, R.A. *Statistical Methods for Research Workers*, Oliver & Boyd, London, 1932. pp. 46, 198 seq.

[6]Burt, C. *The Factors of the Mind*, University of London Press, London, 1940. p. 467.

FIGURE I

Graph to show Distribution of Marks in Examination I



Principles of Education— Psychology & Hygiene— Essay—
Special Method— Educational System— Practical Teaching—
Educational Psychology or Histroy of, or Comparative Education—

FIGURE I contd.

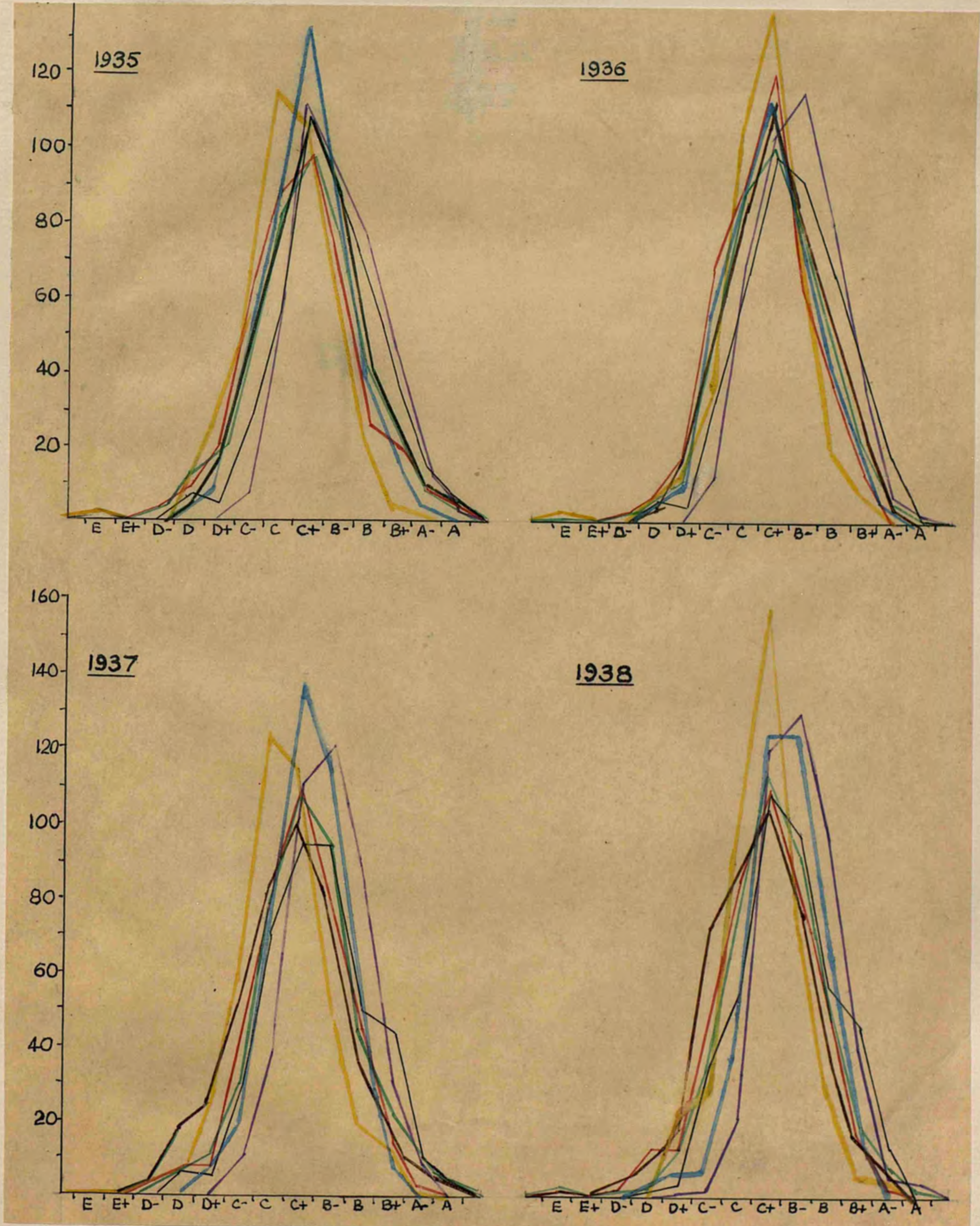


FIGURE II

Graph to show Distribution of Marks in Examination II

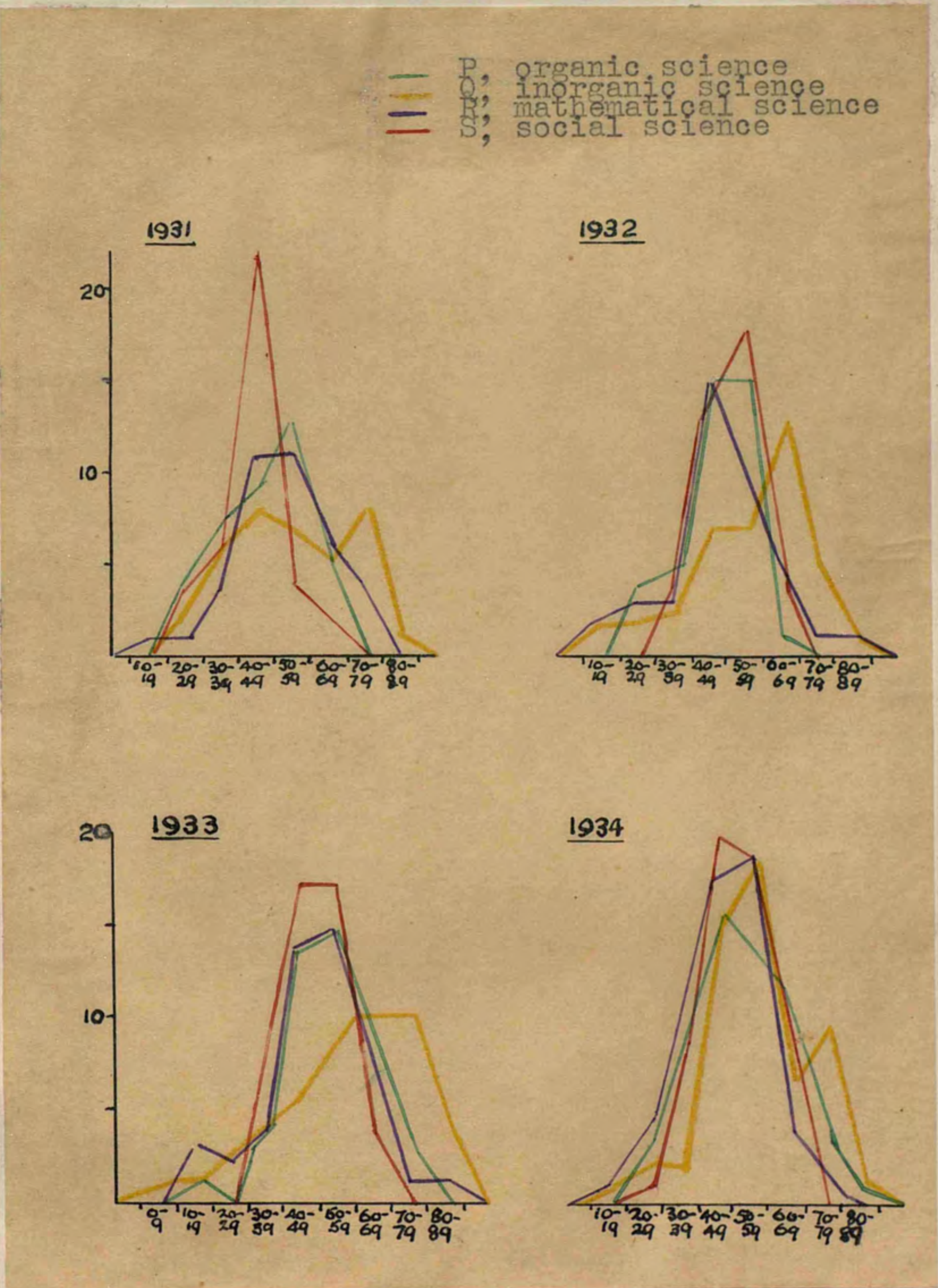


FIGURE II contd.

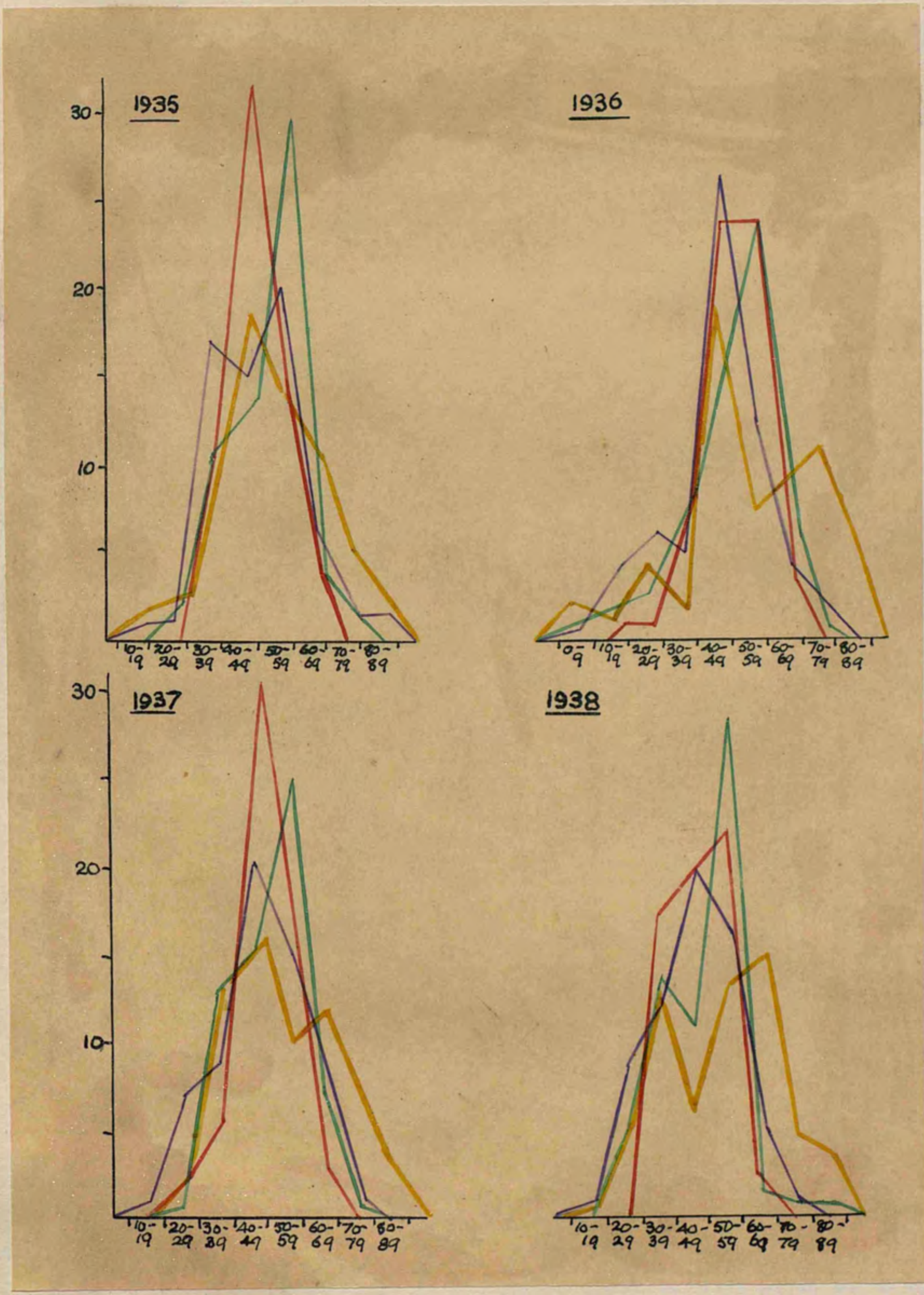
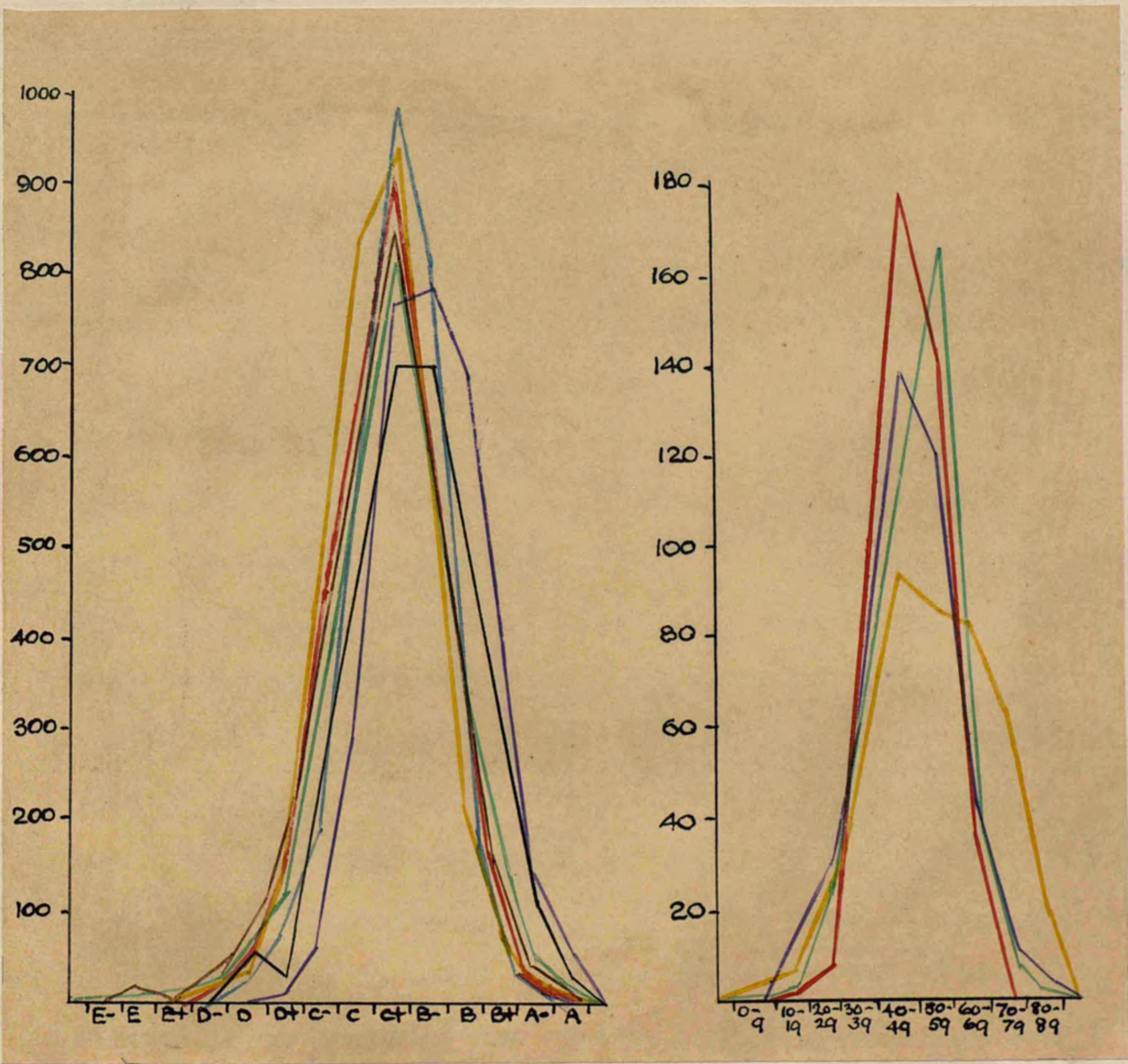


FIGURE III

Graph to show Distribution of Marks Summed for the Years 1931 to 1938

Examination I   Examination II



Results

In Tables III and IV, pages 21 and 22, will be found the correlation coefficients of the marks of Examinations I and II and the saturation coefficients for both examinations obtained by Burt's weighted summation method are given in Tables V and VI, pages 24 and 28. Graphs to illustrate the last two tables are given as Figures IV and V, pages 26 and 29. These tables and figures, together with the distribution curves, will be considered in order.

(a) Distribution of Marks

In Examination I, where it was enjoined that the marks for each subject should fit the normal distribution curve, it will be noticed that the curves for each subject are practically coincident and appear to approximate to the normal distribution shape. The most notable exceptions to this are Special Method and Essay, and the probable explanation for the divergence of these curves from the others is that these subjects are marked by at least six different examiners, though it is highly probable that any one candidate is marked by the same examiner in both subjects. Practical Teaching is marked by the same examiners as the last two mentioned subjects. However, this subject has a curve nearly coincident with the curves of Principles of Education, Psychology and Hygiene, Educational Psychology and Educational System.

In Examination II, although the method of marking is non-restricted, some of the curves appear to approximate fairly well to the normal distribution. Subject Q, inorganic science, is somewhat irregular, five out of the eight curves having more than one peak. Subject S, social science, always has a very narrow range of distribution and the curve is very erratic in 1936 and 1938. This subject is less "exact" than subjects P, Q and R and possibly may be marked by impression.

As the number of candidates for each year in Examination II is small, the mark distributions for the eight years were summed and graphed in Figure III, page 14. The individual pecularities shown in the eight distributions of Examination II, Figure II, pages 12 and 13, are then nearly eliminated and subject Q, inorganic science, has only a mild tendency to be double-peaked. It is still, however, quite markedly different from the others. In Examination I, the Essay distribution, and to a smaller extent, the Practical Teaching distribution, show either that a candidate scores where his choice of subject is wider or that the examiner is influenced by the more personal nature of the examination. One other point is of interest with regard to this set of marks. In 1931, teaching ability was considered by the examiners as the real test of the examination as a whole. The various tutors, therefore, wanted to ensure the success of their individual candidates by giving them a high mark for the subject that influenced the final result. Since 1931 'general suitability for membership of a school staff' which includes a theoretical knowledge of subjects embracing teaching, is considered as a standard for success in the whole examination.

Tables I and II, pages 18 and 19, give the averages and standard deviations of the marks. In Examination I, for each subject throughout the eight years, the average is fairly constant. In Examination II, there is a variation from 42.63 to 55.75 in Subject P, organic science, and from 44.69 to 55.17 in Subject Q, inorganic science. The mathematical science, R, and the social science, S, have a slightly smaller range of variation. A comparison of the standard deviations seems to indicate the restraining influence of marking to fit a normal distribution curve. In Examination I, the greatest variation in the standard deviation is in Essay, namely 5.69 to 9.40 (3.71). Apparently the instructions given out by the head of the examining board did not make it clear that marking to a normal distribution curve was to apply to the students' essays. It will be seen that in 1931 very high marks were given resulting in an average

TABLE I

Table to show Average Marks and Standard Deviations for Subjects in Examination I

Year	A	B	C	D	E	F	G
		Average Marks					
1931	50.20	51.47	48.12	49.07	47.90	55.89	57.31
1932	49.02	50.32	47.36	49.72	50.15	57.53	53.66
1933	50.75	49.61	48.89	50.26	49.84	56.22	52.20
1934	50.60	50.96	48.76	49.55	50.99	55.30	52.83
1935	50.40	50.30	47.73	49,34	48.98	54,75	53,87
1936	49.85	50 27	48.78	46.90	48.33	53.90	52.76
1937	48.74	51.53	47.52	51.06	49.97	54.38	52.74
1938	48.42	53.30	48.88	50.69	49.56	55.39	50.95
		Standard Deviations					
Year	A	B	C	D	E	F	G
1931	7.85	7.22	8.75	8.88	5.69	8.66	9.03
1932	6.63	6.95	7.48	7.77	9.06	8.09	9.22
1933	7.51	6.95	6.80	8.26	8.40	7.46	8.87
1934	7.39	6.38	7.40	8.72	7.11	7.45	8.60
1935	7.96	7.04	6.89	8.99	8.44	7.03	8.36
1936	7.92	6.09	6.86	8.52	7.75	6.63	8.34
1937	8.93	6.05	7.15	9.27	9.40	6.49	8.67
1938	8.27	6.31	6.82	7.97	8.30	5.97	7.45

A Principles of Education
B Psychology and Hygiene
C Special Method
D Educational Psychology or History of Education or Comparative Education
E Educational System
F Essay
G Practical Teaching

TABLE II

Table to show Average Marks and Standard Deviations for Subjects in Examination II

Average Marks				
Year	P	Q	R	S
1931	46.99	48.19	47.31	42.01
1932	42.63	54.97	45.38	51.65
1933	55.23	52.01	54.29	47.01
1934	47.88	55.05	50.78	47.45
1935	51.42	44.69	44.27	47.05
1936	44.29	45.00	43.90	47.17
1937	51.39	55.17	42.87	42.99
1938	55.75	54.75	44.97	44.87
Standard Deviations				
Year	P	Q	R	S
1931	10.97	16.13	13.95	7.40
1932	9.92	16.64	14.35	8.41
1933	11.61	18.73	14.57	8.31
1934	12.85	13.58	11.72	9.54
1935	9.37	14.54	12.33	7.80
1936	13.35	19.67	14.45	8.97
1937	9.77	15.42	13.06	8.35
1938	12.45	17.36	12.45	8.14

P organic science Q inorganic science
R mathematical science S social science

of 55.89. The attention of the examiners was called to this fact and in 1932 a curve with two peaks was produced showing that some of the examiners had complied with the instructions while others had not. The curves and the averages of the marks of the following years, except those of 1936, show that probably the handing back of the essay to the student in question still played a part in the marking of that essay. The standard deviation for Practical Teaching is always high compared with that of other subjects especially in the early years. Students were presumably picked out mainly by academic prowess, so that in theory they are relatively homogeneous. The fairly high number of E marks shows that those with good teaching ability are not necessarily those who have a first class degree.

In Examination II, the standard deviation ranges from 7.40 to 19.67, a difference of 12.27 for the four subjects during the eight years. The standard deviation of Subject Q, inorganic science, varies from 13.58 to 19.67 while the range for social science, S, is only 7.40 to 9.54. It seems fairly obvious from this that marking to fit the normal distribution curve will give the set of marks whose standard is least likely to vary from year to year and whose standard is less likely to vary from that of other subjects in the same year.

TABLE III

Table to show Correlations of Marks in Examination I

	B	C	D	E	F	G	B	C	D	E	F	G
	1931						1932					
A	.437	.353	.425	.550	.095	.127	.486	.410	.342	.395	.143	.222
B		.277	.297	.446	.208	.193		.438	.401	.469	.326	.207
C			.232	.325	.302	.189			.682	.440	.234	.357
D				.455	.082	.151				.457	.300	.405
E					.059	.157					.279	.340
F						.325						.341
	1933						1934					
A	.349	.343	.312	.323	.245	.210	.246	.349	.374	.403	.175	.215
B		.423	.372	.441	.412	.310		.478	.325	.406	.192	.140
C			.431	.362	.304	.443			.330	.354	.265	.295
D				.376	.267	.227				.552	.284	.389
E					.288	.240					.451	.377
F						.406						.423
	1935						1936					
A	.402	.298	.463	.328	.180	.211	.421	.359	.419	.301	.287	.235
B		.315	.323	.274	.086	.133		.316	.360	.340	.208	.147
C			.349	.297	.095	.327			.330	.231	.412	.289
D				.164	.143	.177				.265	,236	.168
E					.121	.087					.218	.158
F						.426						.426
	1937						1938					
A	.287	.354	.336	.456	.244	.169	.409	.363	.372	.344	.272	.273
B		.442	.379	.494	.186	.127		.504	.437	.404	.118	.299
C			.327	.455	.332	.376			.363	.323	.278	.317
D				.434	.246	.268				.483	.243	.255
E					.239	.541					.250	.253
F						.424						.432

A Principles of Education
B Psychology and Hygiene
C Special Method
D Educational Psychology or History of Education or Comparative Education
E Educational System
F Essay
G Practical Teaching

TABLE IV

Table to show Correlations of Marks in Examination II

	Q	R	S	Q	R	S
		1931			1932	
P	.691	.583	.531	.762	.609	.584
Q		.694	.704		.725	.533
R			.584			.355
		1933			1934	
P	.554	.593	.591	.592	.654	.364
Q		.849	.482		.881	.487
R			.515			.482
		1935			1936	
P	.495	.468	.307	.509	.754	.610
Q		.750	.293		.832	.534
R			.423			.609
		1937			1938	
P	.500	.449	.235	.690	.710	.438
Q		.768	.499		.790	.567
R			.545			.357

P organic science
Q inorganic science
R mathematical science
S social science

(b) Correlation Coefficients

These are all positive, and though those of Examination I are mostly small, only three of them are not statistically significant.

(c) Saturation Coefficients

Vernon[7], after analysing his correlation table, stated "in the educational field at least, our data show that there is positive overlapping of widely diverse abilities among training college students." Now Burt states "the relation between the results of the group-factor method and those of the general-factor method can be expressed by a single transformation matrix,"[8] and he has applied this procedure to the factors Vernon obtained. He has reduced the original table of correlations to one positive general factor and three positive group factors showing little or no significant overlap. Burt calls the general factor 'general collegiate ability as judged by college marks and tests'. Table V, page 24, shows that the general factor for Examination I correlates most highly with Principles of Education, Psychology and Hygiene, Special Method, Educational Psychology and Educational System. Burt, in his analysis, finds it correlates most highly with Education, Psychology and English.

---

[7]Vernon, P.E. <u>Loc. cit.</u>, p. 8.

[8]Burt, C. <u>Loc. cit.</u>, p. 6.

TABLE V

Table to show Saturation Coefficients Obtained by Analysis of Correlations of Examination I in Table III

1st Saturation Coefficient							
Year	A	B	C	D	E	F	G
1931	.6972	.6155	.5173	.5543	.7013	.2730	.3024
1932	.5536	.6453	.7405	.7441	.6650	.4214	.4973
1933	.5063	.6818	.6793	.5747	.5888	.5478	.5239
1934	.5070	.5150	.5888	.6702	.7690	.5129	.5256
1935	.6721	.5390	.5787	.5763	.4290	.3107	.4162
1936	.6483	.5644	.6128	.5615	.4608	.5463	.4269
1937	.5284	.5596	.6650	.5683	.7905	.4542	.5483
1938	.5943	.6553	.6349	.6407	.6067	.4330	.5073
Mean	.5884	.5939	.6272	.6113	.6264	.4374	.4685
2nd Saturation Coefficient							
Year	A	B	C	D	E	F	G
1931	-.1846	.0426	.2386	-.1592	-.2151	.4370	.3482
1932	-.2807	-.3168	.0541	.1782	-.0878	.2337	.2818
1933	-.6447	-.0616	.0666	-.2027	-.2187	.2607	.3512
1934	-.1339	-.3128	-.2802	.0505	.0916	.2736	.2844
1935	-.2280	-.3020	.0389	-.1597	-.1818	.5913	.6687
1936	-.1780	-.2631	.1623	-.1885	-.1479	.3573	.3079
1937	-.1849	-.2950	-.0200	-.1314	-.0360	.2980	.3806
1938	-.0163	-.2326	-.0211	-.1521	-.1150	.3836	.3494
Mean	-.2324	-.2175	.0299	-.1093	-.1126	.3544	.3716

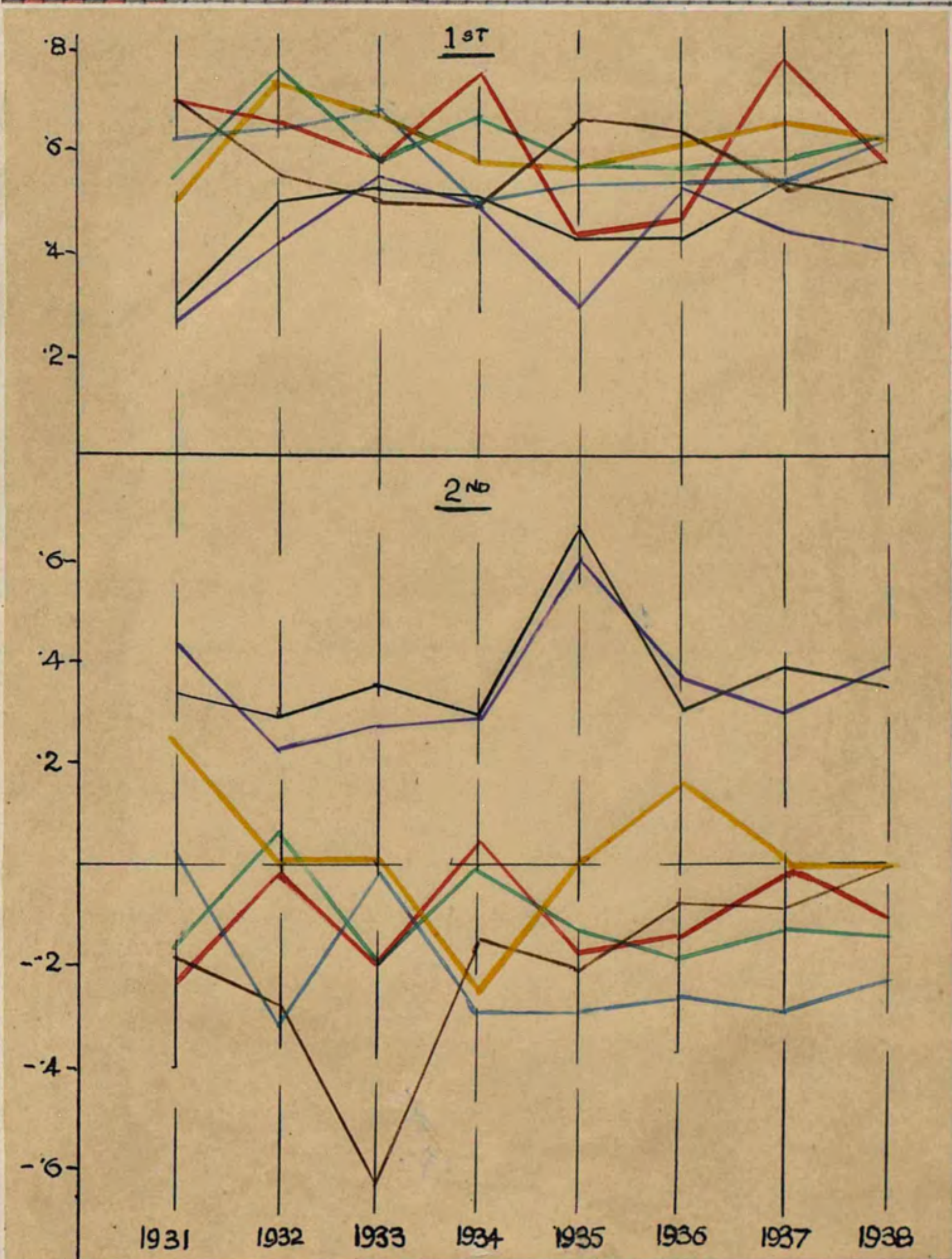
A Principles of Education
B Psychology and Hygiene
C Special Method
D Educational Psychology or History of Education or Comparative Education
E Educational System
F Essay
G Practical Teaching

The second factor divides the seven subjects into two groups, namely those dealing with the theory of education, Principles of Education, Psychology and Hygiene, Educational Psychology and Educational System, as contrasted with the paper on Special Method and the students' own essays which are closely related to the particular subjects they are teaching. The simplest possible explanation is that this second factor is a bipolar factor contrasting theory with practice or one which contrasts general theory with the knowledge of one specific subject. On the other hand, a personal factor may be demonstrated here. It is possible that the marks for Special Method, Essay and Teaching were given individually while no examiner would stop to think which candidate's paper he was marking in the other subjects. In Practical Teaching, the identity of the candidate is obvious. A tutor who supervises the teaching practice of a student usually marks his essay, lectures to him in Special Method and also marks his examination paper in that subject. Hence, there is the possibility of an examiner-candidate relationship which is obscured as far as the other subjects are concerned.

Figure IV, page 26, illustrates the factor pattern. It shows the constancy of this pattern from year to year. The complete isolation of Essay and Teaching from the other

FIGURE IV

Graph to show the Values of the Saturation Coefficients Obtained by Analysing the Marks of Examination I



Principles of Education ——
Special Method ——
Essay ——
Psychology and Hygiene ——
Educational System ——
Practical Teaching ——
Educational Psychology or History of Education or Comparative Education ——

subjects in the second factor is remarkable. Special Method tends to follow a parallel course at a lower level except in 1934 and 1936. There seems to be no apparent reason why the curves for Educational Psychology and Educational System are nearly co-incident in their course. From 1934 onwards, the lecturer (and examiner) in Educational Psychology took over the Principles of Education course. It is noticeable in the second factor that the graphs for these two subjects do tend to converge during the last five years. This may be due to the influence of the teacher. Burt refers to this in the memorandum in which he has analysed Vernon's correlation table. He states that "marked linkages were introduced" under similar circumstances.

In Examination II, (see Table VI, page 28 and Figure V, page 29), the general factor has higher values for each subject than in Examination I. This is probably due to the fact that these candidates are, on an average, three years younger than those of Examination I. Likewise, the standard of the papers of Examination II is lower than that of Examination I. The greater values of the general factor in Examination II is in accordance with Burt's theory that it decreases as the examinees become older. In the second factor, the mathematical science, R, and the social scinece, S, show constancy in the course they follow from year to year.

TABLE VI

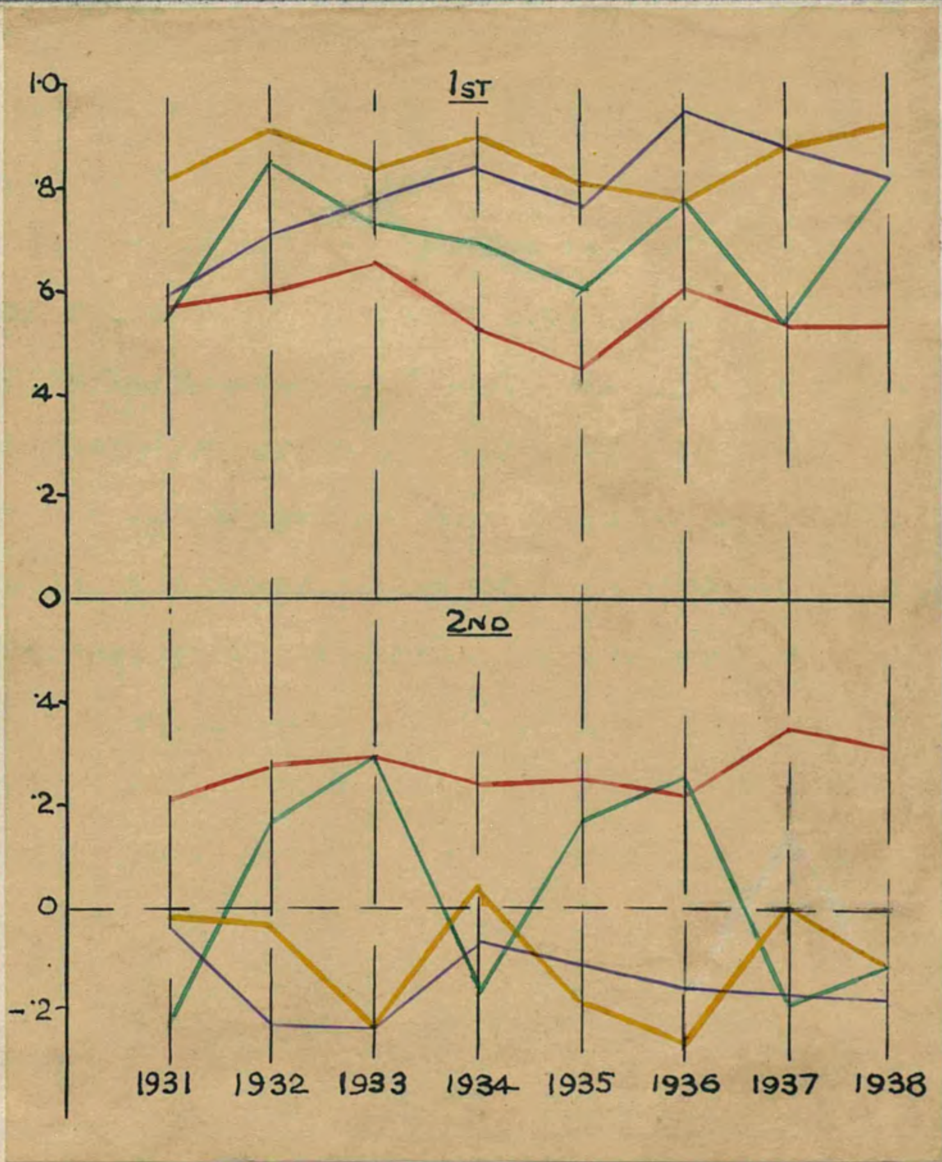
Table to show Saturation Coefficients Obtained by Analysis of Correlations of Examination II in Table IV

1st Saturation Coefficients				
Year	P	Q	R	S
1931	.5699	.8426	.6139	.5820
1932	.8688	.9095	.7325	.6087
1933	.7402	.8408	.8795	.6608
1934	.6971	.9025	.9359	.5448
1935	.6118	.8118	.8704	.4635
1936	.7790	.7878	.9561	.7118
1937	.5315	.8813	.8798	.5374
1938	.8093	.9251	.8326	.5524
Mean	.7010	.8627	.8376	.5827
2nd Saturation Coefficients				
Year	P	Q	R	S
1931	-.2104	..0199	-.0307	.2156
1932	..1656	-.1338	-.2219	.2924
1933	.2756	-.2424	-.2234	.2961
1934	-.1597	.0296	-.0515	.2519
1935	.1827	-.1839	-.0961	.2625
1936	.2425	-.2589	-.1557	.2300
1937	-.1801	.0114	-.1611	.3498
1938	-.1295	.1207	-.2120	.3074
Mean	.0233	-.0797	-.1441	.2745

P organic science
Q inorganic science
R mathematical science
S social science

FIGURE V

Graph to show the Values of the Saturation Coefficients Obtained by Analysing the Marks of Examination II



P organic science ——   Q inorganic science ——
R mathematical science ——   S social science ——

The mathematical science is the stumbling block of the examination. The lecturer shows great consideration to her students and gives them ample opportunity for the explanation of points which are not understood, and, as an examiner she strives to obtain accuracy in marking her papers. It is understood that the social science is marked by impression while the other subjects are marked by points for divisions and sub-divisions of each question in the papers. It is quite reasonable to accept this as an explanation of the fact that the curves of both the saturation coefficients of subject S deviate from those of the other subjects. This explanation makes the examiner's foibles the cause of the difference of the saturation coefficient curve. S is a social subject without a mathematical trend. In the years 1932, 1933, 1935 and 1936, there is a link between it and the organic science, P, which is the least mathematical and the least abstract of the four subjects next after S. All the time, it contrasts markedly with the subjects Q and R, the inorganic and mathematical sciences. Q and R, however, do tend to show a mutual linkage in six out of the eight years.

It was decided to test the stability of the factors, and the values for the subjects were correlated between the years. These are given in Tables VII and VIII, pages 31

TABLE VII

Table to show Correlation Coefficients between the Saturation Coefficients from Year to Year

EXAMINATION I

First Factor

	1932	1933	1934	1935	1936	1937	1938
1931	.6111	.2421	.4411	.6875	.3727	.4104	.8004
1932		.6338	.5957	.5880	.1465	.6298	.9037
1933			.4374	.1681	.2057	.3825	.6000
1934				-.0507	-.3422	.8285	.3811
1935					.7026	.0959	.7666
1936						-.2798	.3582
1937							.5239

Second Factor

	1932	1933	1934	1935	1936	1937	1938
1931	.6024	.8814	.3447	.8606	.9031	.7581	.7899
1932		.7170	.7452	.8047	.7666	.8058	.6944
1933			.4581	.7830	.7953	.7616	.6428
1934				.7472	.5823	.8237	.7481
1935					.9492	.9749	.9641
1936						.9261	.9201
1937							.9421

TABLE VIII

Table to show Correlation Coefficients between the Saturation Coefficients from Year to Year

EXAMINATION II

First Factor

	1932	1933	1934	1935	1936	1937	1938
1931	.7606	.5165	.5821	.5437	.2253	.6953	.6390
1932		.4936	.5498	.5097	.0998	.3887	.8118
1933			.7963	.9994	.8462	.9267	.8453
1934				.9984	.6184	.9336	.8764
1935					.8287	.9331	.8536
1936						.6974	.5201
1937							.7087

Second Factor

	1932	1933	1934	1935	1936	1937	1938
1931	.3073	.3356	.9837	.1647	-.0126	.9227	.8223
1932		.9763	.4235	.9486	.9247	.1604	.6488
1933			.2107	.9821	.9895	.0297	.3316
1934				.2984	.1365	.8744	.8947
1935					.9852	.1872	.3375
1936						.0212	.2154
1937							.5632

and 32. In Examination II, the general factor (except for 1932/1936) appears to be much more stable than in Examination I. On the other hand, in Examination I, the stability of the second factor seems to increase steadily while that of Examination II fluctuates all the time. It is possible that this trend towards a greater degree of stability is due to the tardy acceptance by some of the examiners of the restricted method of marking. In estimating the significance of these results it must be remembered that "a factor is primarily a principle of classification and nothing more: it is expressed in quatitiative form simply because the items, whose characteristic pattern constitutes the distinguishing mark of the class described by it, vary continuously and in degree rather than discontinuously and in kind."[9]

---

[9] Burt, C. The Factorial Analysis of Ability, III, B.J.P., XXX, Oct. 1939. pp. 84 - 93.

CHAPTER III

ANALYSIS OF VARIANCE BETWEEN TWO ESTIMATES OF MARKS FOR A UNIVERSITY EXAMINATION

CHAPTER III

ANALYSIS OF VARIANCE BETWEEN TWO ESTIMATES OF MARKS FOR A UNIVERSITY EXAMINATION

Material

The marks for this section of the research were obtained in the same way and under the same conditions as those discussed in the last chapter. Briefly, they are the marks of one paper of a university examination taken by men and women at the end of a two-year course. To proceed to the course, the students must have matriculated. The examination includes three sets of marks, namely: college assessment, marks awarded at a written examination and those of the revising examiner. The amount of difference between these sets of marks has been determined by analysing the variance. It was found to be more convenient to divide up the marks as follows:-

(i) Set A - men candidates from the colleges I, II, III, IV and V, examiners X, Y and Z.

(ii) Set B - women candidates from the colleges VI, VII, VIII, IX, X, XI, XII, XIII and XIV, examiners S, T, U, V and W.

Statistical Methods

The method adopted is that of Fisher[10] and its application to this work will be explained in a later section. Tests and examinations are assumed to measure some capacity of the individual tested. A test is considered as valid if it measures the capacity for which it was designed. In this research the aim is to examine the accuracy of measurement of the ability of various individuals by an examination and by another separate score. The individuals selected are the examination candidates already mentioned. Two sets of scores are complete; the third is incomplete. The first set is called the college assessment and is understood to be the mark given by the lecturer at the college attended by the candidate in question. No information is available as to the method of making the assessment. The examination marks form the second set and those are independent of the college marks. The papers of the candidates at any one college have been marked by lecturers of other colleges. The third set of marks are those of the revising examiner. An examiner who is not one of the aforementioned lecturers marks the border-line cases and some of the best and worst papers picked at random, which explains why this set of marks is not complete or independent. The examiners

---

[10] Fisher, R.A. Statistical Methods for Research Workers, Oliver & Boyd, London, 1932. p. 216, seq.

marking the written papers often have in front of them the assessment laid down by the college. By inquiry from individual examiners, it would appear that they are very seldom influenced by those marks, since their attention is almost wholly concentrated on the scripts. On the other hand, the revising examiner might, on finding cases in which college assessments differed from the marks awarded at the written examination, give a mark which would not always be independent in the sense required by a statistician or a psychologist.

Results

Calculations are made on the marks according to the following groupings:

(a) As a whole

(b) Men and women separately

(c) Colleges separately

(d) Examiners separately

(e) Per examiner per college

The calculations made for each grouping of marks are the distribution, average mark and standard deviation. The correlation between the college assessments and the marks awarded at the written examination are also found.

(a) and (b) The instructions given to the examiners marking the written papers state the way in which the marks should tend to be distributed. This distribution and the actual distributions are shown in Figure VI, page 38. The presented distribution has an average mark of 57.5 and a standard deviation of 11.15. The adjacent diagrams show the distributions of (a) the marks for all the candidates taken as one group and (b) the marks for men and women grouped separately. These distributions seem to approximate to that prescribed in the instructions. Table IX gives the calculations made on the marks when grouped as (a) and (b).

TABLE IX

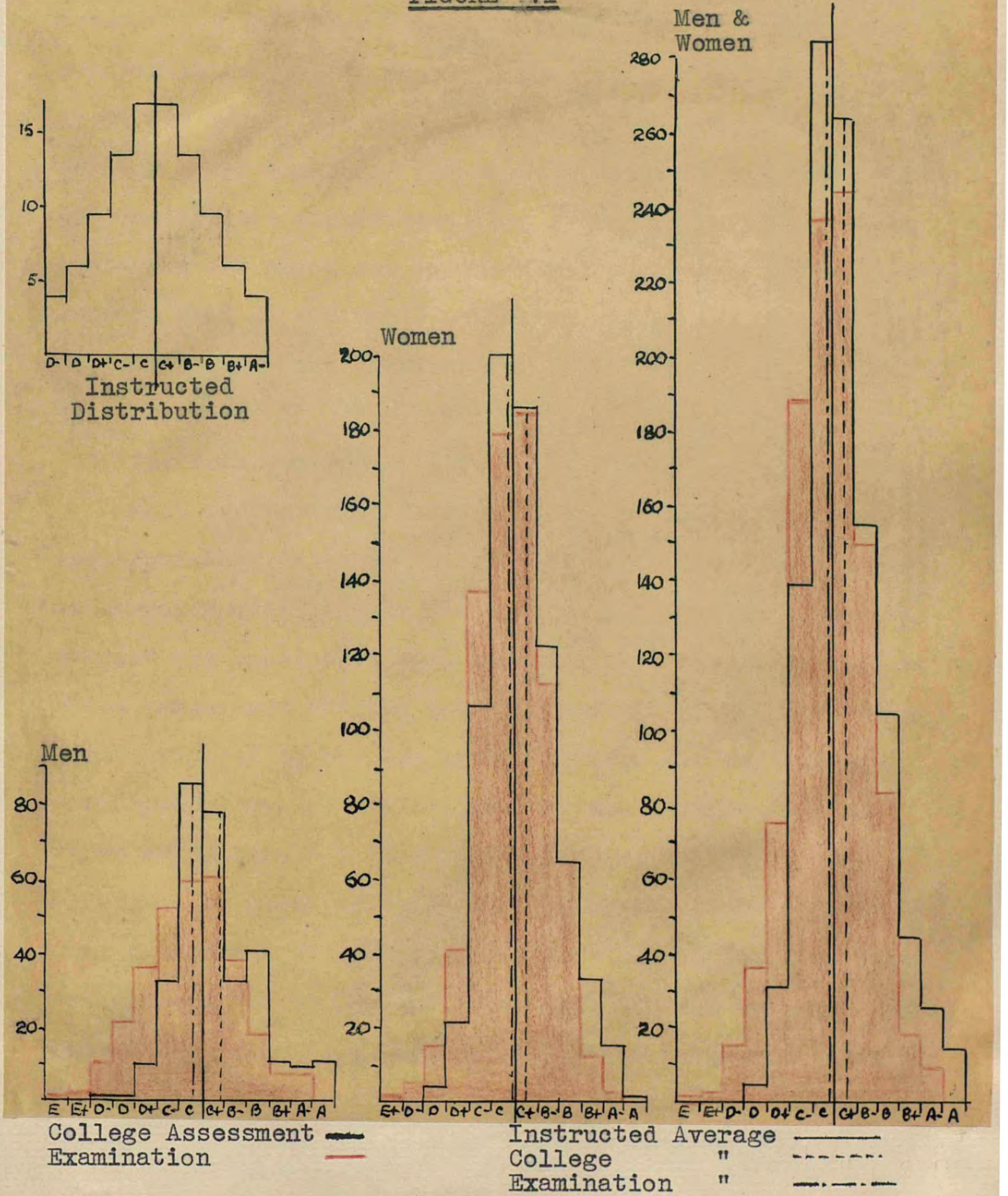
Table to show Calculations on Groups (a) and (b)

		Men & Women	Men only	Women only
Average Mark	Coll. Asst. Examination	59.94 56.77	61.68 55.05	59.57 57.57
Standard Deviation	Coll. Asst. Examination	8.25 8.65	8.70 10.25	7.75 7.75
Correlation between college assessment & examination mark		.497	.623	.431

The average mark of the whole group of candidates for both the college assessment and that of the marks awarded at the written examination is near enough to the specified mean mark (57.5). The mean marks, however, of the men

Graph to show Distributions of Marks of Groups (a) & (b)

FIGURE VI



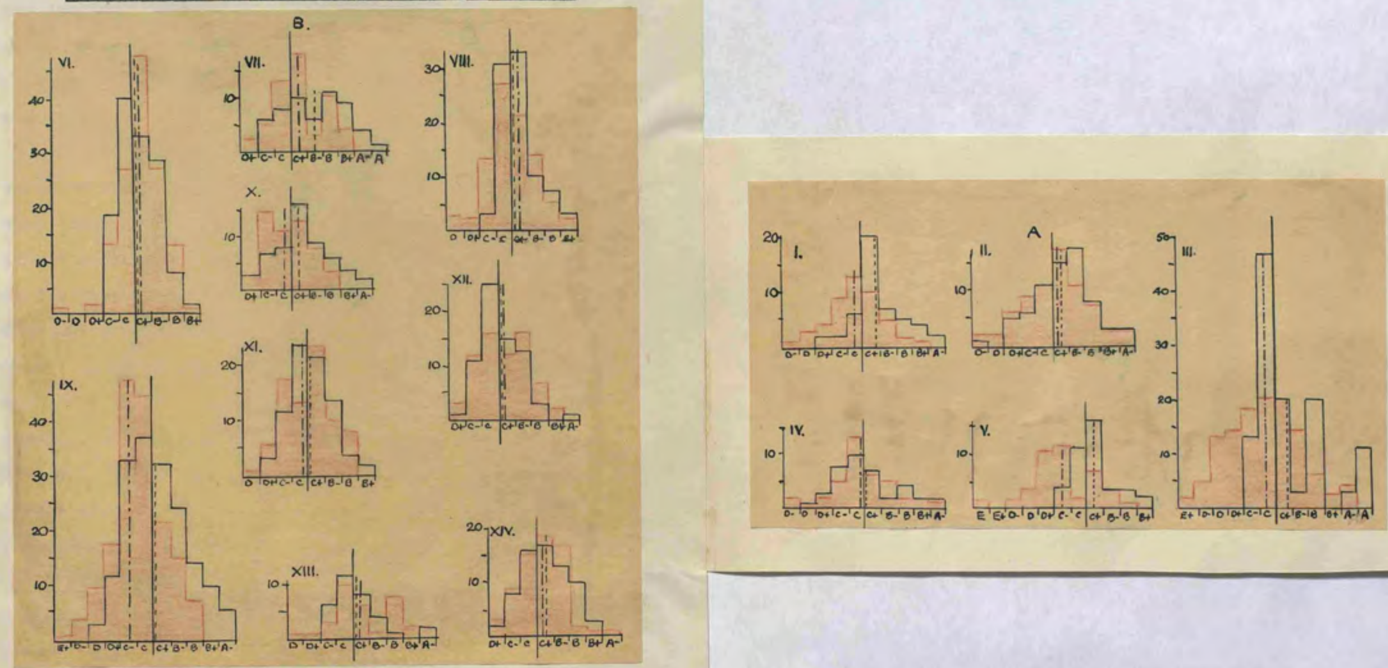
lecturers and examiners diverge more than do those of the women. On the other hand, the standard deviation of the men examiners (10.25) is the only one that approaches the specified value (11.15). The correlations between the college assessments and the marks awarded at the written examination are significant; but it appears that the women lecturers and examiners agree much less closely in their marking than do the men. This sex difference will be discussed when the calculations for individual examiners are considered.

(c) Colleges separately

Most of the distributions given in Figure VII, page 40, are consistent with the notion that the underlying ability is normally distributed. At the same time, the marks awarded are consistent with the view that they belong to a slice taken from the top end of a normal curve of distribution. In Set A, the two irregularities are the college assessment curve of College III and the examiner's mark curve of College V both of which have more than one peak. It should be noted that contrary to instructions, the lecturer puts the peak over B- in College II and the examiner for College V puts the peak over C-. In the college assessment marks of College III, 11 candidates have been given A, 20 the mark B while 47 have C, a total of 78 candidates out

FIGURE VII

Graph to show Distribution of Marks per College



College Assessment	——	Instructed Average	————
Examiner's Mark	—— (red)	College "	- - - - - -
		Examination "	—·—·—

of 117. Here there is a marked disagreement between the college assessment and the examiner's marks. Evidently the lecturer tends to give marks at the whole letters instead of marks involving plus and minus. This will be discussed in greater detail when individual examiners are considered.

In Set B, most of the distribution curves seem to converge to the normal curve; Colleges VII and XIII are somewhat irregular in their college assessments while the examiner's mark curves for Colleges VII, X, XI, XII and XIII each have two very definite peaks. The reason for this does not seem obvious except in the case of College XIII where the marks are peaked over B and C, the examiner having hesitated to use the plus and minus in the literal marking scheme. The candidates in Colleges IX and X are found to be below the general standard by their examiners as the curves show a large proportion of their marks to be below the specified average of 57.5.

One other interesting point arises out of a consideration of the college distributions, namely, the way in which they affect the award of distinctions at one end of the scale and failures at the other end. College III is a very remarkable instance of the phenomenon. The lecturer gives a mean mark of 61.45; he has 11 candidtaes, i.e. 9.4% marked A and no mark lower than C-. At the same time 50% of his students are awarded more than the average mark, ranging from A to C+ while of the other 50% below the average, 40% are given C and the remainder have C-. The examiners give a distribution of marks ranging from A- to E+ and the shape of this curve is very similar to that of the instructed distribution. The lecturer at College V also gives his students a higher range of marks than the examiners do. This lecturer, however, does not claim any outstandingly high marks for his students though his average mark is 8.85% higher than that of the examiner. Colleges I, VIII, IX and X also show a tendency to be more highly rated by their lecturers than by their examiners.

A remarkable sex difference is brought out in these distributions. The men lecturers in set A seem to spread out their candidates much more boldly than the women lecturers do. This fact is supplemented by the low standard deviation of College VIII, 5.55, and of College XII, 5.95. In College VI, 129 marks out of a total of 132 are chosen from the five grades between B and C- whereas in theory ten different grades could be used. The fact seems to be that the untrained person marking by impression can distinguish between those candidates who are just above average and those who are just below average; but he or she seldom stops to distinguish between candidates of outstanding merit or de-merit towards either of the extremes. This is often said to be characteristic of women examiners. They seem afraid to commit themselves to a very high mark or to a very low mark - partly perhaps, they fear criticism from the revising examiner, but more probably because they feel doubtful about deciding the fate of a candidate unless they are very sure of their ground. This is more noticeable towards the failure end.

TABLE X

Table to show Means and Standard Deviations of Marks for the Colleges

College	College Assessment		Examiner's Mark	
	Mean Mark	Standard Deviation	Mean Mark	Standard Deviation
I	62.19	7.80	55.00	8.25
II	60.85	9.02	58.72	9.65
III	61.45	10.25	54.49	11.10
IV	58.08	9.70	56.92	9.75
V	59.49	6.35	50.64	8.90
VI	58.87	6.10	59.70	6.30
VII	65.09	9.40	60.54	7.75
VIII	59.77	5.55	57.65	7.15
IX	58.68	9.20	58.27	7.80
X	60.82	8.75	56.72	6.80
XI	58.87	6.45	57.16	7.20
XII	58.04	5.95	59.35	7.80
XIII	58.59	7.50	60.00	8.75
XIV	60.29	7.25	59.49	7.25
Average	60.08	7.81	57.47	8.16

Table X gives the mean marks and standard deviations of the college assessments and the written examination marks. Several interesting points may be noted here. The college lecturers differ a good deal amongst themselves about the level they assign to their college, a range of 58.04 to 65.09 and every mean mark is above the instructed average of 57.5. The examiners have a still wider range of average marks, i.e. 50.64 to 60.54 but the mean of this range is 57.47, almost the value of the specified average. College VII has the highest mark in the written examination and also the highest college assessment. Colleges I, III and V, however, which have been given the three lowest averages by the examiners, have been awarded some of the highest averages in the college assessments. The lack of agreement between the two sets of average marks is adequately proved by the negligible value of the correlation between them, namely, .016.

A very significant feature may be noted in the standard deviations. Although each value taken individually gives a very narrow range, as a set, they vary considerably, 6.35 to 10.25 in the college assessments and 8.25 to 11.10 in the written examination in set A while set B has a range of 5.55 to 9.40 and 6.30 to 8.75 for the college assessment and the written examination respectively. College III is nearly twice as heterogeneous as College VIII in college assessment though it is true that the written marks tend to agree in the former case. College VII has a standard deviation for lecturer's marks which is almost twice as great as that for College XII but the written marks give the opposite verdict.

The correlations between the college assessment and the marks awarded at the written examination are given in Table XI, below. They are all positive and statistically significant, even those which have a marked difference in the means of the variables correlated.

TABLE XI

Table to show Correlations between College Assessments and Examiner's Marks for Colleges

College	No. in College	Correlation
I	48	.693
II	70	.748
III	117	.501
IV	39	.711
V	39	.565
Mean	313	.644
VI	132	.534
VII	55	.634
VIII	87	.358
IX	171	.307
X	55	.491
XI	81	.406
XII	69	.504
XIII	32	.583
XIV	59	.563
Mean	751	.487
Mean of Total	1064	.543

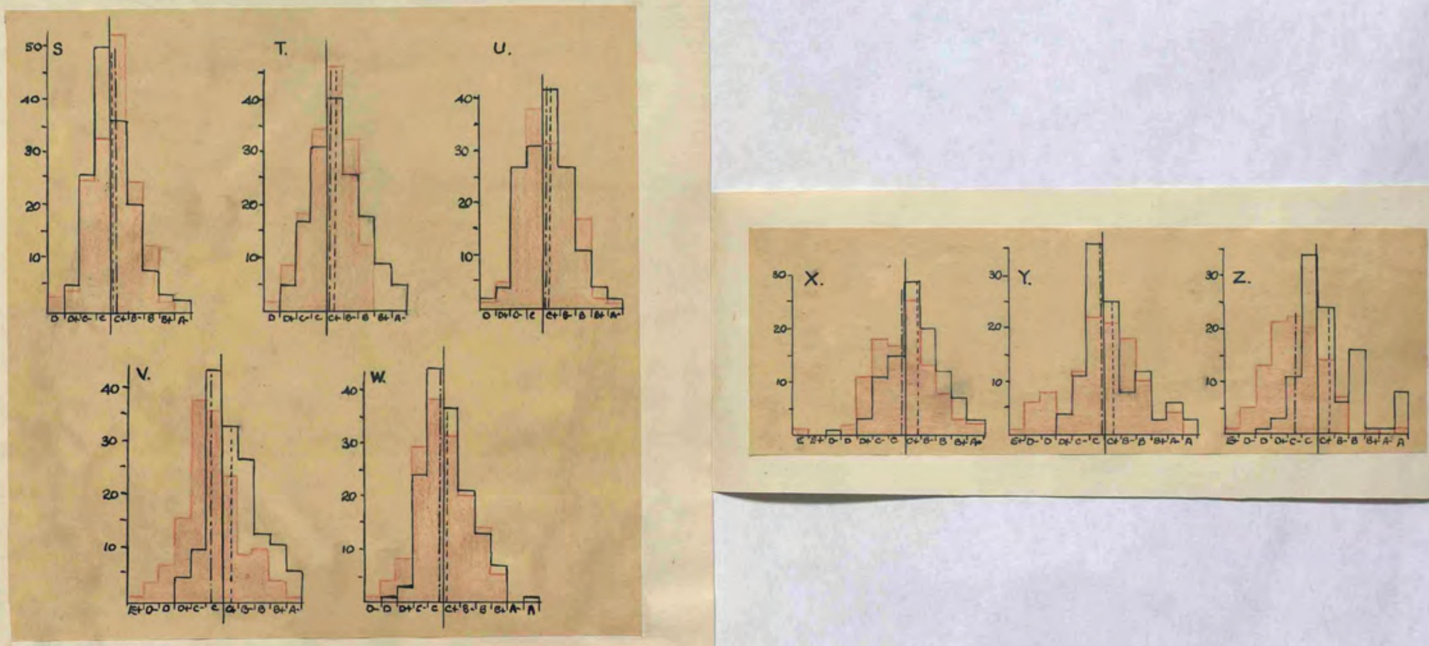
A sex difference is again apparent. Four of the women's colleges have correlations below .4, whereas in two of the men's colleges the correlations are above .7. This point will be discussed later when the colleges are divided up for each examiner. Some of the colleges are affected by the heterogeneity of the groups which is reflected in the standard deviations. The more homegeneous colleges should have the lower correlations. Hence, the average of the lecturer's and examiner's standard deviations was correlated with the corresponding correlation. The values obtained are .494 for men, .360 for women and .473 for the whole group of candidates which might indicate that the women form a more homogeneous group than the men.

(d) Examiners separately

This section deals with the pooled marks of each examiner and the relationship of those marks to the college assessment given. The distribution curves are given in Figure VIII, page 47 where it will be seen that the curve of Examiner Z is the only one that disagrees markedly with that of the college assessment. Correspondingly, this is the only case with a large difference between the two mean marks, viz., 60.74 for the college assessment and 50.71 for the examiner's mark. Examiner V has an examination mean mark of 54.37 while the college assessment is 60.96.

FIGURE VIII

Graph to show Distributions of Marks for Examiners



College Assessment
Examiner's Mark

Instructed Average
College "
Examination "

TABLE XII

Table to show Means and Standard Deviations of Marks for Examiners

Examiner	College Assessment		Examiner's Mark	
	Mean Mark	Standard Deviation	Mean Mark	Standard Deviation
X	61.25	8.41	57.25	9.26
Y	60.46	9.24	57.27	10.97
Z	60.74	9.92	50.71	9.06
Average	60.82	9.19	55.08	9.76
S	57.87	6.86	58.47	6.43
T	60.63	7.94	58.61	6.56
U	58.80	7.48	58.66	7.39
V	60.96	8.18	54.37	9.01
W	58.66	8.12	57.40	6.80
Average	59.38	7.72	57.10	7.24
Average of Total	59.92	8.27	56.59	8.18

The standard deviations are reasonably close in every case but this is probably an artificial effect due to pooling the different colleges.

The correlations between the college assessment and the examiner's mark are given in Table XIII on page 49. Owing to the unusual results obtained from the marks of the candidates at College VIII, the correlations per examiner were calculated both to include and to exclude this college. In the case of the last five examiners, the second set of values in each case are those calculated to exclude

College VIII. It will be noted that these values are higher in each case. All these correlations are statistically significant.

TABLE XIII

Table to show Correlations between College Assessments and Examiner's Mark per Examiner

Examiner	No. in Group	Correlation
X	100	.677
Y	108	.745
Z	105	.447
S	150	.266
	133	.332
T	150	.609
	133	.682
U	149	.241
	132	.244
V	150	.295
	133	.322
W	150	.711
	132	.771

The differences in the values of the correlations are amazing, and it would seem not unfair to say that only the examiners X, Y, T and W are efficient. It is possible that they allowed themselves to be guided by the college assessments which were in front of them while they marked the scripts. This is probably not true of Examiners T and W for in the case of the correlations including College VIII they do not seem to have been influenced by that college assessment nearly as much as Examiners U and V were influenced.

(e) Per Examiner per College

Some interesting facts which are hidden by the averaging involved in Tables X and XI (pages 43 and 45) are revealed in the distribution curves separated out for each examiner in each college. These are given in Figure IX, page 52. A mere glance at the curves shows that many of the examiners do not agree with the college assessments given, most of them find these to be too high. Striking examples of this fact are seen in the curves listed in the following table:-

TABLE XIV

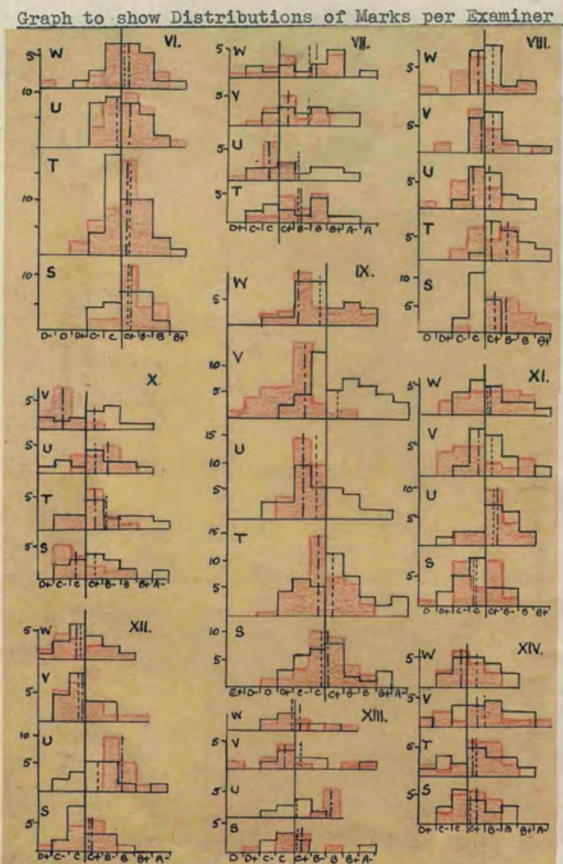
Table to give Examples of Examiner's Disagreement with College Assessment

Examiner	College
X	V
Y	I, III
Z	I, III, V
S	X
T	VIII
U	VII, IX
V	VII, IX, X, XI

On the other hand, Examiner S finds the candidates of College VIII have been marked too low in their college assessment and Examiner U thinks the same for the candidates of College XII, XIII and XIV. It was pointed out on page 49 that only Examiners X, Y, T and W are efficient. Table XIV endorses this statement; Examiner W does not appear in it and Examiner T only disagrees with the college assessment of College VIII which has been criticised previously for its apparently unreliable marking. In Table XV, page 53, the mean mark for the college assessment in the men's colleges is above the instructed average (57.5) and it is only in College IV that the average falls below the specified average for those candidates examined by Y and Z. It will be noted that while Y raises the average of the candidates he examines by 3.85%, Z lowers the average of his candidates by 3.45%. The mean for the examination averages is below the specified value. In the case of the women's colleges, the average of both those of the college assessment and those of the examination are above the instructed value. There are many examples of wide discrepancies between these two averages. In particular, Examiner V gives Colleges VII, IX, X and XIII averages less than those of the college assessment by 6.43%, 13.25%, 9.28% and 5.56%, respectively. Examiner U is more erratic still. She lowers the averages of Colleges VIII

FIGURE IX

Graph to show Distributions of Marks per Examiner per College



Averages
Instructed
College
Examination

College Assessment —— Examiner's Mark ——

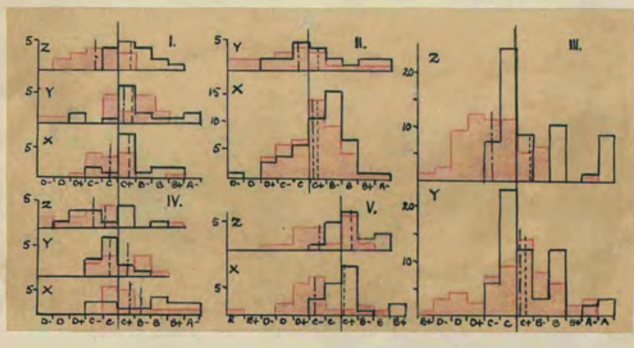


TABLE XV

Table to show Means and Standard Deviations of Marks per Examiner per College

College	Examiner	College Assessment		Examiner's Mark	
		Mean Mark	Standard Deviation	Mean Mark	Standard Deviation
VI	S	59.60	9.58	61.60	6.30
	T	59.24	5.85	58.91	6.20
	U	57.42	5.65	59.85	5.80
	W	58.93	6.45	59.11	7.20
VII	T	63.21	9.15	63.21	6.10
	U	63.85	9.40	59.62	4.15
	V	67.14	8.20	60.71	5.90
	W	67.14	9.95	68.21	9.85
VIII	S	60.88	3.10	63.82	5.55
	T	64.44	5.50	58.89	5.15
	U	58.53	4.95	53.82	5.00
	V	60.29	5.55	57.06	6.65
	W	59.44	4.65	54.94	6.30
IX	S	57.50	8.35	57.50	6.55
	T	60.05	9.60	55.98	6.90
	U	55.17	8.35	51.03	3.15
	V	61.97	9.45	48.72	8.95
	W	56.25	7.60	54.11	8.20
X	S	60.71	7.75	54.64	6.10
	T	62.86	9.20	59.29	5.36
	U	60.38	8.45	63.31	5.00
	V	59.64	8.50	50.36	3.95
XI	S	55.00	7.60	54.76	5.85
	U	59.75	6.00	60.50	5.20
	V	59.00	5.35	55.00	8.20
	W	58.10	6.95	58.50	7.75
XII	S	57.94	5.40	57.35	5.45
	U	61.18	7.15	67.95	4.20
	V	56.39	4.90	56.94	6.85
	W	56.76	6.15	55.29	6.95
XIII	S	58.13	8.95	58.25	6.80
	U	57.86	4.50	69.29	1.75
	V	60.56	8.15	55.00	8.65
	W	57.50	6.60	60.00	7.90
XIV	S	58.61	6.40	58.89	4.85
	T	60.28	8.05	58.89	6.80
	V	62.67	7.90	60.60	9.50
	W	60.28	5.90	56.11	3.90
Average		59.86	7.04	58.37	6.18

College	Examiner	College Assessment		Examiner's Mark	
		Mean Mark	Standard Deviation	Mean Mark	Standard Deviation
I	X	61.86	7.25	55.10	6.55
	Y	62.50	9.95	58.44	9.45
	Z	62.19	5.85	50.94	6.40
II	X	60.90	7.95	59.80	7.90
	Y	60.75	10.40	54.00	12.05
III	Y	60.93	9.05	57.63	11.10
	Z	62.24	11.50	50.43	9.35
IV	X	65.77	10.35	61.92	7.90
	Y	55.38	4.44	59.23	6.15
	Z	53.07	8.20	49.62	10.05
V	X	58.81	6.50	49.52	9.60
	Z	60.28	5.65	51.94	8.15
Average		60.39	8.09	54.88	8.72

and IX by 4.71% and 4.14% but raises those of Colleges XII and XIII by 6.77% and 11.43%, respectively. The correlations between the averages of the college assessment and those of the examination marks are .315 for men and .372 for women. The former is not significant and the latter barely significant for P = .02. The very low values of these correlations prove that the examiners do not agree with the ranges of marks given to the candidates by their lecturers.

Again the standard deviations vary considerably. Of the men, only Examiner Y tends to approach the specified value except in the case of College IV (6.15) but in this case the standard deviation for the college assessment also is very low (4.44). Among the women, College VIII has low standard deviations in both cases, though those of the examination are slightly higher. Examiner U has produced exceedingly low standard deviations for her examination scores; the most remarkable are 4.15 for College VII, 3.15 for College IX, 4.20 for College XII and 1.75 for College XIII. Examiners V and W each have one low value, i.e., 3.95 for College X by V and 3.90 for College XIV by W.

The correlations for this section are given in Table XVI, page 55. In Set A, only 3 correlations out of the 12 are not significant; otherwise the values are satisfactory. In Set B, 21 out of a total of 38 correlations are not significant. Examiner S has 7 out of 8 not significant but

TABLE XVI

Table to show Correlations between College Assessments and Examiner's Marks per Examiner per College

College	Examiner	No. in Group	Correlation
VI	S	25	.449
	T	46	.609
	U	33	.203
	W	28	.768
VII	T	14	.837
	U	13	.342
	V	14	.263
	W	14	.823
VIII	S	17	.146
	T	18	-.119
	U	17	.482
	V	17	-.086
	W	18	.517
IX	S	30	.191
	T	41	.658
	U	29	.125
	V	43	.211
	W	28	.735
X	S	14	.335
	T	14	.884
	U	13	.753
	V	14	.052
XI	S	21	.215
	U	20	-.282
	V	20	.429
	W	20	.726
XII	S	17	.407
	U	17	.032
	V	18	.749
	W	17	.747
XIII	S	8	.849
	U	7	.260
	V	9	.908
	W	8	.839
XIV	S	18	.489
	T	18	.740
	V	15	.776
	W	18	.726

not significant

College	Examiner	No. in Group	Correlation
I	X	16	.695
	Y	16	.920
	Z	16	.447
II	X	50	.649
	Y	20	.854
III	Y	59	.686
	Z	58	.427
IV	X	13	.808
	Y	13	.198
	Z	13	.785
V	X	21	.678
	Z	18	.552

not significant

all are positive. Examiners T, U and V each have one negative correlation. College VIII has all its correlations not significant and two of these are negative. The correlation of -.282 for Examiner U in College XI is twice its probable error and therefore surprisingly high, although not technically significant.

The next section deals with the analysis of variance of the marks. The method used is that of Fisher[11] to which reference has already been made. The fundamental assumption underlying the analysis of college assessment and the marks awarded at the written examination is that these are comparable measures of the same ability or capacity of the individual student. These measures are supposed to be more or less valid estimates of the individual's capacity in (or knowledge of) the subject in question. An assessment or an examination should differentiate between the candidates and show them to be s significantly heterogeneous group. At the same time, the men and women who mark the candidates should form a homogeneous group and the sets of marks given by them should not differentiate significantly between the examiners.

The score of an individual in a particular application of a test such as the college assessment and the written examination, may be expressed as the sum of factors or

---

[11] Fisher, R.A. Loc. cit., p. 36.

components. Let y_{ut} be the score of the t-th individual in the u-th application of a test. Then $y_{ut} = A + B_u + C_t + D_{ut}$, where $u = 1, 2, t = 1, 2, \ldots n$. A is a measure common to all individuals irrespective of the trial under consideration and is regarded as a constant, defined as the arithmetic mean of the effects for all trials: B_u is a measure of the trial effect and is also a constant; C_t is a measure of the individual effect, i.e., a measure of some capacity of the individual measured by the test; D_{ut} is the error of measuring by means of the test applied. Now C_t is the variance for examiners (V_1) and if this value is zero or negligible then it may be concluded that there is no difference in the average mark assigned by the lecturer and by the examiner, i.e., they mark with the same degree of severity. If the results indicate that the variance for examiners is not zero it may be concluded that there is a significant difference between the average marks assigned by the lecturer and examiner. This factor may not seem to be important but actually it may make some difference to the individual student concerned, for instnace, between pass and failure or in obtaining a credit. If an examiner's mean mark and standard deviation are lower than that of the lecturer, then a cnadidate assessed with, say A-, by the lecturer, may have his mark reduced so that he now only has

B or even B-. Likewise, a candidate who has been given, say C- by the lecturer, may be reduced to D or D- if the examiner's standard deviation is greater than and his mean mark less than those of the lecturer. Thus the first candidate is given a pass instead of a distinction and the second is failed by the examiner.

B_u is the variance for the candidates (V_2) and if this is zero it may be seen that the whole set of marks can be "explained" in terms of a constant factor plus a random element - in other words, the marks would seem to bear no relation to the ability of the individual concerned. It may be that the marks assigned by the lecturer and examiner are not measures of the same ability but this is improbable. Even if this were the case, there would seem to be something wrong with either the examination or the method of assigning college assessments, and the net result would be the same. As far as the individual students are concerned, it would seem that the authorities might as well assign the marks at random. The other possibility is that the lecturer and (or) the examiner are very inaccurate measuring instruments and the marks are not measures of the ability of the individuals. It may be, of course, that the examination itself was a poor one and did not truly measure ability. If, on the other hand, the variance is not zero, the marks bear some relation to the ability of the individuals concerned and may be used in distinguishing between the individuals.

Now let V_1 be the variance between lecturer and examiner

V_2 " " " " candidates

V_3 " " the residual variance

$\bar{y}_1$, $\bar{y}_2$ be the means of the variables y_{1t}, y_{2t} $(t = 1, 2, 3, \ldots n)$

$\bar{y}_{..}$ be the common mean of y_{ut} $(u = 1, 2)$

$$V_1 = {}^{n}/_{2}[\bar{y}_1 - \bar{y}_2]^2$$

$$V_2 = [(y_{1t} + y_{2t})/2 - (\bar{y}_1 + \bar{y}_2)/2]^2$$

$$V_3 = [(y_{2t} - y_{1t})/2 + (\bar{y}_2 - \bar{y}_1)/2]$$

Using this set of statistics and Fisher's Z criterion[12], the significance of V_1 and V_2 may now be determined. Let f_1, f_2, f_3 denote the degrees of freedom and correspond to the number of independent differences between the constants in each case. The tests suggested by Fisher are:

(i) $Z_1 = {}^{1}/_{2} \log_e [{}^{V_1}/_{V_3} \times {}^{f_3}/_{f_1}]$

(ii) $Z_2 = {}^{1}/_{2} \log_e [{}^{V_2}/_{V_3} \times {}^{f_3}/_{f_2}$

It is easily seen that:

$$f_1 = u - 1$$

$$f_2 = n - 1$$

$$f_3 = (u - 1)(n - 1)$$

and the calculated Z_1 and Z_2 which are measures of the homogeneity between examiners and the heterogeneity between

---

[12]Fisher, R.A. <u>Statistical Methods for Research Workers</u>, Oliver & Boyd, London, 1932, p. 198 seq. and Table pp. 250 - 53.

candidates can then be compared for significance with either the 5% point or the 1% point given in the tables.

The Analysis of Variance for the marks per examiner per college is performed:

(i) for college assessment and examination marks

(ii) for original and revised marks.

In each case it is desired to test:

(a) the homogeneity of the groups of examiners when the lecturer who gives the college assessment mark is considered as an examiner as well as the person who marks the written examination paper.

(b) the heterogeneity of the groups of candidates. These qualities are considered as the two initial requirements of a good examination procedure. If Z_1 in the first case is less than Z_1 of the tables, the group of examiners is homogeneous; similarly, if Z_2 is greater than the corresponding Z_2 in the tables, the group of candidates is heterogeneous.

(i) College Assessment and Examination Marks

The results of the analysis are given in Table XVII, pages 62 - 67 (incl.) and the tests of significance of the qualities for each examiner in each college are given in Table XVIII, pages 68 and 69. The results indicate that for some lecturers and examiners the measurements of ability are in agreement; and the main errors would seem to be due to the inaccuracies of marking either by the lecturer or by the examiner. Some important facts can be gathered from the tables (pages 62 - 67 [incl.] and pages 68 and 69). As the choice of employing the 5% or the 1% significance criterion is a personal one, when the calculated Z_1 or Z_2 lies between the 5% and the 1% values in the tables, the Z in question will be counted as half significant. The results of Table XVIII, pages 68 and 69 are summarised in Table XIX, page 70, where the number of times in which Z_1 and Z_2 are not significant is given for each examiner and each college. From this table it would appear that Examiner Z is the weakest among the men examiners, while of the women examiners S and U are very weak. Again Examiners X, Y, T and W stand out as being the only efficient examiners. Of the colleges, II and III have the most heterogeneous groups of candidates, but College VI is the most satisfactory when judged on both counts. College VIII is shown again to be unreliable but in all possibility the error lies in the college assessment as all the colleges had the same examination.

TABLE XVII

Table to show Analysis of Variance for College Assessment and Examiner's Mark

College	Examiner	Source of Variation	Sum of Squares	Degree of Freedom	Mean Square
I	X	Bet. examiners	312.5	1	312.5
		" candidates	1350	15	90
		Error	237.5	15	15.83
		Total	1900	15	126.7
	Y	Bet. examiners	22.78	1	22.78
		" candidates	2639	15	175.9
		Error	153.7	15	10.25
		Total	2815	15	187.7
	Z	Bet. examiners	657.0	1	657.0
		" candidates	803.7	15	53.58
		Error	292.5	15	19.50
		Total	1753	15	116.9
II	X	Bet. examiners	30.25	1	30.25
		" candidates	5725	49	116.8
		Error	957.3	49	19.54
		Total	6713	49	137.0
	Y	Bet. examiners	235.2	1	235.2
		" candidates	4644	19	244.4
		Error	371.3	19	19.54
		Total	5251	19	276.4
III	Y	Bet. examiners	47.67	1	47.67
		" candidates	9467	58	113.2
		Error	1850	58	31.89
		Total	11365	58	195.9
	Z	Bet. examiners	3021	1	3021
		" candidates	8438	57	148.0
		Error	3587	57	62.93
		Total	15046	57	264.0

TABLE XVII contd.

College	Examiner	Source of Variation	Sum of Squares	Degree of Freedom	Mean Square
IV	X	Bet. examiners	96.15	1	96.15
		" candidates	1940	12	161.7
		Error	278.9	12	23.24
		Total	2315	12	192.9
	Y	Bet. examiners	252.3	1	252.3
		" candidates	374	12	31.17
		Error	330.1	12	27.51
		Total	956.5	12	79.71
	Z	Bet. examiners	11.12	1	11.12
		" candidates	2021	12	168.4
		Error	302.4	12	25.20
		Total	2334	12	194.5
V	X	Bet. examiners	1001	1	1001
		" candidates	2639	20	132.0
		Error	786.9	20	39.35
		Total	4427	20	221.3
	Z	Bet. examiners	312.1	1	312.1
		" candidates	1480	17	87.06
		Error	463.9	17	27.29
		Total	2256	17	132.7
VI	S	Bet. examiners	24.50	1	24.50
		" candidates	1533	24	63.88
		Error	413	24	17.21
		Total	1971	24	82.10
	T	Bet. examiners	2.44	1	2.44
		" candidates	2684	45	59.64
		Error	610.1	45	13.56
		Total	3296	45	73.25
	U	Bet. examiners	37.88	1	37.88
		" candidates	1823	32	56.98
		Error	387.1	32	12.10
		Total	2248	32	70.26
	W	Bet. examiners	.45	1	.45
		" candidates	2308	27	85.50
		Error	312.1	27	11.56
		Total	2621	27	97.07

TABLE XVII contd.

College	Examiner	Source of Variation	Sum of Squares	Degree of Freedom	Mean Square
VII	T	Bet. examiners	0	1	0
		" candidates	16167	13	1243
		Error	200	13	15.38
		Total	16368	13	1259
	U	Bet. examiners	508.6	1	508.6
		" candidates	888.5	12	74.04
		Error	503.8	12	41.99
		Total	1901	12	158.4
	V	Bet. examiners	170.0	1	170.0
		" candidates	961.6	13	73.97
		Error	524.5	13	40.34
		Total	1656	13	127.4
	W	Bet. examiners	108.0	1	108.0
		" candidates	2587	13	199.0
		Error	229.5	13	17.65
		Total	2924	13	224.9
VIII	S	Bet. examiners	567.5	1	567.5
		" candidates	325.0	16	20.31
		Error	248.5	16	15.53
		Total	1150	16	71.88
	T	Bet. examiners	277.8	1	277.8
		" candidates	672.2	17	33.66
		Error	450.0	17	26.47
		Total	1300	17	76.47
	U	Bet. examiners	212.5	1	212.5
		" candidates	503.0	16	31.43
		Error	250	16	15.63
		Total	965.5	16	60.34
	V	Bet. examiners	12.97	1	12.97
		" candidates	638.0	16	39.87
		Error	748.5	16	46.78
		Total	1399	16	87.47
	W	Bet. examiners	200.7	1	200.7
		" candidates	1056	17	62.13
		Error	511.8	17	30.11
		Total	1769	17	104.0

TABLE XVII contd.

College	Examiner	Source of Variation	Sum of Squares	Degree of Freedom	Mean Square
IX	S	Bet. examiners	0	1	0
		" candidates	2000	29	68.97
		Error	1375	29	47.41
		Total	3375	29	116.4
	T	Bet. examiners	395.1	1	395.1
		" candidates	4813	40	120.3
		Error	1080	40	27.00
		Total	6288	40	157.2
	U	Bet. examiners	269.4	1	269.4
		" candidates	1247	28	44.55
		Error	1068	28	38.15
		Total	2585	28	92.32
	V	Bet. examiners	2872	1	2872
		" candidates	4285	42	102.0
		Error	2986	42	71.10
		Total	10143	42	241.5
	W	Bet. examiners	64.29	1	64.29
		" candidates	3048	27	112.9
		Error	485.7	27	17.99
		Total	2598	27	133.3
X	S	Bet. examiners	200.9	1	200.9
		" candidates	904.5	13	69.57
		Error	461.6	13	38.51
		Total	1567	13	120.5
	T	Bet. examiners	89.29	1	89.29
		" candidates	1393	13	107.1
		Error	185.7	13	14.29
		Total	1668	13	128.3
	U	Bet. examiners	34.62	1	34.62
		" candidates	1140	12	95.03
		Error	290.4	12	24.20
		Total	1465	12	122.1
	V	Bet. examiners	378.9	1	378.9
		" candidates	740.5	13	56.96
		Error	527.6	13	40.59
		Total	1650	13	126.9

TABLE XVII contd.

College	Examiner	Source of Variation	Sum of Squares	Degree of Freedom	Mean Square
XI	S	Bet. examiners	.60	1	.60
		" candidates	911.9	20	45.60
		Error	511.9	20	25.60
		Total	1424	20	71.20
	U	Bet. examiners	.625	1	.625
		" candidates	556.9	19	29.31
		Error	811.9	19	42.73
		Total	1369	19	72.07
	V	Bet. examiners	48.40	1	48.40
		" candidates	1246	19	65.56
		Error	573.6	19	30.19
		Total	1868	19	98.29
	W	Bet. examiners	.625	1	.625
		" candidates	1887	19	99.31
		Error	311.9	19	16.41
		Total	2199	19	115.8
XII	S	Bet. examiners	2.94	1	2.94
		" candidates	611.8	16	38.24
		Error	397.1	16	24.82
		Total	1012	16	63.24
	U	Bet. examiners	389	1	389
		" candidates	605.9	16	37.87
		Error	573.3	16	35.85
		Total	1568	16	98.02
	V	Bet. examiners	56.25	1	56.25
		" candidates	1067	17	62.75
		Error	231.2	17	13.60
		Total	1354	17	79.67
	W	Bet. examiners	11.77	1	11.77
		" candidates	1378	16	86.12
		Error	263.2	16	16.45
		Total	1653	16	103.3

TABLE XVII contd.

College	Examiner	Source of Variation	Sum of Squares	Degree of Freedom	Mean Square
XIII	S	Bet. examiners	1.563	1	1.563
		" candidates	885.9	7	126.6
		Error	110.9	7	15.85
		Total	998.4	7	142.6
	U	Bet. examiners	457.1	1	457.1
		" candidates	96.43	6	16.07
		Error	67.86	6	11.31
		Total	621.4	6	103.6
	V	Bet. examiners	46.72	1	46.72
		" candidates	1256	8	157.1
		Error	49.78	8	6.222
		Total	1353	8	168.1
	W	Bet. examiners	25	1	25
		" candidates	775	7	110.7
		Error	75	7	10.71
		Total	875	7	125.0
XIV	S	Bet. examiners	0	1	0
		" candidates	755.6	17	44.44
		Error	450.0	17	26.47
		Total	1206	17	70.92
	T	Bet. examiners	6.25	1	6.25
		" candidates	1737	14	124.1
		Error	281.2	14	20.09
		Total	2024	14	144.6
	V	Bet. examiners	0	1	0
		" candidates	1680	14	120.0
		Error	583	14	41.64
		Total	2263	14	161.6
	W	Bet. examiners	25	1	25
		" candidates	880.6	17	51.80
		Error	150.0	17	8.824
		Total	1056	17	62.09

TABLE XVIII

Table to show Tests of Significance Criterion

College	Examiner	Calculated		Theoretical			
				Z_1		Z_2	
		Z_1	Z_2	5%	1%	5%	1%
I	X	1.49	.660	.757	1.08	.453	.650
	Y	.399	1.42	.757	1.08	.453	.650
	Z	1.76	.506	.757	1.08	.453	.650
II	X	.219	.894	.700	.990	.265	.375
	Y	1.25	1.26	.739	1.05	.374	.537
III	Y	.201	.816	.693	.978	.265	.375
	Z	1.94	.428	.693	.978	.265	.375
IV	X	.731	.970	.779	1.12	.494	.712
	Y	1.11	.051	.779	1.12	.494	.712
	Z	.409	.948	.779	1.12	.494	.712
V	X	1.62	.605	.732	1.04	.360	.515
	Z	1.22	.580	.746	1.06	.400	.591
VI	S	.178	.656	.725	1.03	.343	.498
	T	.858	.741	.700	.990	.265	.375
	U	.571	.775	.714	1.01	.318	.452
	W	1.62	1.00	.717	1.01	.325	.463
VII	T	0	2.20	.770	1.10	.479	.688
	U	1.25	.284	.779	1.12	.494	.712
	V	1.87	.303	.770	1.10	.488	.689
	W	.906	1.21	.770	1.10	.488	.689
VIII	S	1.80	.134	.751	1.07	.443	.634
	T	1.16	.120	.746	1.06	.434	.620
	U	1.31	.350	.751	1.07	.440	.630
	V	.642	.080	.751	1.07	.443	.634
	W	.949	.362	.747	1.06	.434	.620

TABLE XVIII contd.

Table to show Tests of Significance Criterion

College	Examiner	Calculated		Theoretical			
				Z_1		Z_2	
		Z_1	Z_2	5%	1%	5%	1%
IX	S	0	1.46	.716	1.01	.321	.457
	T	1.34	.747	.714	1.01	.318	.452
	U	.977	.078	.717	1.02	.329	.463
	V	1.85	.180	.714	1.01	.318	.452
	W	.637	.918	.719	1.01	.329	.469
X	S	.867	.336	.770	1.10	.479	.688
	T	.916	1.01	.770	1.10	.479	.688
	U	.179	.684	.789	1.12	.494	.712
	V	1.12	.170	.770	1.10	.479	.688
XI	S	1.88	.289	.735	1.05	.369	.525
	U	2.11	.189	.739	1.05	.374	.537
	V	.236	.422	.739	1.05	.374	.537
	W	1.63	.900	.739	1.05	.374	.537
XII	S	1.07	.216	.751	1.07	.443	.634
	U	1.19	.027	.751	1.07	.443	.634
	V	.710	.765	.747	1.06	.434	.620
	W	.167	.828	.751	1.07	.443	.634
XIII	S	1.16	1.04	.861	1.25	.660	.974
	U	1.85	.176	.895	1.31	.727	1.07
	V	1.01	1.61	.836	1.21	.618	.898
	W	1.58	1.17	.861	1.25	.658	.961
XIV	S	0	1.68	.747	1.06	.434	.620
	T	.584	.710	.763	1.09	.465	.668
	V	0	2.88	.763	1.09	.465	.668
	W	.521	.885	.747	1.06	.434	.620

TABLE XIX

Table to show Frequency of Non-Significance of Z Criterion

Examiner	No. of Calculated Z's				No. of Colleges Examined
	$> Z_1$ of tables	%	$< Z_2$ of tables	%	
X	2	50	0	0	4
Y	1.5	38	1	25	4
Z	3	75	1	25	4
S	3.5	44	6	75	8
T	2.5	42	1.5	25	6
U	5.5	69	6.5	83	8
V	3.5	44	5	63	8
W	4	50	1	13	8
College					**No. of Examiners Examining**
I	2	67	0.5	17	3
II	1	50	0	0	2
III	1	50	0	0	2
IV	0.5	17	1	33	3
V	2	100	0.5	25	2
VI	1	25	0.5	13	4
VII	2.5	63	2	50	4
VIII	3.5	70	2	40	5
IX	2.5	50	3.5	70	5
X	2	50	2.5	63	4
XI	2	50	2.5	63	4
XII	1.5	38	2	50	4
XIII	3	75	1	25	4
XIV	0	0	1.5	38	4

(ii) Original and Revised Examination Marks

The results of the analysis are given in Table XX, page 72; and the tests of significance of the qualities for each examiner are given in Table XXI, page 73. It will be seen that there is no significant difference in the average mark assigned by the first and the revising examiners as in every case the calculated Z_1 is less than the 5% value of the tables. There is only one value of Z_2 which is not significant. It is that of Examiner X whose Z_2 lies between the 5% and the 1% points in the Z_2 tables. It is fair, then, to say that from the point of view of the first and revising examiners the marks do bear some relation to the ability of the individuals concerned.

Table XXII, page 73 shows the examiners arranged in order of merit from three points of view:

(a) The ratio $^{V_1}/_{V_3}$ may be considered as a measure of agreement between the first and revising examiners, the least value of the ratio giving the greatest degree of agreement.

(b) The greater the ratio $^{V_2}/_{V_3}$, the greater the differentiation between the candidates.

(c) Random fluctuation in marking may be measured by the ratio $^{V_3}/_{V_4}$, where V_4 is the total of the sum of squares or total variance. Thus the greatest extent to which the

TABLE XX

Table to show Tests of Significance Criterion

Examiner	Source of Variation	Sum of Squares	Degree of Freedom	Mean Square
X	Between examiners	4.17	1	4.17
	" candidates	2800	2	1400
	Error	33.33	2	16.67
	Total	2838	2	1419
Y	Between examiners	1.8	1	1.8
	" candidates	7072	9	785.8
	Error	82.7	9	9.13
	Total	7156	9	794
Z	Between examiners	2.25	1	2.25
	" candidates	5291	7	755.8
	Error	14.75	7	2.107
	Total	5308	7	758.3
S	Between examiners	40.04	1	40.04
	" candidates	11593	11	1054
	Error	126.5	11	11.50
	Total	11760	11	1069
T	Between examiners	18	1	18
	" candidates	2817	3	939
	Error	67	3	22.33
	Total	2902	3	967.3
U	Between examiners	72,32	1	72,32
	" candidates	4915	13	378.1
	Error	265.2	13	20.40
	Total	5253	13	404.1
V	Between examiners	21.3	1	21.3
	" candidates	15550	33	471.2
	Error	244.7	33	7.415
	Total	15816	33	479.3
W	Between examiners	5.063	1	5.063
	" candidates	5152	7	736.0
	Error	73.44	7	10.49
	Total	5230	7	761.4

TABLE XXI

Table to show Tests of Significance Criterion

Examiner	Calculated		Theoretical			
	Z_1	Z_2	Z_1		Z_2	
			5%	1%	5%	1%
X	.693	2.22	1.45	2.30	1.47	2.30
Y	.812	2.23	.820	1.18	.590	.850
Z	.033	2.94	.860	1.25	.670	.980
S	.624	2.26	.790	1.13	.510	.740
T	.108	1.87	1.16	1.76	1.11	1.69
U	.633	1.46	.770	1.10	.480	.690
V	.528	2.08	.710	1.01	.320	.450
W	.364	2.13	.860	1.25	.670	.970

TABLE XXII

Table to show Examiners Arranged in Order of Merit

Examiner	V_1/V_3	V_2/V_3	V_3/V_4
Y	.197	86.0	.0114
X	.251	84.0	.0118
W	.482	70.2	.0141
T	.808	42.0	.0231
Z	1.07	359	.0028
V	2.87	63.5	.0154
S	3.48	91.7	.0108
U	3.55	18.5	.0507

examiners have been able to reduce random fluctuation is assessed by the smallest value of V_3/V_4.

In this analysis of original and revised examination marks, many of the conclusions already found regarding the examiners are confirmed.

(i) The men examiners (X, Y, Z) are, on the whole, better examiners than the women.

(ii) Examiners X, Y, T and W, when grouped with either the lecturer who gives the college assessment or the revising examiner, form the most homogeneous sub-groups.

(iii) Examiner U is by far the most unsatisfactory of all the examiners in every respect.

CHAPTER IV

ANALYSIS OF MARKS AWARDED TO CHILDREN WHO WERE EXAMINED TWICE

CHAPTER IV

ANALYSIS OF MARKS AWARDED TO CHILDREN WHO WERE EXAMINED TWICE

Problem

Until recently factor-analysis was generally applied to correlations between persons rather than to correlations between tests. In earlier investigations the number of testees or candidates was generally much larger than the number of tests and examinations; and, therefore, in the interests of economy it was natural to begin by correlating tests or examinations rather than candidates or testees. The results of such an analysis yield factors of four different types: first, one or more genreal factors; secondly, a number of group-factors or bipolar factors according to the mathematical method employed; thirdly, specific factors each peculiar to a single test or examination; and, fourthly, error factors. With a single correlation matrix it is not always possible to distinguish factors of all these different types; and in earlier work, owing to the small size of the groups examined, only a single general factor was as a rule significant.

These factors may be regarded as indicating the possible structure of the tests or examination. If the factors are identified with concrete mental abilities such as innate intelligence, memory, verbal facility, and the like, then the factor-variances show how much each of these abilities contributes to the variance displayed by the correlated or covariated sets of scores; and the factor-measurements yield an estimate of the amount of intelligence, memory, verbal facility, etc., presumably possessed by each examinee.

As Burt has pointed out, however, it is always possible with any set of scores to start by correlating or covariating, not tests or examinations, but testees or examinees. With such a procedure approximately the same factors can be demonstrated (exactly the same factors if certain condidtions are fulfilled). Since, however, the number of persons examined is usually far greater than the number of examinations or tests, the task of correlating persons would as a rule be much more formidable than the task of correlating tests.

If, however, Burt's reciprocity principle is accepted, it is possible to reach the same results that would have been obtained by correlating persons if the more usual method of correlating tests had been used. It is true that the existence of an exact reciprocity has been denied by one or two writers, notably Thomson; but he seems now to agree that even if the results of the two methods are not exactly identical they will nevertheless be approximately identical.

The following analysis aims at measuring the factors for the candidates examined and at determining the change, if any, in their factor-measurements after the lapse of a year. Since the examination is the "same" on both occasions, i.e., that of 1937 and of 1938, the question papers of the examination being designed, as far as is humanly possible, to measure the same mental qualities and to have the same standard, any change in the factor-measurements for persons should measure some change in the mental content of each individual concerned. Factor-measurements for individuals as actually calculated by analysis depend very largely upon the way the battery of tests is set up. It is agreed that intelligence as actually measured does change, for people increase in their mental age. There is apparently no means whereby factor-analysis can distinguish a mental age from a mental ratio or I.Q. The I.Q. in theory does not change. Hence, any change in the factor-measurement for an individual will probably measure the change in some quality other than intelligence.

It is the purpose of this analysis to determine whether there is (a) any change in the factor pattern, and, (b) any change in the factor-measurements for persons who are entered for the "same" examination on two occasions with an interval of a year between the two examinations.

Material

Each child in the elementary schools in certain of the counties is examined twice for promotion to the secondary schools, i.e., at the age of 10+ and again at 11+. Some of the most intelligent of these children may go to the secondary school after the first examination which means that those who are left to take the examination for the second time form a relatively homogeneous group. In a particular county, the children are card-indexed with the marks for all subjects for each of the two years. Out of some 2000 cards a random sample of 100 pairs of such marks from examinations in 1937 and in 1938 has been used for analysis in this section of the research. The subjects of the examination are arithmetic, general intelligence and English. The average age of the children was 11.6 at the time of the examination in 1938.

Statistical Methods

Since it is required to test the stability of the factor-pattern of the scores for each of the two years, the raw marks of the children were correlated between subjects. The matrix of correlations thus obtained for each year was then analysed separately by Burt's Weighted Summation Method.[13]

---

[13]Burt, C. The Factors of the Mind, U. of L. Press, London, 1940. p. 467, seq.

To measure the change in the factor-measurements of each child after the lapse of a year, the first and second factor-measurements for each of the 100 children for the two years were computed by Burt's method.[14] To insure that the tests are consistent the factor-measurements of the tests were also correlated. To illustrate the use of the statistical methods applied, the marks of 5 children only were first analysed. The random selection of these children was made as follows: first, the cards of the 100 children were arranged numerically according to the numbers given to them by the county card-indexing system. Then a colleague was asked to name any five numbers between 1 and 100. These were 15, 29, 47, 63 and 89. Finally the 15th, 29th, 47th, 63rd and 89th cards starting from the first of the pile of cards already arranged as described above were taken out. The marks on these cards were used for the preliminary analyses of the stability of the factor-pattern and of the investigation of the change, if any, in the factor-measurements for each child. Having obtained satisfactory results by the analyses of these children, the 5 cards were re-inserted in the pile and the two separate analyses for each year were made on the marks of the original 100 children.

---

[14] Burt, C. Correlations between Persons, B.J.P., XXVIII, July 1937, pp. 59 - 96.

Results

(a) Tests

The correlations between the two sets of raw marks are given below (i) for 5 children, (ii) for 100 children:

TABLE XXIII

Table to show Correlations of Marks

(i) 5 children

	1937		1938	
	Intelligence	English	Intelligence	English
Arithmetic	.806	.828	.765	.659
Intelligence		.844		.730

(ii) 100 children

	1937		1938	
	Intelligence	English	Intelligence	English
Arithmetic	.608	.595	.765	.730
Intelligence		.712		.659

The extraordinary similarity between the correlations with (i) 5 children and (ii) 100 children in 1938 is almost uncanny but is perfectly genuine. The value of .659 for Arithmetic/English for the 5 children in 1938 is just below the significance level for $P = .1$; all the others are statistically significant.

The saturation coefficients obtained by the method of weighted summation are given in Table XXIV, page 81.

TABLE XXIV

Table to show Saturation Coefficients by Weighted Summation

(i) 5 children

Saturation Coefficients	1937			1938		
	1	2	3	1	2	3
Arithmetic	.9029	.1632	-.0834	.8473	-.1740	.0793
Intelligence	.9147	-.1278	-.0397	.9123	-.0332	-.1089
English	.9312	-.0347	.1186	.8174	.2243	.0803

(ii) 100 children

Saturation Coefficients	1937			1938		
	1	2	3	1	2	3
Arithmetic	.8406	-.1299	.0107	.8455	-.1805	.0482
Intelligence	.8516	-.1009	-.1342	.9061	-.0283	-.1165
English	.7466	.2612	.1268	.8188	.2193	.0775

There are some differences in the saturation coefficients of the tests with five children only. They are: (i) smaller values for the first, probably 'g'; (ii) the second which shows antithesis between arithmetic and English in both years links English with intelligence in 1937 and may be considered as 'v' but it links arithmetic with intelligence in 1938 when it is probably a 'number' factor; (iii) the third saturation coefficient in 1937 is comparable to the second in 1938, the latter having larger values, but the third in 1938 links arithmetic and English and may be a memory factor.

It is the results of the tests with the 100 children that are important, those with five children only having been made for the purpose of comparison at a later stage in the research. With the 100 children there is a remarkable correspondence between the saturation coefficients of the two years shown in Table XXIV, page 81. In all cases, the signs are the same and the subjects keep the same order in the size of their weightings. The first factor in both cases gives the weightings of 'g', the heaviest being that of intelligence as would be expected, and the least, that of English. The second factor is probably verbal, 'v', being relatively larger for English and negative for arithmetic and intelligence. The third factor is probably a memory, 'm', factor as it is negative for intelligence, small and positive for arithmetic and slightly larger and positive for English.

Table XXV, page 83, gives the test factor-measurements that are obtained by using Burt's method of correlations between persons. Since a double-centred matrix has been used, a general factor, which might be identified with general intelligence, has been eliminated. There are three specific factors which can be represented by tow bipolar factors. The first factor deals with the antithesis between arithmetic and English and as the factor-measurement for intelligence falls into line with that of English, this may be called a verbal factor. The measurements of the second factor for each year are alike in sign and order of magnitude as they are in the first. The second factor again shows antithesis between English and arithmetic but as intelligence lines up with arithmetic, in this case the factor is probably a numerical one. The remarkable similarity between the test factor-measurements for the two years shows that the tests themselves are measuring the same mental qualities each year and that related to each other, subject for subject, the examinations are accurate instruments.

TABLE XXV

Table to show Test Factor-Measurements

Test Factor Measurement	1937		1938	
Arithmetic	-.8158	.0339	-.7842	.2272
Intelligence	.4373	.6895	.5889	.5655
English	.3785	-.7234	.1953	-.7928

The two sets of raw marks for each of the three subjects were correlated to obtain a measure of reliability of the tests themselves. The correlation coefficients are as follows: Arithmetic, .691, Intelligence, .787 and English, .798.

(b) Persons

The test factor-measurements ($P_p = L_t'$ according to Burt's notation) which were used in section (a) for the purpose of showing agreement between the tests are the direction cosines obtained in the process of computing the factor-measurements for persons. The marks for both years were dealt with and Table XXVI, page 85, gives the two factor measurements (P_t) for each of the 100 children for both years.

The correlation between the first factor-measurements for the years 1937 and 1938 is .396 and that of the second factor-measurement is .375. The values are significant for the level $P = .01$ when $n = 100$. This significance of the correlations between corresponding factor-measurements for persons of the two years shows, in general, that the children did not alter relatively in their test results after the lapse of a year. At the same time, there are 49 children who have a change of sign in one of their factor-measurements and 11 children who have the sign changed for both factor-measurements. Hence, though the correlations between the factor-measurements are significant for the group, nearly 50% of the children show a change in some mental quality or qualities and about 10% of them change

TABLE XXVI

Table to show Factor-Measurements for the Same Children for years 1937 and 1938

Children	1937		1938	
	First	Second	First	Second
1	.0356	-.0508	.0235	.1383
2	.0396	.0946	.0632	.2021
3	-.0495	-.1182	.0377	-.0816
4	-.0356	.0508	.0593	-.0887
5	-.1379	.0329	-.1691	-.0211
6	-.1586	-.0466	-.1117	-.1630
7	-.0397	-.0426	-.1189	-.1914
8	-.0851	-.0674	-.0756	-.0212
9	-.0604	.1575	.0271	.1525
10	-.0857	.2480	-.1025	.0108
11	-.1757	.0030	-.0865	-.1559
12	.0820	-.0456	.0522	.0673
13	-.0437	.0917	-.1205	.0320
14	-.0816	.0618	.0144	.0567
15	.1194	-.0318	.2862	.1593
16	-.1045	-.0820	-.1403	.0002
17	-.0270	.0260	.0701	-.0462
18	.0451	-.0433	.1153	.1772
19	-.0369	.0023	.1134	.0318
20	-.0284	-.0225	-.0360	-.0496
21	-.0113	-.0721	-.0216	.0071
22	.1197	.3160	.0522	.0673
23	-.0334	.1315	-.0810	.0036
24	-.0284	-.0225	.0035	-.0780
25	-.0445	.0271	-.0451	-.1312
26	.1955	.0443	-.0898	.1987
27	.2699	.0558	-.0125	.0887
28	-.2076	-.1652	-.2159	-.0210
29	.0977	.1880	-.0863	.0285
30	-.1550	.0825	-.1061	-.0054
31	.1117	.0253	-.0396	-.0638
32	-.0563	-.0288	-.1080	-.0566
33	-.1027	-.0340	-.0252	-.0071
34	.1032	.0501	.2159	.0210
35	-.0293	-.0548	.2033	.0175
36	-.0576	-.0773	.0845	-.0818
37	-.0550	.0196	-.0379	-.1028
38	.0712	-.1015	.0504	.0141
39	-.1460	.0739	-.0377	.0816
40	-.0343	.0992	.0631	.1099
41	-.0238	-.1926	.1259	-.0569
42	.1420	-.2192	.0665	-.0604
43	.0266	-.0421	.0862	-.1207
44	-.0374	-.0138	.0504	.0414
45	.0338	-.1153	.2571	-.1386
46	-.0590	-.1257	.1079	-.0356
47	.0658	.0363	-.0539	.0639
48	-.0838	-.0190	.1240	-.1101
49	.0270	-.0260	-.0127	-.0957
50	.0207	.0796	.0522	.0673
51	.0748	.0276	.0774	.0744
52	-.0635	.0444	-.0448	.1455
53	-.2186	.1270	-.0611	.0355
54	.0876	.1643	.0144	.0567
55	.2316	.0097	.0252	.0071
56	-.0825	.0295	-.0523	-.1595
57	.1564	-.0341	-.0200	-.2163
58	-.0684	-.1332	-.0234	-.0461
59	.0477	.0536	-.1258	.1491
60	.1451	-.1062	.0666	.0318
61	-.0216	-.1119	.0161	-.0745
62	.2685	.0074	.0701	-.0462
63	-.0955	-.1072	-.1530	-.0956
64	.0055	-.1378	.1045	.1345
65	-.1212	-.0328	-.1260	-.0353
66	-.0681	.2146	-.2698	.0429
67	-.0013	-.0484	-.0037	-.1064
68	-.1131	-.0738	-.0702	-.0460
69	.0182	-.3489	-.0865	-.1559
70	.0563	.0288	-.0162	-.0177
71	-.0013	-.0484	.1007	-.0640
72	.0838	.0190	-.0163	-.1099
73	-.1077	.1200	-.0558	.0107
74	.0838	.0190	-.0700	.1384
75	-.0794	.1425	-.0198	-.0319
76	.0134	.1529	-.0287	.0710
77	.0418	.1753	-.0502	.1703
78	.1762	.0132	-.0072	-.0284
79	.1144	.1222	-.0754	.1632
80	.0748	.0276	.0558	.0815
81	.0653	.0202	.0665	-.0604
82	.1199	-.0156	.1276	-.0959
83	.0361	-.0346	-.1132	.1526
84	-.0297	-.0709	.1458	.0672
85	-.0392	-.0784	-.0503	.0781
86	.0897	-.1027	.1295	-.0427
87	-.0202	-.0634	.0467	-.0923
88	-.0365	.0187	.0577	.1347
89	.0627	-.0767	-.0128	-.1880
90	.0653	.0202	-.0485	.1313
91	-.1469	.0416	-.1312	.0817
92	-.0293	-.0548	-.0542	-.2127
93	-.0068	.0894	-.0611	.0355
94	.1113	.0092	.0882	.0247
95	.1068	-.1523	.0774	.0744
96	.1190	-.0479	.1601	-.0605
97	-.1856	-.0207	-.0826	.1349
98	-.0996	.0791	-.1024	.1030
99	-.0640	.0283	.2302	-.1067
100	.0081	-.0410	-.0990	-.0673

in still more mental qualities. Since it is not possible to determine these qualities, nor the direction of the change, it cannot be said whether the children in quastion profit by another year's schooling.

This chapter is not complete unless the test factor-measurements and the factor-measurements for the five children for both years are recorded. These are given in Tables XXVII and XXVIII.

TABLE XVII

Table to show Test Factor-Measurements for the 5 Children for both Years

Test Factor Measurement	1937		1938	
Arithmetic	.7964	-.1800	-.5648	.5897
Intelligence	-.5541	-.5997	-.2283	-.7839
English	-.2423	.7797	.7931	.1942

These test factor-measurements agree in sign with the second and third saturation coefficients of Table XXIV (i), p. 81, and in order of weighting with one exception in each year.

TABLE XXVIII

Table to show Factor-Measurements for Each Child

Year		Factor-Measurements for Children				
		(a)	(b)	(c)	(d)	(e)
1937	(i)	.058286	-.705303	-.029110	-.029176	.705303
	(ii)	-.080996	-.004987	.744100	-.663103	.004987
1938	(i)	-.543485	.053608	.404108	-.474208	.559978
	(ii)	-.484848	.753387	.002127	.148125	-.418791

CHAPTER V

CONCLUSION

CHAPTER V

CONCLUSION

The conclusions to each of the parts of the research will be summed up under separate headings.

I. Some of the Effects Caused by Marking in Accordance with a Normal Distribution

(a) A factor pattern is produced which is fairly constant from year to year.

(b) There is evidence that the enjoining of examiners to mark their papers in accordance with a normal distribution produces:

(i) curves which approximate to the normal distribution curve,

(ii) averages which are fairly constant among the various subjects and also from year to year,

(iii) smaller fluctuations of the standard deviation.

(c) Examination I, the subjects of which are Principles of Education, Psychology and Hygiene, Special Method, Educational Psychology, Educational System, Essay and Practical Teaching, is confined to post-graduate candidates. It yields figures which show that its candidates form a more homogeneous group than do those of

Examination II whose candidates are in their first year of college life and the subjects of which are Inorganic, Organic, Mathematical and Social Sciences.

(d) (i) The general factor in the sense of the factor which contributes most to the variance may be roughly identified with general intelligence.

(ii) The contribution of this general factor is smaller with the post-graduate candidates of Examination I.

(iii) The other factors of Examination I make a greater contribution to the variance than do those of the undergraduate candidates of Examination II.

II. Analysis of Variance between Two Estimates of Marks for a University Examination

(a) There appears to be a general tendency for a lecturer who is attached to the same college as the candidates:

(i) to be more generous in his marking of them than another examiner who is not attached to the same college as the candidates,

(ii) to be more reluctant to award low marks which entail failures because he or she would have advised candidates expected to fail to withdraw,

(iii) to be less accurate than the person marking the written examination if judged by the standards of the revising examiner.

due to the way in which the women are examined or taught. Possibly it is due to the fact that women as a group are more homogeneous than men as a group.

(f) The individual differences in accuracy among the college lecturers is very noticeable in Colleges III and IX but less so in Colleges I, VII and X. On the other hand, according to Examiner Y for College IV, Examiner S for College VIII and Examiner U for Colleges X, XII and XIII, the lecturers have been excessively diffident and have awarded an insufficient number of high marks.

(g) The analysis of the results of these examinations shows clearly that the difference in efficiency between examiners is larger than might have been expected.

III. Analysis of Marks Awarded to Children Who Were Examined Twice

(a) (i) The general factor, which may be identified with intelligence in 1937 and in 1938, contributes approximately the same values to the variance each year.

(ii) The two group-factors on both occasions can each be identified with memory and verbal ability respectively, and the factor saturations of the tests remain approximately the same.

(b) (i) Considering the whole group, the correlations between the corresponding factor-measurements for individuals are low but statistically significant.

(ii) There is a change of sign in both the factor-measurements of 10% of the individuals.

APPENDIX

TWO ALTERNATIVE METHODS TO OBTAIN FACTOR MEASUREMENTS OF PERSONS

APPENDIX

TWO ALTERNATIVE METHODS TO OBTAIN FACTOR MEASUREMENTS OF PERSONS

Problem and Material

Two methods for solving the equations the coefficients in which the unknown quantities form a double-centred matrix, were tried out. The purpose of these analyses is to show the agreement of the results with those obtained by Burt's method used in Chapter IV and described by him in his article entitled "Correlations between Persons" to which reference has already been made. To avoid making the enormous number of computations necessitated by analyses with 100 persons, the marks of the 5 children who were picked at random for use in Chapter IV, were used again for the analyses now to be described. It is understood that these methods are original and not yet published. They will, therefore, be described in full as they were used to compute the results. The actual figures used in the computations will be given in so far as they are relevant.

A. The K-Method for Solving Characteristic Equations with Multiple Latent Roots

This method was worked out by the author of this research in collaboration with a colleague[16]. It expresses the five factor-measurements for each of the five children in terms of two quantities, k_1 and k_2. The existence of quantities like k_1 and k_2 is due to the fact that the matrix M'M has a triple zero latent root. The number of quantities k_1 and k_2 (in this case two) is equal to the number of non-zero roots of M'M.

The working of this method is most easily explained by applying it to find the factor-measurements of the matrix M'M, i.e., the product sums for the columns of the matrix M (below). The original marks, expressed as a deviation from the mean mark per test and per person are:-

	Person	Arithmetic	Intelligence	English
M =	(a)	-1	0	1
	(b)	10	-7	-3
	(c)	2	5	-7
	(d)	-1	-5	6
	(e)	-10	7	3

	Person	(a)	(b)	(c)	(d)	(e)
R_p = M'M,	(a)	2	-13	-9	7	13
	(b)	-13	158	6	7	-158
	(c)	-9	6	78	-69	-6
	(d)	7	7	-69	62	-7
	(e)	13	-158	-6	-7	158

---

[16] Head, J.W. & Harwood, M.K.B. The K-Method for Solving Characteristic Equations with Multiple Latent Roots, 1942. Vibration Dept., de Havilland Aircraft Co. Ltd. Stag Lane, Edgware, Mdx. (Unpublished)

Let p_i, q_i, r_i, s_i, t_i, be the solutions of the equations:

$$\left.\begin{array}{rrrrrl}
(2-\lambda_i)p_i & - \; 13q_i & - \; 9r_i & + \; 7s_i & + \; 13t_i & = 0 \\
-13p_i & + \; (158-\lambda_i)q_i & + \; 6r_i & + \; 7s_i & - \; 158t_i & = 0 \\
-9p_i & + \; 6q_i & + \; (78-\lambda_i)r_i & - \; 69s_i & - \; 6t_i & = 0 \\
7p_i & + \; 7q_i & - \; 69r_i & + \; (62-\lambda_i)s_i & - \; 7t_i & = 0 \\
13p_i & - \; 158q_i & - \; 6r_i & - \; 7s_i & + \; (158-\lambda_i)t_i & = 0 \\
 & & p_i^2 + q_i^2 + r_i^2 + s_i^2 + t_i^2 & & & = 1
\end{array}\right\} [A]$$

It is required to find the p's and q's corresponding to non-zero latent roots. Since the rank of the above five-fold matrix is only 2, the columns of the matrix can be expressed in terms of any two columns. Taking the first two columns, the equations [B] are constructed:

$$\left.\begin{array}{rcl}
p_i & = & 2k_1 - 13k_2 \\
q_i & = & -13k_1 + 158k_2 \\
r_i & = & -9k_1 + 6k_2 \\
s_i & = & 7k_1 + 7k_2 \\
t_i & = & 13k_1 - 158k_2
\end{array}\right\} [B]$$

Substituting these values in [A],

$$2(2k_1 - 13k_2) - 13(-13k_1 + 158k_2) - 9(-9k_1 + 6k_2) + 7(7k_1 + 7k_2) + 13(13k_1 - 158k_2) = \lambda_i(2k_1 - 13k_2)$$

etc. [C]

$$k_1(2^2 + 13^2 + 9^2 + 7^2 + 13^2 + 2\lambda_i) + k_2[2(-13) - 13.158 - 9.6 + 7.7 + 13(-158) + 13\lambda_i] = 0$$

and

$$\left.\begin{array}{rrl}
(472 - 2\lambda_i)k_1 + & (-4139 + 13\lambda_i)k_2 & = 0 \\
(-4139 + 13\lambda_i)k_1 + & (50182 - 158\lambda_i)k_2 & = 0 \\
(-1359 + 9\lambda_i)k_1 + & (1998 - 6\lambda_i)k_2 & = 0 \\
(887 - 7\lambda_i)k_1 + & (2141 - 7\lambda_i)k_2 & = 0 \\
(4139 - 13\lambda_i)k_1 + & (-50182 - 158\lambda_i)k_2 & = 0
\end{array}\right\} [D]$$

when $\lambda = \lambda_i$

On expanding the determinant M'M in powers of λ in terms of the sums of the principal minors, the characteristic equation for R_p is:

$$-\lambda^3 + 458\lambda^2 - 44589\lambda + 0 = 0$$

whence $\lambda = 0$, 317.61151 and 140.38849

When $\lambda_i = \lambda_1 = 317.61151$, is substituted in [D],

$$k_1 = -.0615743k_2$$

This gives:

$$\begin{aligned} p_1 &= -13.1231486k_2 \\ q_1 &= 158.8004659k_2 \\ r_1 &= 6.5541687k_2 \\ s_1 &= 6.5689799k_2 \\ t_1 &= -158.8004659k_2 \end{aligned}$$

and using $p_1^2 + q_1^2 + r_1^2 + s_1^2 + t_1^2 = 1$

$$\begin{aligned} p_1 &= .058286 \\ q_1 &= -.705303 \\ r_1 &= -.029110 \\ s_1 &= -.029176 \\ t_1 &= .705303 \end{aligned}$$

By the same process when $\lambda_i = \lambda_2 = 140.38849$,

$$k_1 = 12.10078k_2$$

and

$$\begin{aligned} p_2 &= -.080996 \\ q_2 &= -.004988 \\ r_2 &= .744099 \\ s_2 &= -.663103 \\ t_2 &= .004988 \end{aligned}$$

B. On an "Escalator" Method for the Numerical Solution of Lagrangian Frequency Equations

The second method, which was evolved jointly by Morris and Head[17] will be referred to as the "Escalator" Method.

---

[17] Morris, J. & Head, J.W. On an "Escalator" Method for the Numerical Solution of Lagrangian Frequency Equations, to be published in Aircraft Engineering.

Consider the equations:

$$(a_{11} - \lambda)x + a_{12}y + a_{13}z = 0 \quad . \; . \; . \; . \; (i)$$
$$a_{12}x + (a_{22} - \lambda)y + a_{23}z = 0 \quad . \; . \; . \; . \; (ii)$$
$$a_{13}x + a_{23}y + (a_{33} - \lambda)z = 0 \quad . \; . \; . \; . \; (iii)$$

There are three values $\lambda_1 > \lambda_2 > \lambda_3$ of λ for which these are consistent. Only the ratios of x:y:z are determinate for these values of λ, but a set of values of x, y, z which at the same time satisfies both the above equations and also the equation $x^2 + y^2 + z^2 = 1$ will be referred to as a 'rectified set'. There is a unique rectified set corresponding to the value λ_i which will be denoted by (x_i, y_i, z_i) $(i = 1, 2, 3)$ if $(-x_i, -y_i, -z_i)$ and (x_i, y_i, z_i) are regarded as the same set. The object of the Escalator method is to obtain λ_i, x_i, y_i, z_i for the equations:

$$\left.\begin{aligned}(a_{11} - \lambda)x + a_{12}y &= 0\\ a_{12}x + (a_{22} - \lambda)y &= 0\end{aligned}\right\} \quad [B]$$

Suppose that [B] are consistent when $\lambda = \Lambda_1$ or Λ_2 $(\Lambda_1 > \Lambda_2)$ and that X_i, Y_i are the rectified set corresponding, (that is X_i, Y_i satisfy [B] and $X_i^2 + Y_i^2 = 1$ when $\lambda = \Lambda_i$), then it can be proved that

$$\left.\begin{aligned}X_iX_j + Y_iY_j &= 0 \text{ when } i \neq j\\ X_1Y_1 + X_2Y_2 &= 0\\ X_1^2 + X_2^2 &= 1\\ Y_1^2 + Y_2^2 &= 1\end{aligned}\right\} \quad [D]$$

Now let
$$P_1 = a_{13}X_1 + a_{23}Y_1$$
$$P_2 = a_{13}X_2 + a_{23}Y_2$$

Multiply (i) by X_1 and (ii) by Y_1 and add

and $(\Lambda_1 - \lambda)(X_1 x + Y_1 y) + P_1 z = 0$ from [B]

Multiply (i) by X_2 and (ii) by Y_2 and add

and $(\Lambda_2 - \lambda)(X_2 x + Y_2 y) + P_2 z = 0$

$$\left.\begin{aligned} \therefore X_1 x + Y_1 y &= -\frac{P_1 z}{\Lambda_1 - \lambda} \\ X_2 x + Y_2 y &= -\frac{P_2 z}{\Lambda_2 - \lambda} \end{aligned}\right\} \quad [C]$$

Next multiply [C] by $X_1 X_2$ and add and from [D]

$$\left.\begin{aligned} x &= -z\left[\frac{P_1 X_1}{\Lambda_1 - \lambda} + \frac{P_2 X_2}{\Lambda_2 - \lambda}\right] \\ y &= -z\left[\frac{P_1 Y_1}{\Lambda_1 - \lambda} + \frac{P_2 Y_2}{\Lambda_2 - \lambda}\right] \end{aligned}\right\} \quad [E]$$

Substituting in (iii)

$$(a_{33} - \lambda)z = a_{13} z\left[\frac{P_1 X_1}{\Lambda_1 - \lambda} + \frac{P_2 X_2}{\Lambda_2 - \lambda}\right] + a_{23} z\left[\frac{P_1 Y_1}{\Lambda_1 - \lambda} + \frac{P_2 Y_2}{\Lambda_2 - \lambda}\right]$$

$$\text{and } \therefore F(\lambda) = a_{33} + \lambda + \frac{P_1^2}{\Lambda_1 - \lambda} + \frac{P_2^2}{\Lambda_2 - \lambda} = 0$$

This last equation is the same as the determinant of the equations (i), (ii) and (iii) but is in a much more convenient form for computation.

To obtain a rectified set it is noted that

$$x^2 + y^2 + z^2 = z^2\left[1 + \frac{P_1^2}{(\Lambda_1 - \lambda)^2} + \frac{P_2^2}{(\Lambda_2 - \lambda)^2}\right]$$

is obtained by direct squaring and adding and using [D].

$1 + \frac{P_1^2}{(\Lambda_1 - \lambda)^2} + \frac{P_2^2}{(\Lambda_2 - \lambda)^2}$ is $\frac{dF(\lambda)}{d\lambda}$ which quantity is used in the method described by Manley for the approximation to the solution of a numerical equation[18].

Having obtained the rectified sets x_i, y_i, z_i corresponding to $\Lambda = \lambda_i$ $(i = 1, 2, 3)$ the equations:

$$(a_{11} - \lambda)x + a_{12}y + a_{13}z + a_{14}u = 0, \text{ etc.},$$

can be solved in a similar manner, obtaining the solution:

$$x = -u\Sigma \frac{P_i x_i}{\Lambda_1 - \lambda}$$

$$y = -u\Sigma \frac{P_i y_i}{\Lambda_2 - \lambda}$$

$$z = -u\Sigma \frac{P_i z_i}{\Lambda_3 - \lambda}$$

where $P_i = a_{14}x_i + a_{24}y_i + a_{34}z_i$ and λ is a root of

$$\frac{P_1^2}{\Lambda_1 - \lambda} + \frac{P_2^2}{\Lambda_2 - \lambda} + \frac{P_3^2}{\Lambda_3 - \lambda} = a_{44} - \lambda,$$

and so on.

The actual computations made to obtain the factor-measurements of the five children already mentioned will now be given to demonstrate the use of the Escalator Method.

---

[18] Manley, R.G. Fundamentals of Vibration Study, Chapman & Hall, London, 1942. p. 118.

The matrix M'M, i.e., the product sums for the columns of the matrix M (see p. 93) is used.

R_p = M'M,

Person	(a)	(b)	(c)	(d)	(e)
(a)	2	-13	-9	7	13
(b)	-13	158	6	7	-158
(c)	-9	6	78	-69	-6
(d)	7	7	-69	62	-7
(e)	13	-158	-6	-7	158

2nd Order

Form quadratic from top left-hand corner (2 columns and 2 rows)

$$(2 - \lambda)(158 - \lambda) = (-13)^2$$
$$\text{whence } \lambda = 159.0759,\ .9241$$

Using λ_1 = 159.0759 in

$$\left.\begin{array}{r} (2 - \lambda)p - 13q = 0 \\ -13p + (158 - \lambda)q = 0 \\ p^2 + q^2 = 1 \end{array}\right\} \text{[A]}$$

$$p_1 = -.082480r_1$$
$$q_1 = .996593r_1$$

Using λ_2 .9241 in [A],

$$p_2 = .996593r_2$$
$$q_2 = .082480r_2$$

$[\sqrt{\beta_1} = \sqrt{\beta_2} = 1]$ (see page 100)

3rd Order

(using 3 columns and 3 rows, top left-hand corner)

$P_1 = -9p_1 + 6q_1 = 6.721878, \quad P_1^2 = 45.183644$

$P_2 = -9p_2 + 6q_2 = -8.474457, \quad P_2^2 = 71.816421$

[Check

$P_1p_1 + P_2p_2$ should = -9, actual value = -9.000005

$P_1^2q_1 + P_2^2q_2$ " = 6, " " = 6.000003]

Equation

$$\begin{vmatrix} 2 - \lambda & -13 & -9 \\ -13 & 158 - \lambda & 6 \\ -9 & 6 & 78 - \lambda \end{vmatrix} = 0 \text{ is the same as}$$

$$\frac{6.721878^2}{159.0759 - \lambda} + \frac{(-8.474457)^2}{.9241 - \lambda} = 78 - \lambda = \alpha$$

Approximations (a) to λ's of 3rd order

a	0	160	159.3943	159.5308	159.6140	159.6319	159.6325
$\Lambda_1 - a$	159.0759	-.9241	-.3184	-.4549	-.5381	-.5560	-.5566
$\Lambda_2 - a$	.9241	-159.0759	-158.4702	-158.6067	-158.6899	-158.7079	-158.7084
$78 - a$	78	-82	-81.3943	-81.5308	-81.6140	-81.6319	-81.6325
$P_1^2/(\Lambda_1 - a)$	.284038	-48.894755	-141.905429	-99.326542	-83.968860	-81.265546	-81.177944
$P_2^2/(\Lambda_2 - a)$	77.714988	-.451460	-.453185	-.452795	-.452558	-.452507	-.452505
α	.000794	-32.653785	60.967314	18.248537	2.807418	.086153	-.002051
$P_1^2/(\Lambda_1 - a)^2$	.001785	52.910675	445.692302	218.348080	156.046942	146.161053	145.846108
$P_2^2/(\Lambda_2 - a)^2$	84.098028	.002838	.002859	.002854	.002851	.002581	.002581
β	85.099813	53.913513	446.695161	219.350934	157.049793	147.163904	146.848959
$\delta = \alpha/\beta$	.000011	-.6057	.1365	.0832	.0179	.0006	.00001

δ = gradient and is added to last value of a to get new a

$$\beta = P_1^2/(\Lambda_1 - a)^2 + P_2^2/(\Lambda_2 - a)^2 + 1$$

$\lambda_1 = 0$
$\lambda_2 = 159.6325$
$\lambda_3 = 78.3675$ (by similar method)

$$p_1 = \left[-\frac{P_1p_1}{(\Lambda_1 - \lambda)\sqrt{\beta_1}} - \frac{P_2p_2}{(\Lambda_2 - \lambda)\sqrt{\beta_2}}\right]r_1 = 9.142739r_1$$

$q_1 = .714271r_1$ $\quad \sqrt{\beta_1} = 9.224956$

$p_2 = -1.049297r_2$

$$q_2 = \left[-\frac{P_1q_1}{(\Lambda_1 - \lambda)\sqrt{\beta_1}} - \frac{P_2q_2}{(\Lambda_2 - \lambda)\sqrt{\beta_2}}\right]r_2 = 12.031127r_3$$

$\sqrt{\beta_2} = 12.118125$

$p_3 = -.102185r_3$
$q_3 = -.092027r_3$ $\quad \sqrt{\beta_3} = 1.009411$

Check:
$$p_1^2 + q_1^2 + r_1^2 = \beta_1$$
$$p_2^2 + q_2^2 + r_2^2 = \beta_2$$
$$p_3^2 + q_3^2 + r_3^2 = \beta_3$$

Note - P_1p_1, P_1q_1, etc. in brackets are taken from previous order.

4th Order

$P_1 = -.000100$ (to be neglected from now on)
$P_2 = .649672$, $P_2^2 = .422074$
$P_3 = -69.703504$, $P_3^2 = 4858.578470$

[Check

$\frac{P_2 p_2}{\sqrt{\beta_2}} + \frac{P_3 p_3}{\sqrt{\beta_3}}$ should = 7, actual value = 6.999992

$\frac{P_2 q_2}{\sqrt{\beta_2}} + \frac{P_3 q_3}{\sqrt{\beta_3}}$ " = 7, " " = 6.999806

$\frac{P_2 r_2}{\sqrt{\beta_2}} + \frac{P_3 r_3}{\sqrt{\beta_3}}$ " = -69, " " = -69.000029]

Equation

$$\begin{vmatrix} 2-\lambda & -13 & -9 & 7 \\ -13 & 158-\lambda & 6 & 7 \\ -9 & 6 & 78-\lambda & -69 \\ 7 & 7 & -69 & 62-\lambda \end{vmatrix} = 0$$

is the same as $\frac{.422074}{159.6325-\lambda} + \frac{4858.578470}{78.3675-\lambda} = 62 - \lambda$

By similar approximations as in 3rd order:

$\lambda_1 = 0$
$\lambda_2 = 140.3564$
$\lambda_3 = 159.64365$

$p_2 = .116748 s_2$
$q_2 = .069054 s_2$
$r_2 = -1.116748 s_2$

$\beta_2 = 1.505166$

$p_3 = -4.958383 s_3$
$q_3 = 57.926348 s_3$
$r_3 = 5.958544 s_3$

$\beta_3 = 58.281458$

5th Order

$P_2 = -6.439369$, $P_2^2 = 41.465473$

$P_3 = -158.670931$, $P_3^2 = 25176.464344$

Equation
$$\begin{vmatrix} 2-\lambda & -13 & -9 & 7 & 13 \\ -13 & 158-\lambda & 6 & 7 & -158 \\ -9 & 6 & 78-\lambda & -69 & -6 \\ 7 & 7 & -69 & 62-\lambda & -7 \\ 13 & -158 & -6 & -7 & 158-\lambda \end{vmatrix} = 0$$

is the same as $\frac{41.465473}{140.3564 - \lambda} + \frac{25176.464344}{159.64365 - \lambda} = 158 - \lambda$

By similar approximations as in 3rd order:

$\lambda_1 = 0$
$\lambda_2 = 317.61129$
$\lambda_3 = 140.38854$

$p_2 = .082367t_2$	$p_3 = -16.241454t_3$
$q_2 = -.999998t_2$	$q_3 = -1.001568t_3$
$r_2 = -.041269t_2$	$r_3 = 149.210821t_3$
$s_2 = -.041370t_2$	$s_3 = -132.969313t_3$
$\beta_2 = 1.417830$	$\beta_3 = 200.525618$

By using the equation $p^2 + q^2 + r^2 + s^2 + t^2 = 1$ the values of p, q, r, s and t are determined; they are given in the table below.

TABLE XXIX

Table to show Factor-Measurements for 5 Children as Determined by the Three Different Methods

Method		Factor-Measurements				
Burt's	(i)	.058286	-.705303	-.029110	-.029176	.705303
	(ii)	-.080996	-.004987	.744100	-.663103	.004987
K	(i)	.058286	-.705303	-.029110	-.029176	.705303
	(ii)	-.080996	-.004988	.744099	-.663103	.004988
Escalator	(i)	.058284	-.705302	-.029107	-.029178	.705303
	(ii)	-.080999	-.004989	.744099	-.663104	.004987

Table XXIX gives the factor-measurements for each of the five children as obtained by the three different methods of computation. It will be noted that the agreement is perfect for five significant figures. It is not intended to claim that any one of these methods is better than either of the other two. Experience in another field of reserach, which must not be disclosed by reason of the Official Secrets Act, has shown that when the number of equations to be solved is six or more, the time and labour necessitated by the expansion of the determinant formed by the coefficients of the unknown quantities of the equations, are cut down enormously by use of the Escalator method. It is understood that such exponents of the use and application of matrix algebra as Duncan and Collar, have admitted the superiority of the Escalator method in this respect.